



Communications  
Canada

BIBLIOTHEQUE  
Industrie Canada

THE OVERLOAD

OF

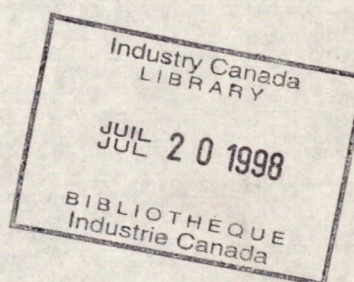
COMMUNICATION SYSTEMS

James R. Taylor

P  
91  
.C655  
.T39  
1972



P  
91  
.C655  
.T39  
1972

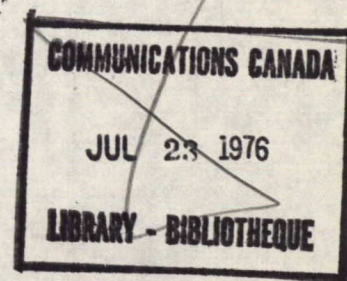


THE OVERLOAD

OF

COMMUNICATION SYSTEMS

James R. Taylor



THE OVERLOAD  
OF  
COMMUNICATION SYSTEMS

James R. Taylor

Section de communication  
Université de Montréal

December 13, 1972



THE JOURNAL  
OF  
COMMUNICATIONS

Volume 10, Number 1

P91  
.C655  
.T39  
1972  
C.1



### ACKNOWLEDGEMENT

The work leading up to the preparation of the present report was supported in part by a contract with the Social-economic Research Branch of the Department of Communication in Ottawa.



## TABLE OF CONTENTS

---

### CHAPTER I: INTRODUCTION

The Prospect of Information Overload in Technologically Advanced Societies.....	I - 1
Channel Capacity .....	I - 5
Switching Capability .....	I - 6
The Effect on the Individual .....	I - 7
The Motivation for the Present Study .....	I - 8
Situation the Present Report with respect to Previous Research .....	I - 11
Communication and Information: a brief review of some basic concepts .....	I - 16
Communication and Networks .....	I - 16
Three Principal Questions .....	I - 19
Shannon's Theory of Information Transmission .....	I - 19
Information and Probability .....	I - 20
Essential Notions of Probability .....	I - 23
Redundancy .....	I - 27
Channel Capacity and Coding .....	I - 28
Coding .....	I - 33
Noise .....	I - 36
Modular Networks and Communication Theory .....	I - 39
Communication and Control .....	I - 45
Plan of Succeeding Chapters .....	I - 51



## CHAPTER II: HYPOTHESES AND FINDINGS

Organization of the Chapter .....	II - 1
Reaction Time Experiments and the Concept of Channel Capacity ...	II - 3
The Information-Theoretic Interpretation of Choice-reaction Time Experiments: early experiments ....	II - 9
The Hypothesis of Information Overload .....	II - 13
Overload at the Level of the Cell .....	II - 17
Direct Evidence of Overload at the Level of the Individual Organism ....	II - 22
Tracking Experiments .....	II - 23
The Theoretical Basis of the Information Overload Hypothesis .....	II - 27
Signal Rate of Arrival as an Information Parameter .....	II - 27
Size of Ensemble as an Information Parameter ...	II - 29
Modelling the Human as a Communication Channel: an assesment .	II - 32
Problems of Interpretation .....	II - 33
Experimental Findings: problems and inconsistencies .....	II - 35
The Communication Channel Model in Perspective .	II - 41
Modelling the Organism as a Multi-stage Adaptive System .....	II - 43
The Principle of Central Intermittency .....	II - 43
The Principle of Decomposability .....	II - 46



Attention .....	II - 49
The Operation of the Buffer Storage .....	II - 52
The Operation of the Selector .....	II - 54
Pre-attentive Processes: some conclusions .....	II - 62
The Identification of Stimuli .....	II - 66
Fitting the Stimulus to the Memorial Representation .....	II - 66
The Memory Search Process .....	II - 67
The Stage of Response Selection .....	II - 74
The Stage of Response Execution .....	II - 76
Summary .....	II - 80
Implications for Further Research .....	II - 81

### CHAPTER III: COMMUNICATION OVERLOAD

Limitations of the Concept of Transfer Function .....	III - 2
Information Conservation .....	III - 3
Information Reduction .....	III - 4
Information Creation .....	III - 6
Plans and Behavior .....	III - 12
The Efferent Control of Sensory Input .....	III - 14
Pre-attentive Processes .....	III - 18
Perceptual Analyzers .....	III - 20

Symbolic Representation of Stimuli .....	III - 23
Feedback and Information .....	III - 26
Hypothesis Testing and the Formation of Concepts .....	III - 29
The Internal Environment .....	III - 34
The Role of Communication in the Information Processing System .....	III - 37
Problems of the Determination of Referential Information Content..	III - 39
The Internal Structure of The Symbol System ..	III - 40
Problems of the Determination of Relational Information Content...	III - 45
Network Overload .....	III - 54
The Transmission of Information in Network....	III - 58
Experimental Results .....	III - 58
Relationship and Information Transmission ....	III - 60
Experimental Support .....	III - 63
Perspective .....	III - 64

## BIBLIOGRAPHY



CHAPTER I

I N T R O D U C T I O N

Few technological changes have had so profound an effect on the human condition as the development of telecommunications. Man today lives in a maze of electronic signals; it is certain that their influence on the quality of his environment will be even more important in the future than is the case today.

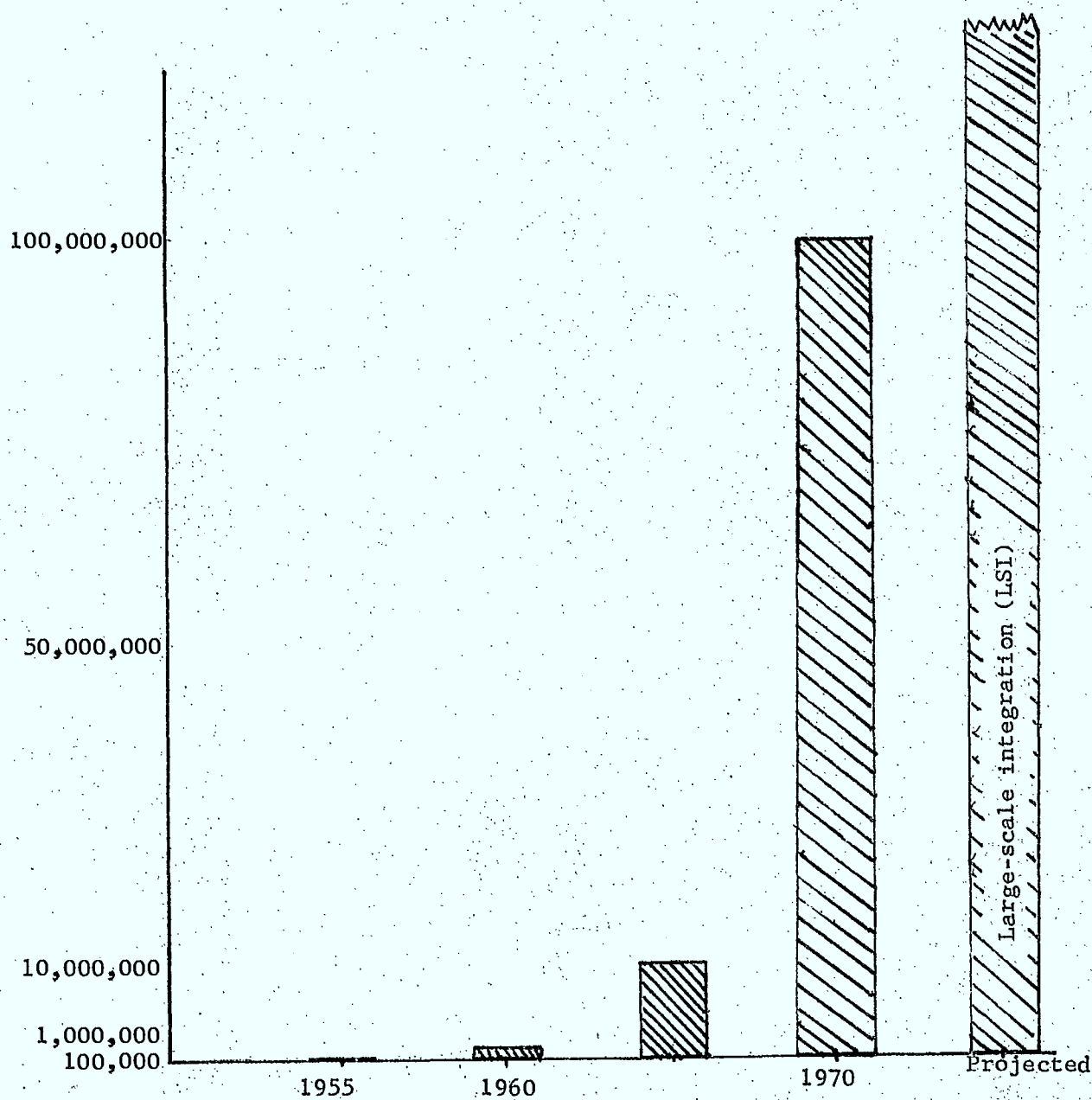
Final Report President's (U.S.) Task  
Force on Communications Policy, (1968)

### The Prospect of Information Overload in Technologically Advanced Societies

An extraordinary feature of modern telecommunications systems, present and projected, is the extent to which they are characterized by rapid and accelerating technological change. The introduction of innovations in the field of communication, particularly for systems of information transmission, has expanded from a trickle to a stream and is now turning into a torrent. In some respects telecommunication appears to stand now in terms of comparative development where transportation stood nearly a half century ago, although the comparison may actually underestimate the dimensions of the present transformation. One or two examples will help to show the extent of the present rate of change. In the field of computing where the entire development has occurred since World War II, the growth is illustrated in Figure I-1.

FIGURE I- 1

Number of  
computer  
instructions  
executed  
per dollar  
cost



Computing capacity / cost ratio , 1955-1970

Source: James T. Martin, Future Developments in Telecommunications.



What Figure I shows is that the computing-power-to-cost ratio has been increasing by a factor of 10 (exponentially) every five years! The introduction of mass-produced LSI (large-scale integration) circuitry will assure the continuation of this trend, at least in the immediate future.<sup>1</sup>

The transmission of digital information still lags behind computing capacity; however at least one company (Datran) is proposing a system which offers up to 14,400 bits per second, and ITT has announced the development of a Digital Data Network. Other transmission rates have been exhibiting trends somewhat comparable to those for data computation rates. Figure II shows trends in the transmission of telephone conversations.

A similar pattern can be observed with respect to increases in radio bandwidth: successively, we have seen the introduction of long wave (frequencies up to about 100 kHz), the "broadcast" band (around 1 MHz), shortwaves (around 10 MHz), millimeter waves (large-scale use of frequencies above 10,000 Mhz).

Satellite transmission is of even more recent origin (from one satellite in 1965 with a capacity of 240 voice circuits to 11 in 1971 with a total capacity of 10,000 circuits). The present capacity is predicted to expand very rapidly in the foreseeable future, and satellites have the additional important characteristic that they largely eliminate the usual association

---

1. For discussion, see, for example, the February 1970 issue of Scientific American.

FIGURE I-2

I-4

Number of  
telephone  
calls carried  
long-  
distance  
cables

Laser  
Optical Waveguide  
No Practical  
Limit

200,000

100,000

50,000

10,000

1,000

100

1940

1950

1960

1970

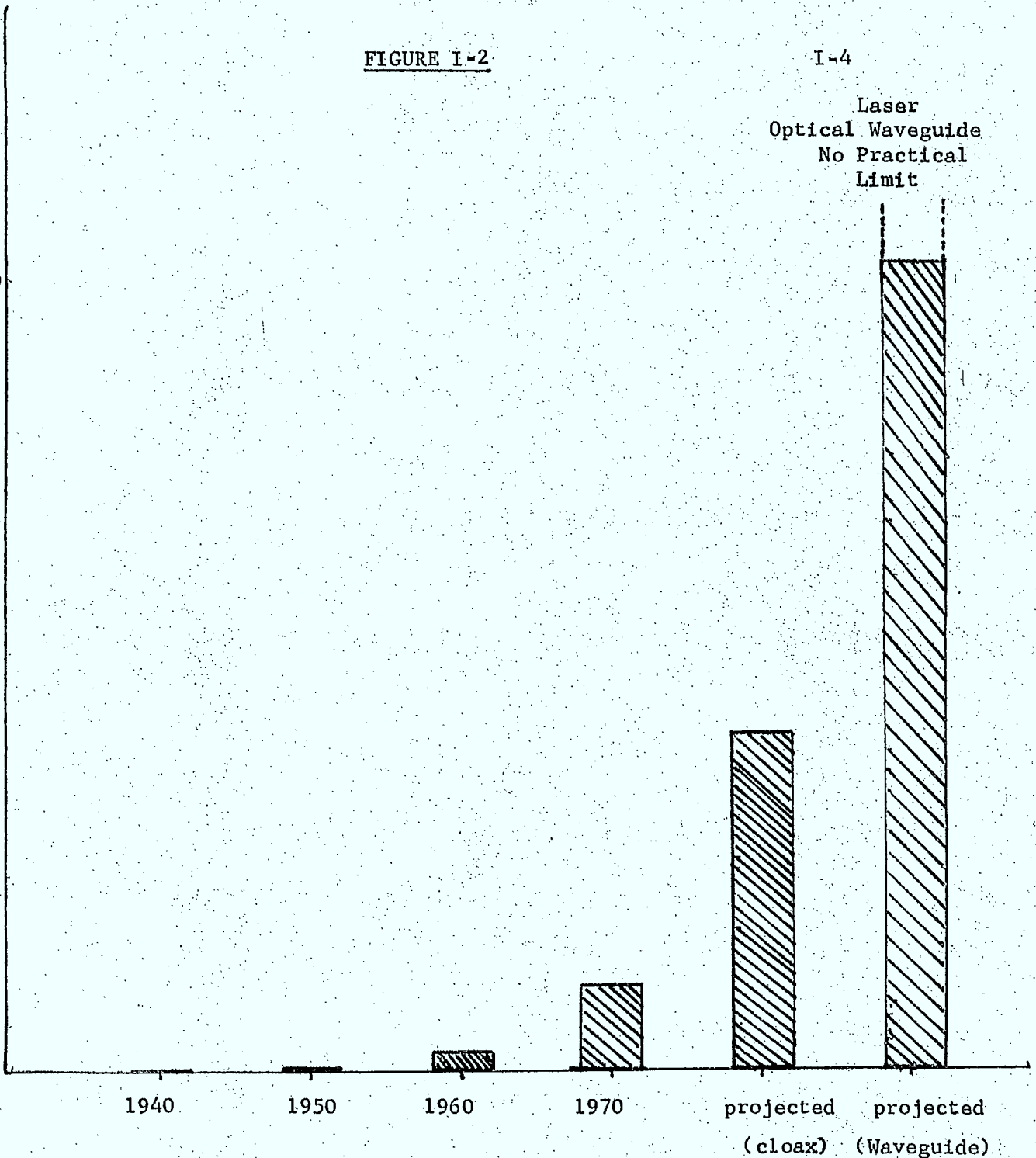
projected

projected

(coax) (Waveguide)

Carrying capacity, long-distance telephone cables, 1940-1970.

Source: James T. Martin, Future Developments  
in Telecommunications.



between cost and distance.

The examples given here are simply indicative of an overall phenomenon, the consequence of which will be to facilitate immediate and universal information availability.

Increases in the rate of transmission of information involving long distance communication have resulted from two related types of technological development:

- a) increases in channel capacity, i.e., net additions in the volume of information which can be transmitted from one point to another in a given time, and
- b) increases in switching capability, i.e., an expansion in the means available to form combinations of inter-nodal links into discriminably different networks.

#### Channel capacity

-----

Growth in channel capacity has resulted from changes in the speed and volume of transmission of signal, additions to the kind of signal that can be effectively transmitted at high speed over long distances - - voice, print, image - - and, by derivation, in the quality of representation of the original message which can now be achieved. Changes in quantity and kind of transmittable signals came about with the introduction of new modes such as radio (voice), television (image), and facsimile transmission (print). The establishment of standards with respect to quality of transmission has been a gradual process. In telephone transmission, the goal of engineers has been typically



limited to the "simulation of presence", or a level at which sufficient supralinguistic information such as tone of voice, inflection, etc... is transmitted, as well as more basic content information, to create somewhat the illusion that another person is present, at least in part, depending on the sense modality employed; the visual equivalent of simulation of presence remains to be accomplished.

In every field, however, the constraints of cost are gradually yielding to technological advance.<sup>1</sup>

#### Switching capability

-----

Among telecommunication systems, fully switched networks are represented by the postal service and the telephone system. Their flexibility is obtained at a cost: both systems have limited local channel capacity. In the past there seems in fact to have been a tradeoff between channel capacity and switching capability. Television, with greater channel capacity than either telephone, telegraph, or post, has been until recently considered primarily an area-wide distribution system. However while it may have once appeared that greater information transmission could be gained either through increases in channel capacity, or through more flexible switching patterns, but not through both, this constraint in turn seems now in the process of being slowly pushed back. Eventually, we may look forward to fully-

---

1 It is worth noting, parenthetically, that increases in telecommunicative channel capacity have been accompanied by increases in memory capacity as the variety and fidelity of recording mechanisms has evolved.

switched maximum capacity systems with, for practical purposes, unlimited range of geographical distribution.

The effect on the individual

For the individual, the effect of these developments has been that (a) he is each year (as part of an accelerating process) the target of more messages, each with augmented informational content, and (b) because of his access to larger and more differentiated networks, he tends to interact within larger systems having greater complexity of organization, which in turn is correlated with increased variability, and hence with greater information. Thus, on two counts, technological advance in the field of telecommunication is closely associated with the amount of information available to the individual. It seems reasonable to assume that such a process must eventually reach a ceiling, which is the limit of individual human beings to accept and process information.

"In a metropolitan area of 5 million population, about 4,800 hours per year per capita (or about thirteen hours per day) are devoted to various modes of reception of social communications, such as reading, television, lecture and discussion, observation of environment, radio, film and miscellaneous... At various estimated receiving rates of non-redundant bits per minute, the per capita average reception of information is 100 millions bits per year or roughly, 300 bits per minute. It is startling to note how close this is -- within a factor of 5--to the 1,500 bits per minute taken as the limit of human capacity to absorb information....

As the amount of necessary information per capita grows, the limits of human capacity may be pressed at least for many". (2)

---

(2) Meier (1962), also cited in a report of the American Committee on Telecommunication, National Academy of Engineers (1969)

It has been more than once asserted, and recently strongly reiterated, that the individual's capacity to function and to make decisions must eventually break down under the strain of environmental overstimulation.

"The striking signs of confusional breakdown we see around us -- the spreading use of drugs, the rise of mysticism, the recurrent outbreaks of vandalism and undirected violence, the politics of nihilism and nostalgia, the sick apathy of millions -- ... may well reflect the deterioration of individual decision-making under conditions of environmental overstimulation". (3)

Lipowski (1971) has similarly attributed unrest, anomie, and violence to the influence of the widespread exposure of individuals living in affluent, technological and open society to what he terms an "overload of attractive stimuli". The condition in which extreme information processing demands result in temporary or permanent system breakdown has been termed information overload.

The motivation for the present study

The present study was undertaken in response to the evident need to obtain a clear idea of the danger of information overload, insofar as it can be assessed now from scientific investigations conducted either in a psychological laboratory or in natural settings. Initially, the objective was in essence to conduct a simple review of the available literature on the subject. The report of this work takes up Chapter 2: there is in fact an extensive literature on the topic.

---

(3) Toffler, 1971.



However, while a number of interesting and useful conclusions can be drawn from the literature, most of the relevant research, with rather few exceptions, has been conducted within the limits of psychological laboratories, rather than in the field. For a report which is intended to have practical implications with respect to the development of social policy, this strongly psychological (and theoretical) orientation presents something of a dilemma. As we shall find in the second chapter, the kind of individual information processing studied within the psychological laboratory may often appear to constitute a very special subset of the whole spectrum of information transmission behaviors to be found in naturalistic social settings. The term "information" itself is assigned a very particular meaning, which only very partially reflects what we usually think of as information.

This deliberate restriction of the field of inquiry has disadvantages and advantages. The main disadvantage has already been alluded to: the great difficulty in generalizing from laboratory findings to naturalistic contexts. This problem is not unique to the questions of information overload, of course, but in no other field is the constraint more irksome.

The great advantage of the experimental method, in spite of its substantive limitations, is that the concepts which have evolved are (remarkably) more precise than those developed in field investigations, and conclusions are more solidly documented by firm evidence. The domain of cognitive psychology, which

is that part of psychology concerned with the investigation of information processing systems within the individual, is presently in a state of very lively development. Exciting new ideas and approaches are rapidly being realised. The very theoretical bases on which the literature on information overload was first constructed are themselves being modified. It thus is unthinkable that we should ignore these developments.

How then have we attempted to meet the dilemma? It would be dishonest to pretend that the whole difficulty has either been resolved or can so easily be wished away. However we believe we have found at least a partial resolution: the discovery of the crucial link between an approach narrowly based on classical information theory (in its more limited sense) and one which aims to say something about real persons in actual social contexts is a task for communication theory. The linchpin between individual and society is communication. In the third chapter, therefore, an attempt is made to restate the original psychologically-based theory in terms of communication theory, and to reexamine some evidence concerning information overload in the light of this statement.

The net effect of the method adopted here may be a document which appears somewhat abstract: this is unfortunate but it seems preferable to attempt a rigorous, if very incomplete, attack on the question of information overload, rather than to

become mired in doubtful if exciting generalizations about trends in society, based on dubious and sketchy information.

Situating the present report with respect to previous research

Modern discussions of the problem of information overload reflect the influence of two basically different approaches to the study of the human organism as an information processor, one derived from the field of engineering, one psychological in origin.

Among engineers, interest has for some time been focussed on the role of humans as components within larger systems. By the beginning of the twentieth century, this concern was already well established. As the coordination of work activities in industrial society became functionally more complex, there was a strengthening of interest in the possibility that the operations performed by men could be analysed in terms compatible with those used for machines. The school of "scientific management" associated with the name of Frederick Taylor (1907, 1911, 1919, 1947) became interested in the problem of how to describe with exacting precision the task-performance characteristics of the human component. Through "time and motion" studies it was hoped that physical tasks performed by humans could be specified in the form of detailed programs of behavior, or "methods", and eventually, through systematic training, that the work efficiency of individuals could be radically upgraded and their notorious unreliability reduced. The



emphasis of Taylor and his associates was on energy-transforming tasks.

The original scientific management movement was gradually absorbed into other management schools of thought preoccupied with the growing trend to automation (Diebold, 1952). It soon became evident in any case that Taylor and his followers had seriously underestimated the importance of motivational factors in work performance (Mayo, 1933; Roethlisberger & Dickson, 1939). However, human factors engineering received a fresh new impetus during the period of the second World War with the emergence of cybernetics as a domain in its own right. (Wiener, 1948; Shannon, 1948). The attention of engineers accordingly turned from the energy-transforming capabilities of the human organism to his information-processing characteristics. Thus, while Sinaiko and Buckley (1957, 1961) state their objectives in terms reminiscent of those of Taylor:

Machines do not operate by themselves. Even in an age of automation men will be involved in one way or another in every system... Men as well as machines are components of systems. Since mechanical and electronic components are now available with very high speeds and capacities, the design engineer's task of integrating men and machines into smoothly functioning systems has become more difficult. If the characteristics -- limitations and capabilities -- of humans are known and understood, better man-machine systems will be designed and built,

nevertheless, the list of variables to be considered now includes

a concern not only for energy-transforming constraints (the physical dimensions of the individual, his capability for motor activity, his physical needs) and motivational factors (psychological needs, capability for learning, sensitivities to social environment), but also for information-processing constraints (capability for data sensing and processing).

Psychologists, on the other hand, have been interested in the processes of perception and the informational determinants of reactions by humans since the nineteenth century, indeed since the emergence of the discipline in its modern form.

Psychologists' interest in the problem of human information processing appears to have been further encouraged by a convergence of the paradigm of information transmission used by Shannon with models of the organism employed by psychologists in the field of learning theory. Both the philosophy and the methodology of behavioristic S-R psychology encouraged the investigation of humans as input-output systems with essentially linear transformational properties (or some stochastic approximation thereof); the role of learning or conditioning is to assure that, presented with a given stimulus, the organism will produce an associated set of responses (with the association between stimulus and response meeting certain standards of reliability). The study of human behavior, in this perspective, then reduces to the proper description of ensembles of stimuli

and responses, and the observation of behavioral correlations between them.

It seems to have been in part because the communication model of Shannon could not only be made to conform to this traditional psychological paradigm, but also promised a more accurate quantification of stimulus-response ensembles, that the work of Wiener and Shannon was so quickly absorbed into the mainstream of psychological research. Much of the work reported in this study is set within this integrated S-R and information-transmission model.

Information theory, however, is by no means the only contribution of cybernetics to the illumination of the problem we are studying in this inquiry. For one thing, cybernetic theory provides an alternative model to that of the reflex arc, which, even long after being officially discredited in psychology, seems to have continued to exercise a persistent (and often hidden) influence on the design and interpretation of psychological experiments. In simplest form the cybernetic model of the individual information-processor is illustrated by a computer program which is executed step by step and which utilizes input in order to perform tests and operations in order to produce an output. In this model, output is a function both of input and of the internal states of the organism. The application of this alternative cybernetic model for purposes of experimental

research has occurred relatively slowly, presumably because of the rather great change of perspective which is required by comparison with the use of the simple S-R model.

A major goal of the present study is to reexamine the literature on information overload in the light of the general cybernetic model. The remainder of this introductory chapter is accordingly devoted to a review of some of the main insights which have animated subsequent discussions set within the cybernetic framework.



## Communication and Information: a brief review of some basic concepts

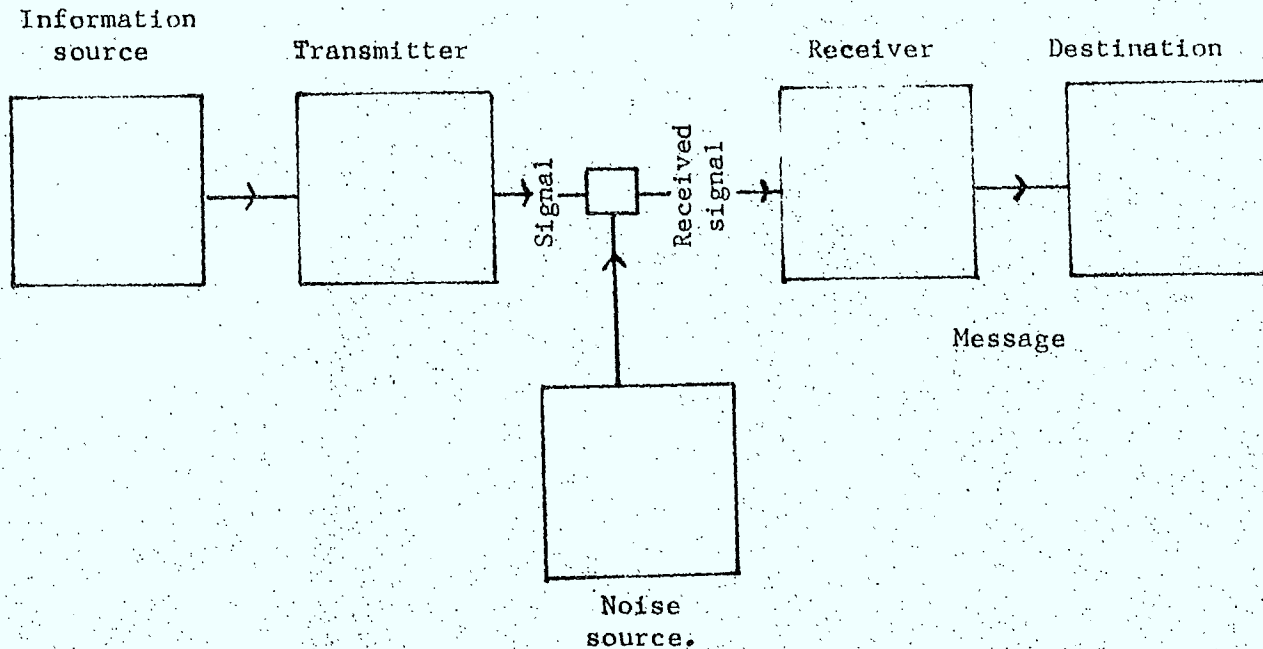
Much of the succeeding discussion, particularly Chapter 2, assumes a basic knowledge of communication theory. It may be useful therefore in this introductory chapter to digress briefly in order to sketch in some of the principles underlying the modern theory of communication and control. For those interested in a fuller explanation, many excellent treatments exist. On the other hand, readers already familiar with the theory of communication and control may want to omit this section, which rehearses only quite well known material.

The contemporary approach to the study of communication and control centers on the use of mathematical models to describe biological, social and artificial systems. The mathematical models are neutral in the sense that they are axiomatic in their development, and hence, while they may be internally consistent, in and of themselves they make no statement about any world. Their application to actual systems requires an act of interpretation. We shall see in Chapter 2 that while such theoretical models may provide powerful explanatory tools, they may also impose their own limitations, depending on how they are applied.

### Communication and Networks

The concept of communication implies a network. In its reduced form, the communication situation contains a sender and a receiver (though on occasion sender and receiver may be the same individual: eg. someone writing himself a memorandum, to be read at a later time). Claude Shannon (1948) proposed the following schematic diagram of a general communication system:

FIGURE 1-3

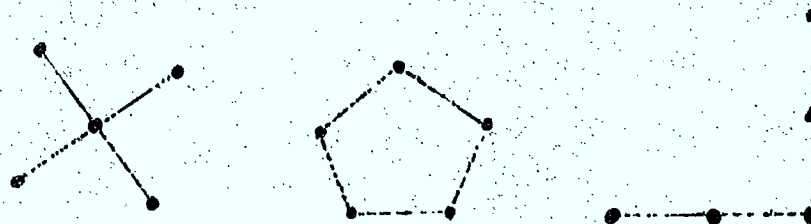


The Shannon schematic diagram of a general communication system

In Shannon's model, the information source is responsible for the production of a message, or messages, which are to be communicated. The transmitter operates on the message in order to transform it into a signal, which is then transmitted through a channel to a receiver, which in turn operates on the signal to produce a message for the destination, or the intended recipient of the message. Noise, or interference, is conceived as an additional (unwanted) information source which introduces signals into the system which compete with the main signal transmission, and may thus make it more difficult for the intended message to be clearly inter-

The communication situation can also be conceptualised in another way. Let  $X = (x_1, x_2, \dots, x_n)$  be a set of potential communicators.

Then a potential network exists if there is a possibility for at least one communicator to exchange messages with another communicator, i.e., if there exists a communication channel (channels) by which messages can be transmitted from  $x_i$  to  $x_j$ , and from  $x_j$  to  $x_i$  (assuming for the sake of simplicity that all channels are bi-directional, although the assumption of symmetry is not required).



### Formally Distinct Communication Networks

One way in which these intuitive notions can be formalized mathematically is to define a mapping,  $\Gamma$ , on the Cartesian product,  $X^2$  (where  $X$  is defined as before), such that  $(x_i, x_j) \in \Gamma$ , if and only if  $x_i$  can communicate with  $x_j$ . Then  $N = (X, \Gamma)$  is a graph, and  $(x_i, x_j)$  is called an arc. In this way, the notion "graph" has been equated with that of "network", and that of "arc" with the concept of a "channel".

1. Vitiations due to noise are shown as entering the system at the point of the channel; this is a convenient abstraction, since interference may equally well be due to characteristics of the encoding and decoding process.

### Three principal questions

This development of the idea of channel of communication is highly restrictive in that it deals only with the possibility of communication, i. e., the existence of a channel by which to communicate: nevertheless, it allows us to ask a number of questions which will serve to introduce certain essential notions of communication theory. First, we might want to know how to describe the capacity of the channel, remembering always that we are to measure the transmission, not of matter, but of information. This is essentially the question which was posed by Shannon.

Second, supposing that we know something about how individual elements of the network behave when presented with certain messages, what can we usefully say about the behavior of the network as a whole, i.e., what are the probable outputs of the whole network for given inputs, and for given internal states? This is the question which was asked by McCulloch and Pitts, in their classic paper (1943).

Third, what can be said with respect to the explanation of purposive behavior if we suppose an element or a network to be joined to an environment by a return loop? This is the question which interested Wiener (1948), and Ashby (1952). Together these three topics constitute the starting point for the modern theory of general systems of communication and control.

### Shannon's theory of information transmission

Weaver (Shannon and Weaver, 1949) identified three levels



on which the communication of messages can be analyzed, each of which raises a somewhat different order of problem:

- LEVEL A. How accurately can the symbols of communication be transmitted?  
(The technical problem)
- LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem)
- LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem)

Shannon originally entitled his paper The Mathematical Theory of Communication, and in it he was uniquely concerned with Level A problems, that is to say problems associated with the transmission of signals. The definition he gives to the term "information" has therefore a limited domain of application, so much so, indeed that it has since been proposed that he might more properly have termed his paper "The Theory of Signal Transmission and Coding" (Bar-Hillel, 1964). In similar vein, MacKay (1969) discriminates between "information" and "amount-of-information" or "information-content", where the latter terms refer to what Shannon calls "information". While these distinctions will become important in our later discussion, for the moment all that is necessary to bear in mind is that Wiener and Shannon have a restrictive sense in mind when they use the term, one that in no way captures all of the meaning usually assigned to the word in our ordinary discourse.

#### Information and probability

The mathematical theory of communication, as developed by Wiener,

Shannon and others is closely related to the mathematical theory of probability, from which it derives most of its postulates. As in elementary probability theory, it is convenient to define an ensemble of possible outcomes of an observation (such as an observation of the behavior of a potential communicator). Then a communication event (i.e. the transmission of a signal) is simply the occurrence of a subset of the ensemble of possible events. To illustrate, suppose a traffic policeman has manual control over a traffic light. The ensemble in which the motorist is for the moment interested consists of the possible states (three) of the traffic light: red, green and amber. If only one light occurs at a time, i.e., if the possible communication events are mutually exclusive, the traffic policeman can signal permission to oncoming motorists to pass by causing green to appear on the light standard, caution by amber, and non-permission by red. The meaning of the events, "go", "slow", "stop" is however not of immediate relevance to the mathematical theory of communication. In this example, the ensemble consists of the available symbols, ("red", "amber", "green"). Message-transmission depends upon the occurrence of a signal event, i.e., the occurrence of some possible combination of symbols, (eg. "red"). It is signals which are transmitted through channels, and our task is to determine the capacity of channels. Information is simply a measure of the freedom of choice in the selection of a message from an available set of messages (where a message is construed to mean an acceptable combination of symbols).

What we want to know is how such freedom of choice is to be measured: i.e. what is an appropriate metric, or measuring stick?

Suppose the ensemble (or repertoire) to contain  $n$  symbols, and

suppose further that each symbol is equally likely to occur, independent of the occurrence of any previous symbol. In this case it is obvious that the measure of amount of information conveyed by the occurrence of any message-event (containing just one symbol) should be a direct function of  $n$ , since the freedom of choice increases (decreases) as  $n$  increases (decreases). It might even seem reasonable at first to take  $n$  itself as the measure of amount of information.

This solution is, however, not adequate when we relax the constraints to permit compound message-events, i.e. those containing more than one symbol. Suppose the ensemble of "simple" messages (i.e. messages having just one symbol) consist of two signals, eg., "yes" and "no". Suppose that all transmitted messages actually contain three symbols. In this new ensemble of compound messages (combinations of symbols) there are eight possible occurrences of message-events:

1. Yes, yes, yes
2. Yes, yes, no
3. Yes, no, no
4. No, no, no
5. No, no yes
6. No, yes, yes
7. Yes, no, yes
8. No, yes, no

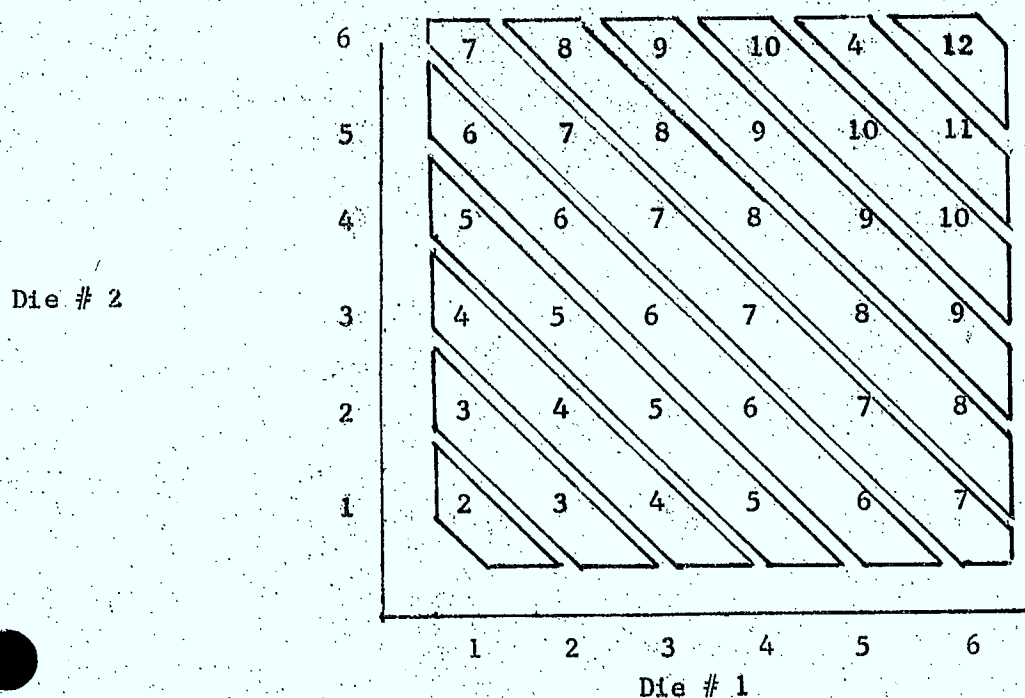
Since a "complex" message containing three symbols can be thought of as being built up out of elementary components, equivalent to simple messages drawn from an ensemble of  $n = 2$ , and assuming our measure of information to be the size of the ensemble, the combined information content might be presumed to be  $3 \times 2 = 6$ . The actual  $n$ , and hence the information content, of the compound-event ensemble is, however, 8. Thus to take  $n$  itself as a measure of information leads to certain complications.

This inconsistency can be resolved by using an appropriate mathematical transformation of  $n$ , (in fact a logarithmic transformation), Any number can be expressed as a logarithmic transformation of another number, so that the use of logarithms as a measure of amount of information implies no loss of information; in the present case such a transformation proves to be convenient. Shannon, for intuitively good reasons, chose the number 2 as the logarithmic base of information theory. Thus, returning to the example given above, the amount of information (measure of freedom of choice) of an ensemble with two symbols is  $\log_2 8 = 3$ , so that messages which are compound events consisting of combinations of three symbols have exactly three times the information content of those made up of a single symbol, as intuitively they should. The term bit was coined to express the idea of the amount of information associated with the occurrence of one of two equally probable signals. The occurrence of a message-event drawn from the ensemble of eight possible events, in the example given above, thus has an information-content of three bits. A "bit" therefore is the basic measurement unit of information theory.



# Essential notions of probability

In the discussion above, the notion of probability has been introduced by the back door. Since it cannot freely be assumed that message-events always have equal likelihood of occurrence or are unaffected by the prior occurrence of other message-events, it is important to develop more fully our concept of probability. We can do so with the following example: suppose the basic message-event to be the throw of a set of dice. The source of the message is "Chance", the mechanism by which the message is transmitted ("the channel") is the throw of the dice, and the actual message-event is the combined value which shows on the two faces of the dice which face upwards, i.e., a natural number between 2 and 12. The message is drawn from an ensemble which is illustrated in the following diagram:



From this diagram, it is evident that certain combinations of simple messages are lumped together to create compound messages: thus a 1-6, 2-5, 3-4, 4-3, 5-2, or a 6-1 are all called a "seven". From this fact, and assuming that the dice are "unbiased" to start with (i.e. that the chance of occurrence of every simple message is equal), we can draw up the following table of probabilities:

<u>Compound message-event</u>	<u>Probability</u>
2	1/36
3	1/18
4	1/12
5	1/9
6	5/36
7	1/6
8	5/36
9	1/9
10	1/12
11	1/18
12	1/36

Since we have supposed the probability of occurrence of each symbol (i.e. 1, 2, 3, 4, 5, or 6) to be equal for each die taken individually, then the amount of information transmitted by each throw of a single die is  $\log_2 6 = 2.585$  bits (or 5.17 bits for two throws of a single die).

To find the information-content of the compound event associated with the throw of the two dice together, we use the formula:

$$H = - \sum_{i=1}^N p_i \log_2 p_i, \text{ or } (2(1/36 \log 1/36) + 2(1/18 \log 1/18)$$

$$+ 2(1/12 \log 1/12) + 2(1/9 \log 1/9) + 2(5/36 \log 5/36) + 1/6 \log 1/6) \text{ or}$$

$$(1.0612 + .8906 + .7342 + .6254 + .2435 + .2732) = 3.8381 \text{ bits.}$$

The difference between 5.17 bits, which is the amount of information which would have been transmitted if each of the 36 "simple" message-events in the original ensemble of potential symbols had been assigned an individual message value, and the actual value of 3.8381 bits obtained, is a measure of the constraint which results from an unequal distribution of probabilities, and the assignment of structure to the original message-space.

Not only may separate message-events have unequal probabilities which have to be taken into account in the evaluation of amount of information transmitted, there is a further possibility that the probability of occurrence of a certain message-events is affected by the neighborhood of other symbols in which it occurs. The equivalent concept in probability theory is that of conditional probability, which is defined formally as follows:

for two events, A and B, the conditional probability of B, given A, is defined to be

$$\frac{P(B|A)}{P(A)} = \frac{P(B \cap A)}{P(A)}$$

where  $P(A)$  is the probability of occurrence of A,  
 $P(B|A)$  is the probability of occurrence of B, given that A has occurred,  
 and  $P(B \cap A)$  is the probability that both B and A occur.

#### Redundancy

-----

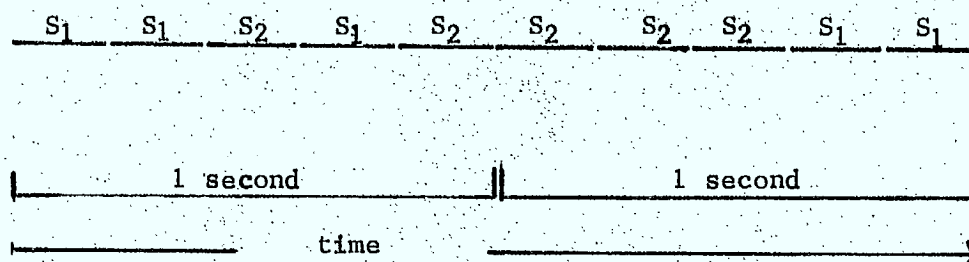
In natural languages, some letters are associated with others in special patterns: in French and English, for example, the letter "u" follows the letter "q" with a very high probability. In the context of "q" the occurrence of "u" has negligible information-content, irrespective of its usual value if we were to consider only the relative frequency of occurrence of the letter in the language (i.e. averaged over all its environments). Such serial dependencies are not limited to first-order correlations of the type just described, i.e. where one symbol follows, precedes, or is in some other way in the immediate neighborhood of the other. From the letters "blank, t, h, e, i, ..., blank" we could without difficulty fill in the "missing" letter, "r". In this instance, no single preceding or succeeding symbol establishes the constraint which allows us to supply the absent letter; rather our sense of pattern is due to our sensitivity to higher-order contingencies. Inter-signal dependencies of this character are referred to as redundancy, and can be dealt with in the measurement of information without difficulty by the use of the probability calculus described above.

Channel capacity and coding  
-----

To this point, we have been concerned with developing a means to measure the quantity of information transmitted by a given sequence of message-events. We now turn to two slightly different questions: (1) how can we ascertain the capacity of a given channel to transmit information? and (2) how can messages be re-coded in order to make the best possible use of a given channel capacity? This latter question will lead in turn to the problem of interference, or noise, and means to combat it.

Messages are transmitted over channels (physical media of some kind) by mechanisms which are capable of being in one or the other of a certain number of different states (eg., a telegraph transmitter, the human speech production system, etc). Such a mechanism is capable of transmitting a certain set of signals,  $S_1, S_2, \dots, S_n$ , each with a corresponding duration,  $t_1, t_2, \dots, t_n$ . Signals might take the form, for example, of pulses which vary in (1) intensity and (2) duration. The simplest case is that of an ensemble of two signals,  $S_1$  and  $S_2$ , of equal duration, i.e.  $t_1 = t_2$ . Then the channel capacity is  $\log_2 2/t$ , or  $1/t$  bits per second. Suppose  $t$  to be equal to  $1/5$  second; then this channel can transmit 5 bits per second.



FIGURE I-4Messages:

Capacity =  $\log_2 N/t$  bits per second = 5 bits per second.

$N = 2$

$t = 1/5$  second

Illustration of Transmission of Signals in a Given Time Period

The capacity of a channel, on the assumption of equal duration of signals, is limited by two factors: (a) the number of discriminably different signals it can recognize and produce, and (b) the number of separate pulses, or signals, it can handle in any given time period. The importance of these two parameters will be seen in Chapter 2.

What of the case where the signals are of unequal duration? Here we cannot simply compute the information transmitted per time interval and multiply by the number of time intervals as we did for the easier case: recall that when signals are of equal length, for some time period  $T$  made up of  $m$  time periods each of length  $t$ , the total information transmitted is  $\log_2 N(T)$ , where  $N(T)$  is equal to  $n^m$  and " $n$ " indicates as before the size of the ensemble. This follows from the fact that the information transmitted for the time period  $T$  is identical whether we think of it as  $m$  separate messages, each of duration  $t$  and drawn from a repertoire of  $n$  symbols, or as 1 message of length  $T$ <sup>1</sup>. In the latter case the size of the repertoire is  $n^m$ . (An example will make this clearer: suppose the time period  $T$  is made up of two smaller intervals each of length  $t$ , and the available symbols to be  $S_1$  or  $S_2$ . Then whether there are (a) two messages each of length  $t$ , in which case the first one may be either  $S_1$  or  $S_2$  (and similarly the second), so that there are  $2^2 = 4$  possible messages, or (b).

---

1 See again the example shown on page I-22 above.

one message of length  $T$ , drawn from the repertoire  $S_1S_2, S_1S_1, S_2S_1, S_2S_2 =$   
 4 possible messages, the variety (and hence the information transmitted)  
 is identical.

To arrive at the information which could be transmitted when  
 the signals are of unequal duration we need to find  $N(T)$ , i.e., the size  
 of the repertoire from which a single message of length  $T$  might be drawn.  
 This we can arrive at by a process of "ladder-climbing", i.e. through  
 finite difference equations. Thus suppose the initial ensemble to consist  
 of two signals,  $S_1$  of length  $t$ ,  $S_2$  of length  $2t$ . Then in the first inter-  
 val, it is possible to transmit only one signal,  $S_1$ .  $N(T)$  is therefore  
 equal to 1. In two intervals the repertoire of possible messages has three  
 elements:  $S_1, S_1S_1, S_2$ . The variety in  $N(T)$  when  $T = 2t$  is thus 3.  
 Higher values can be computed by the use of the following equation:

$$N(T) = N(T - 1) + N(T - 2)$$

Then, from the earlier discussion, we can compute the amount  
 of information transmitted for period  $T$  as  $\log_2 N(T)$  bits, and the theo-  
 retical rate of flow of information over the channel as  $\log_2 N(T)/T$  bits  
 per second. Table I-1 shows some values for our example of a two-symbol  
 channel, using signals of unequal duration:

TABLE I-1

Value of $t$	$N(T)$	Information in bits $\log_2 N(T)$	Rate of flow in bits per unit of time $\log_2 N(T)/T$
1	1	0	0
2	3	1.58	0.79
3	4	2.00	0.66
4	7	2.78	0.69
5	11	3.45	0.69
6	18	4.7	0.69
7	29	4.85	0.69
8	47	5.55	0.69
9	76	6.24	0.69
10	123	6.94	0.69

Theoretical Rate of Information Flow for a Two-symbol Channel with  
signals of uneven Duration.

Rather quickly, it will be noted, the rate of flow stabilizes at the figure 0.69, which is the capacity for this channel (Note that it is lower than for the earlier 2-symbol equal-duration example where exactly 1 bit per unit of time could be transmitted).

#### Coding

-----

Information theory thus is seen to provide a means to measure actual amount of information transmitted, and potential capacity of a given channel. It also assures that in principle a given message can be adapted to a given (noiseless) channel in such a way that the actual information transmitted is arbitrarily close to the channel capacity (though never in excess of channel capacity). The adaptation of message to channel, in order to take advantage of the peculiar characteristics of the latter, implies an ability to recode the message. An example of one coding system based on a binary fission process will serve to illustrate some of the salient points.

Our task is as follows: we wish to convey to someone at a distance the results of a series of throws of the dice (following our earlier example). There are, it will be recalled, eleven different discriminable "messages" which may occur, with quite unequal probabilities of occurrence. Our transmitter however is capable of taking one of only two different states, "0" and "1" (possibly standing for "open" and "closed"). How can we maximize the amount of information conveyed per signal transmitted? One procedure which comes close to achieving a maximum consists first in arranging the messages to be transmitted in order of probability, and second, successively partitioning them into two parts of roughly equal probability. The process is illustrated in Figure VII:



Possible mes-  
sages (Values  
of Dice)

Probabilities of  
messages

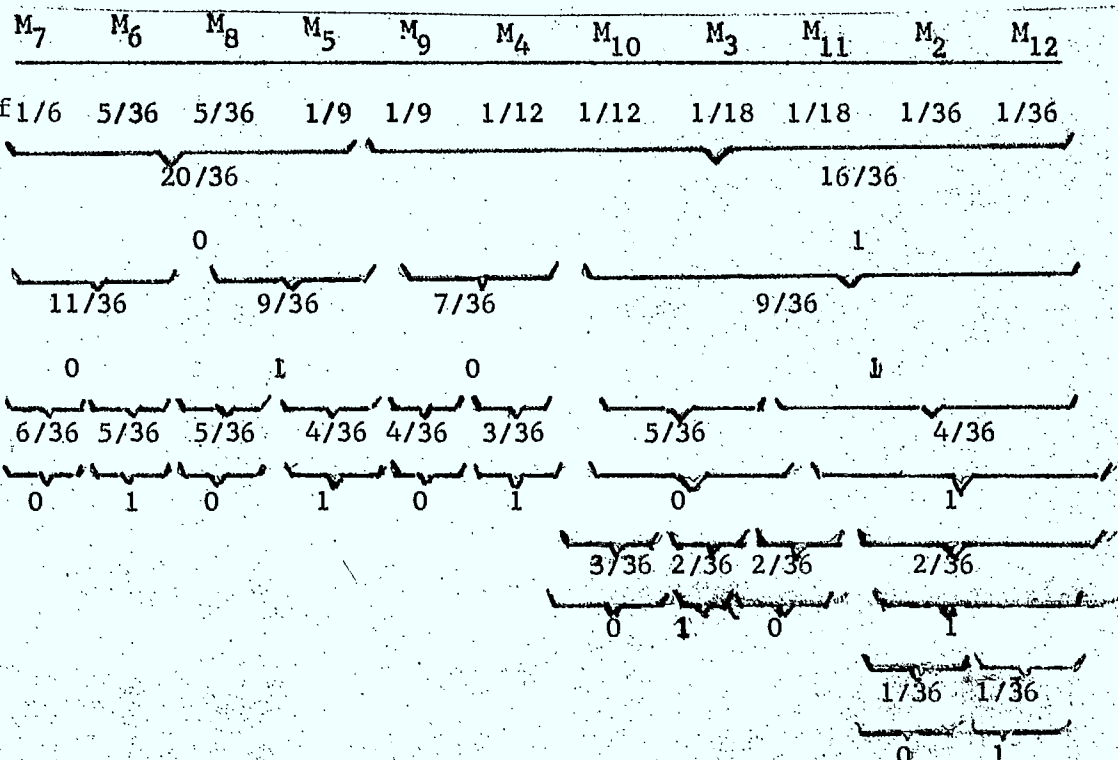


TABLE I-2: Example of a Binary Fission Process

The messages are then re-coded in appropriate form as follows:

Original message

Coded message

M<sub>7</sub>  
M<sub>6</sub>  
M<sub>8</sub>  
M<sub>5</sub>  
M<sub>9</sub>  
M<sub>4</sub>  
M<sub>10</sub>  
M<sub>3</sub>  
M<sub>11</sub>  
M<sub>2</sub>  
M<sub>12</sub>

000  
001  
010  
011  
100  
101  
1100  
1101  
1110  
11110  
11111

The coding is "optimal" in the sense that the most probable and hence the most frequently transmitted messages ( $M_1, M_6, M_8, M_5, M_9, M_4$ ) are the shortest while the least probable messages ( $M_2, M_{12}$ ) are the longest. The average length of message with these probabilities, and using this coding method, is about 3.3 digits. Using an alternative coding system, such as the following:

$M_7$	0000
$M_6$	0001
$M_8$	0011
$M_5$	0111
$M_9$	1111
$M_4$	1110
$M_{10}$	1100
$M_3$	1000
$M_{11}$	1001
$M_{12}$	1011
	1100
	1101
	1110
	0100
	0101
	0110

-- unused spare capacity

where the repertoire of available symbols is sufficiently large but where there is also "spare" capacity, the average number of digits per message transmitted is 4, which is inferior to the earlier coding system.

Optimal coding systems using signals having unequal durations can easily be devised, and in fact one of Shannon's theorems states that there exists a code which precisely matches the channel so that channel capacity can always, in principle (since the theorem does not specify how to obtain the appropriate code), be utilized to the full. The difficulty, however, is that there is a payoff between optimization of the code and

the time required to encode: the more nearly optimal the coding, the longer the potential delay in the act of coding.

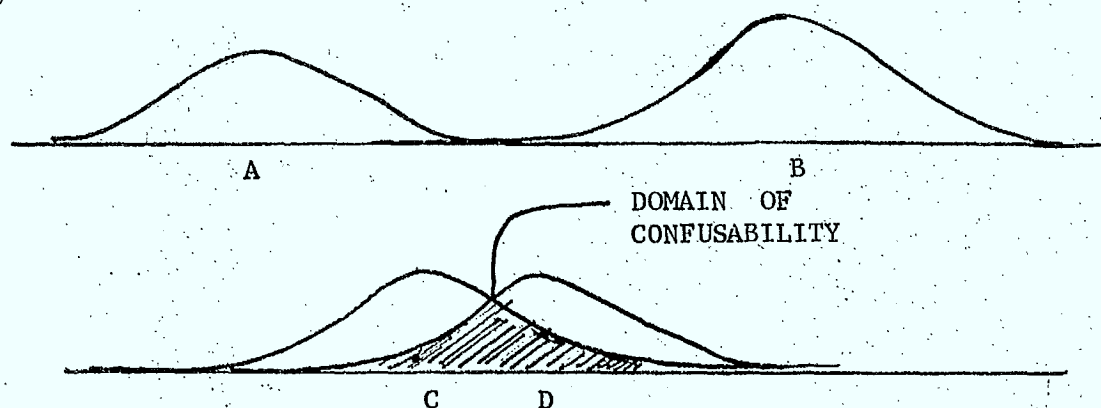
#### Noise

-----

The final distinction due to Shannon which will be employed in the sequel is that which distinguishes "noiseless" from "noisy" channels. The difference between the theory of noiseless and noisy channels is comparable to that between the theory of probability and statistics. The presence of "noise" or interference on a channel is roughly equivalent to the statistical concept of "error", i. e., effects due to variables other than those in which the statistician is immediately interested. The goal of communication (and of statistical inference) is to separate out the effects due to "irrelevant" sources from those which constitute the "real" message (the main effect). A second major theorem of Shannon ensures that even in the presence of noise, information can be transmitted over a noisy channel at a rate which is arbitrarily close to its capacity, and with an arbitrarily small error.

The principle involved is that of the appropriate use of redundancy. It is employed in statistics, in the sense that our certainty that we have correctly identified the value of a parameter is positively related to the number of observations which we take. As in statistics, where there is a tradeoff between the cost of taking additional measures in order to become more certain, and the worth of the information which is obtained by each additional observation, so in information theory there is an optimal coding procedure. The principle is simple: we reduce the likelihood of confusion between the messages by increasing "distance"

between them. (As in statistics, there is less danger of confusion between variables A and B below than there is between variables C and D):



One method of combatting noise can be illustrated as follows: suppose we wish to transmit the information that the toss of a coin has come up "heads". We choose a code which assigns to "heads" the symbol "1" and to "tails" the symbol "0". If the signal "1" is sent over a noisy channel, there is a certain probability that the wrong message will be received, i.e., that the receiver will be led to believe that "tails" came up when in fact it was "heads". We can minimize this danger by using another code. Let "heads" be represented by "11" and "tails" by "00". Now in the event of noise, if one signal only is affected, the receiver will know that there has been interference, since he will receive either an "01" or a "10", neither of which have any assignment within the given code. There is still a certain probability that he will receive a "11" when a "00" was intended, or a "00" when a "11" was actually transmitted, but this possibility has been reduced. If the coding: "heads" = "111", "tails" = "000" is adopted, the probability of absolute confusion is further reduced. If one error only occurs in transmission, the distance between signals is now sufficient that messages can still be identified correctly in spite of the effects of noise: "001",

"010", or "100" would be obtained from a "000" if one error only occurred; a "110", "101" or "011" from a "111" in the same case. The occurrence of two errors will still produce a faulty interpretation. However if the alternative coding "heads" = "11111", "tails" = "00000" is introduced the risk is further considerably reduced since three errors would have to occur in order to result in an erroneous reading of the original message. Suppose that initially there is a 25% probability of an error in transmission occurring for each digit: the probability that a mistaken interpretation will occur has now been reduced to 1 %, a very considerable gain in reliability of transmission.

This completes our cursory (and partial) examination of information theory. The meaning of the term information, as employed by Shannon is related to the degree to which a given message reduces the uncertainty of the receiver: more explicitly, it is based on the average surprisingness of messages drawn from a certain ensemble, and hence on the average prior likelihood of occurrence of symbols in the ensemble. It is the subjective state of the receiver before the arrival of the message which determines its informativeness; thus information theory, like probability theory, is concerned with expectation.<sup>1</sup>

Nowhere in this formulation is any reference made to the meaning of the signals transmitted. This has two implications with which later chapters will be concerned: (1) the model is of very wide application in the sense that it applies to phenomena we might not want to think of as communication; (2) it deals with

---

1 It is wise to note, however, that around the concept of "expectation" there has centered a controversy among mathematicians dating back to Bernoulli and some of the issues are still unresolved; the interpretation of information theory suggested here is thus not necessarily non-controversial.



one aspect only of the phenomenon of human communication, namely that which can be explained by the use of a model which is based on probability theory.

#### Modular networks and communication theory

-----

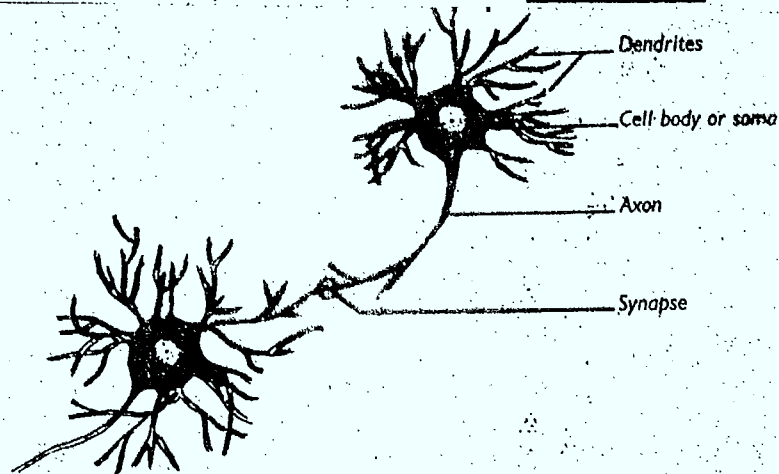
The Shannon model deals explicitly only with communication in one-way two-node (i.e. simple sender-receiver) networks. It takes us up to the point of reception of a message by receiver; it does not explicitly take into account processes of re-transmission, and hence patterns of information flow through networks composed of many elements. In this section we take a very brief look at some approaches to the study of the behavior of complete networks.

In the succeeding discussion, we will be considering information transmission and overload in two kinds of networks - the human nervous system, and social groupings of humans. In this section we will look at two classic papers having to do with the former, reserving our discussion of models of groups for Chapter 3.

The nervous system of a human being is a network of formidable size and complexity of connection. On the other hand, the information-retransmission characteristics of individual cells in the nervous system are quite straight-forward and rather well understood. This latter fact means that by imposing relatively simple restrictions it is possible to produce a model using only quite well-known mathematical operations which nevertheless captures some of the characteristics of the neural system and evokes the possibility of grasping the principles on which the functioning of the brain is based. (We shall see later that modelling of groups processes presents exactly the opposite problem: while the complexity of connection is not potentially very great for small groups, the actual elementary information transmission

processes of evolved are enormous.)

Single cells of the neural system (neurons) are composed of a central cell body (called a soma), to which are attached fibers called dendrites, and one single long fiber (called on axon) which in turn branches into a network of smaller fibers (called telodendria).



Cells are interconnected with each other in such a way that the telodendria of one cell terminate at (synapse upon) either a dendrite or the soma of another cell. Each cell operates by transmitting electrical impulses (or spikes) from the soma along the axon until they reach a synapse, or synapses, with other cells. Whether or not a cell emits an impulse at a given time, depends on the strength of the impulses which input upon it from other cells, that is, the number and intensity of signals arriving from adjacent cells at the several synapses. If in some finite interval the incoming signals to a cell together summate to a given value (called a threshold) the cell fires; otherwise it is silent.<sup>1</sup> Input signals arriving during the refractory period will not cause the cell to respond.

---

1. Actually there is some (non-zero) probability that the cell will fire in the absence of external stimulation. In addition, thresholds values may vary at different times. In general the explanation given here is highly idealized. For a more thorough treatment see for example Pribram (1971)

The interpretation of a neural net in terms of the Shannon model of an ideal communication system would suggest the following: each cell functions in turn as a receiver and a transmitter of messages. Messages are constructed by the use of a simple binary code ("fire - not fire"). Differences in intensity at the input level to the cell are reflected in the frequencies at which output impulses are emitted (the greater the input intensity, the higher the output spike frequency).

To understand how the nervous system can represent the external world in all its complexity, we have to turn to a higher level of explanation.

In a paper which has since become a classic, McCulloch and Pitts (1943) showed that the characteristics of individual cells described above, with the imposition of certain not excessively limiting further assumptions, lend themselves to precise modelling of neural networks, using only the instrument of the ordinary propositional calculus of modern mathematical logic.

They showed that by assuming, for example, that all cells fire in phase only at specific time intervals, and by ignoring differences in the time required to transmit impulses from one cell to another, a "modular network" (i.e. one composed of modules having well-defined mathematical properties which represent certain essential neuronal characteristics) can be made to perform operations similar to what we usually would call "thinking" i.e., such networks are capable of "reasoning".

In a basic propositional logic, two major types of linguistic elements may be discriminated: a) a set of elementary propositions which may be mapped to states of a possible real world ("The sun is shining", "Roses and red", "I missed my breakfast this morning"); and b) a set of

logical connectives, by means of which complex sentences may be built up out of simple units, to any extent desired ("I missed my breakfast this morning since the sun was shining and I wanted to look at my red roses"). Certain of these logical connectives ("not", "and", "or", "if", "then"...) have excited the interest of modern logicians, since their mathematical specification has provided a powerful tool for the analysis of comprehensive systems of thought based on complicated trains of reasoning. It has become common practice to use what are termed "truth tables" to define some of these logical connectives. Some examples are given below: (the column(s) on the left specify the assumed conditions in the real world, i.e. "true", "false", corresponding to the extension of the propositions linked together by that particular logical connective is still to be true in the real world):

A. INVERSION (NEGATION)

A	A
true	false
false	true

B. CONJUNCTION

A	B	A & B
true	true	true
true	false	false
false	true	false
false	false	false

## C. DISJUNCTION

A	B	$A \vee B$
true	true	true
true	false	true
false	true	true
false	false	false

## D. IMPLICATION

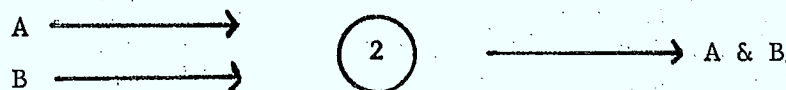
A	B	$A \rightarrow B$
true	true	true
true	false	false
false	true	true
false	false	true

To illustrate: if the statement "Roses are not red" is true, then to say "Roses are red" must be false, and vice versa (Negation). Similarly the statement "It is cold and it is raining" can be true only when both statements "It is cold" and "It is raining" are true (conjunction). However if we say "It is either cold or it is raining (or both)" then the only condition under which the statement is false is when both "It is cold" and "It is raining" are untrue (disjunction). The statement "If it rains then it must be cold" is false only in the circumstances where "It is raining" is true but "It is cold" is not.<sup>1</sup>

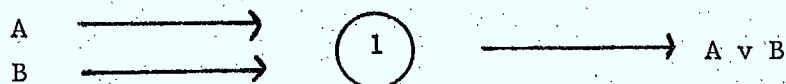
Any one of these "truth table" definitions of the logical connectives of negation, conjunction, disjunction and implication can be easily realized by a modular net. For example, the following net, with threshold "2", with two inputs, and one output, is sufficient to represent the connective "conjunction":

1 N.B.: this definition of the connective "if... then..." is rather different from that normally employed in ordinary language.





The cell fires (i.e.,  $A \& B$  is true) only under the condition that both A and B fire simultaneously (i.e., A is true and B is true). Similarly we can represent "disjunction" as follows:



Since the cell has a threshold of 1, an output occurs whenever either input is activated. This is equivalent to saying that  $A \vee B$  is true whenever A is true or B is true: Hence this type of cellular arrangement can be said to represent isomorphically the logical connective of disjunction.

By joining cells in more complex assemblies, logical ideas of considerable complexity can be achieved.

The principles embodied in the McCulloch-Pitts article are exactly those which were being employed at the time in the development of modern high speed digital computers. A digital computer is in fact a realization of a McCulloch-Pitts modular net. It can perform simple logical operations, which can be combined into complex "programs", and depending on the ingenuity of the programmer, the result is an behavior of very considerable complexity, which begins to rival , if not surpass in certain respects, that of man. The temptation is to identify machine "thinking" and human "thinking". There exists in fact a theorem which states that a suitable modular net can be made to reproduce the behavior of any finite

automation(i.e., any organism having a finite number of internal states) providing of course that we could specify exactly the input to the automation, its internal states, and the output which occurs for any given combination of input and internal states.

McCulloch and Pitts never claimed of course that their modular net in fact really represented the functioning of the actual nervous system, and it has become increasingly clear over the years that while computers and humans may both be said to "think" they hardly do so in anything like the some kind of way. To some extent indeed the strengths of one are the weaknesses of the other. In most contemporary approaches to the modelling of the human nervous system, the wavelike character of neural activity and the patterning of showers of neuronal impulses is given greater emphasis. Von Neumann argued, on the basis of Shannon's theory of optimal coding and because of the known unreliability of individual neurons, that more attention should be paid to the statistics of the neural system. What McCulloch and Pitts successfully demonstrated, thus, was not so much how the human nervous system actually worked, but rather how to address the general problem of modeling complex networks.

A second paper of equal influence to that of the McCulloch and Pitts, is that of Turing (1936). A Turing Machine has some control over its own input, and as a consequence is able to determine to an extent its own pattern of stimulation, and can access its own memory and stored programs of behavior. A Turing machine, in conception, is an automaton to which is attached a potentially infinite tape upon which are inscribed a certain number of symbols (including the null symbol). These are read

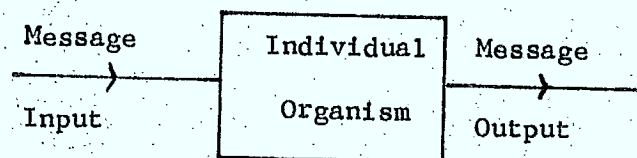
by the machine , and the machine in its turn subsequently inscribes symbols on the tape (if necessary erasing those which already appear there). In some respects the organization of a modern computer is simply a realization of the principles of the Turing machine. The computer reads in data, accesses information stored in memory, performs (one at a time) instructions which are part of a predetermined program, calls up subroutines, and writes output data. The resemblance of these operations to those performed by humans is suggestive: indeed Turing himself proposed that it should be possible to produce a Turing machine which could so successfully "mimic" human behavior that protocols of their behavior would be undistinguishable. In fact Turing proposed as a criterion for the design of such a machine not that it be "really" like a human being, but only that it be capable of reproducing human-like behavior with sufficient similitude to be indistinguishable from the former.

Two characteristics of the Turing machine should be noted since they are related to some of the discussion which follows: first, unlike the network of McCulloch-Pitts neurons, the Turing machine is able to take information only one step at a time, i.e. it proceeds by following a sequence of operations one after the other. This will appear to be equivalent to a "single-channel" capacity as the concept is to be used in the sequel. Secondly, the Turing machine executes "programs", which only in part depend on the state of the environment. It is not totally "stimulus-dependent".

#### Communication and Control

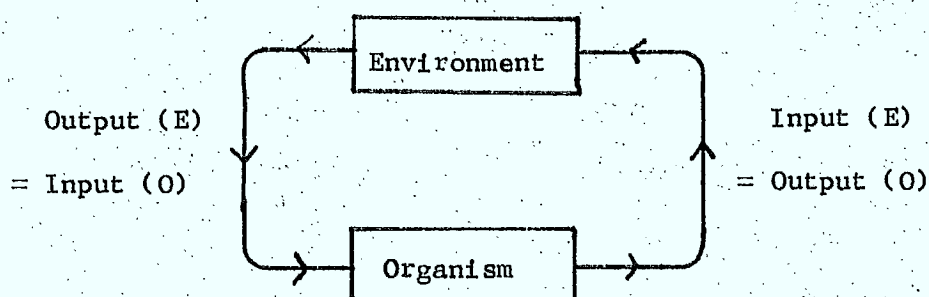
In our discussion of communication in networks, we have freely used the notions of "inputs" to an organism, which might consist of

a single cell or a network of elements, and of outputs. The inputs and outputs considered have been thought of as the reception and transmission of messages, which are a special subset of all the events which might constitute inputs and outputs to an organism. We have been led to think of this process somewhat as represented below:



In this model the output of the organism at time  $t + 1$  would usually be expected to show some influence of the input at time  $t$ . This assumption has been basic to all the discussion to this point.

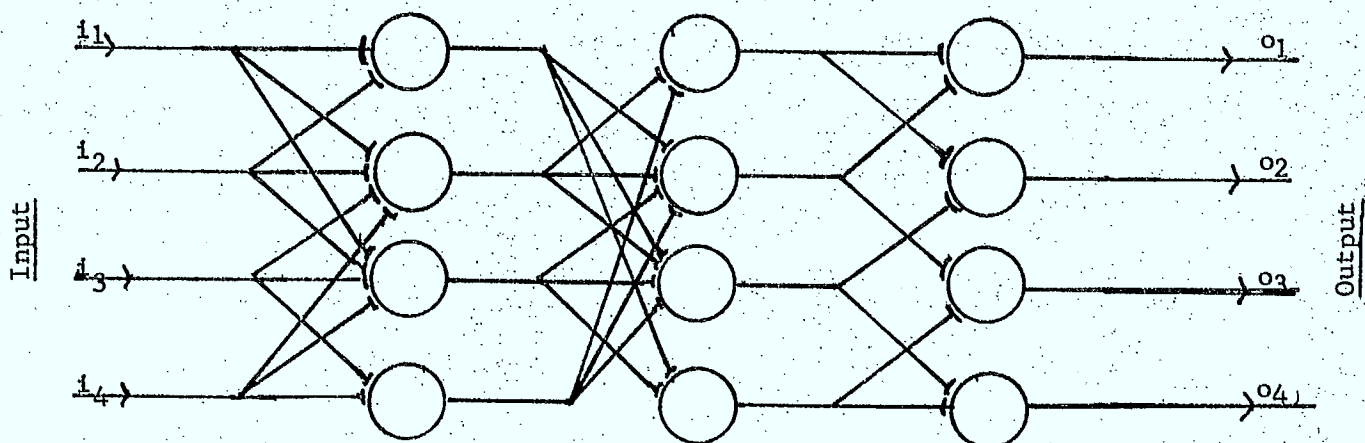
It seems equally reasonable, however, to ask whether the input at time  $t + 2$  is not in turn likely to show an effect due to the output at time  $t + 1$ . If we strike a piano key, we normally expect to hear a sound afterwards. If we say "hello" to someone we expect him to return the greeting. The reason we are able to make this assumption is that we may assume the organism(s) to be coupled to another entity (which we term "environment", (E), in order to leave open the question as to what kind of system is concerned: animate, inanimate, human, non-human, etc.), in such a way that the output of "O", constitutes the input of "E", and vice versa. In this way Organism and Environment constitute a closed loop:



Such a system exhibits "feedback".

Two systems so coupled are now in a position to affect each other. Some of the inputs to "O" are likely to be judged more favorable than others. This may be because some activity of "E" threatens the security (what Ashby calls the essential variables) of "O". It may equally be that keeping "E" in a certain state, as signaled by "E"'s output, figures in a plan of "O". If "O" is a pilot coupled to a plane flying across the Atlantic, keeping the plane in a certain state (flying at a certain altitude, headed in a certain direction) is necessary both for reasons of security and intentionality. His problem is this to maintain guidance, or control, over the behavior of the plane. This he can only hope to achieve by being able to vary his own outputs.

To illustrate, suppose that "E" is a network having the following form (in which every cell has a threshold of "2"):



Note now that if inputs  $i_1$  and  $i_2$  occur, and at the same time inputs  $i_3$  and  $i_4$  do not, then outputs  $o_3$  and  $o_4$  will fire, while outputs  $o_1$  and  $o_2$  will not. Since any automator, as we have seen, can be represented by a modular net, it only requires sufficient access to the inputs of the automator to assure that its outputs are maintained within acceptable limits (always assuming there is a subset of acceptable outputs available to the



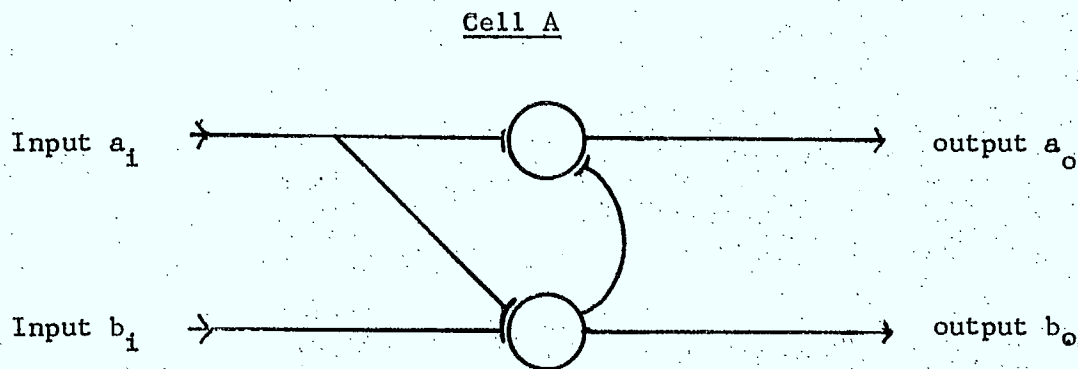
automaton). We may have to perform a number of experiments to see which inputs to it will produce the desired outputs. What this means is that eventually we should be able to learn to fly a plane by ourselves, providing that a) it will fly and b) we don't crash while learning to fly.

In terms of the earlier discussion, we may think of the actual outputs of "E" as drawn from a repertoire of available outputs. The size of the repertoire is related to the complexity of the environment "E"; the potential information value of the messages transmitted by "E" (i.e. its outputs considered now in the perspective of information theory) is thus in part also a function of environmental complexity. The information value of communications from "E" is also, following the argument developed earlier, a function not only of the size of repertoire but also so the probability of occurrence of each to the outputs. However since "O" and "E" are two elements in a coupled system, the outputs of "E" are not independent of the behavior of "O": hence, each output of "E" can be expressed as a probability conditional on the outputs of "O".<sup>1</sup>

One difficulty which arises is that the environment may display memory, i.e., it may reflect in its output not only the most recent behavior of "O" but also earlier activity. Furthermore, if the environment is organized in such a way that it has return, or feedback, loops in its internal organization, the actual functional relation between input and output may become very difficult to comprehend. A very simple example is shown in the following net:

---

1 This is the principle of the "black box", i.e., the method of understanding systems by examining their behavior under varying conditions of external stimulation without regard to an analysis of their internal working parts.



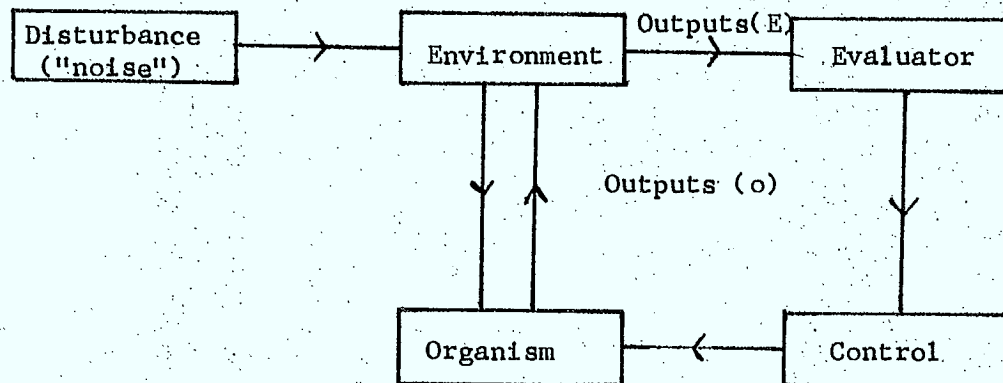
Here, if inputs  $a_i$  and  $b_i$  occur simultaneously, Cell A with threshold "2" will not fire, but Cell B, also with threshold "2" will, and output  $b_o$  alone will occur. If, in the next time period, the same inputs,  $a_i$  and  $b_i$ , are repeated, the output of the net will now be different: the output of Cell B, as a result of the first input, is sufficient, when combined with input  $a_i$ , to cause cell A to fire, and hence this time output  $a_o$  will occur as well as  $b_o$ .

A second problem is that "E" may exhibit learning. Over time its organization may change, and in turn the contingent probabilities of environmental output, for given output of "O", may alter.

Finally, we must take into account the possibility of "noisy" communication from "E" to "O". In part what happens in the environment is due to external disturbances, external in the sense that they are due to unexplained factors, which nevertheless may act on the coupled organism-environment system. Such external disturbances result in increases in the information content of messages received from "E".

We are now in a position to state the principal of a cybernetic control mechanism. Since "O" has determined goals, and since the outputs of "E" can be evaluated against the criterion of whether or not they are

consistent with these goals, we may schematise the complete system as follows:



This is a model of a control system based on the principle of error-control.<sup>1</sup> It operates to reduce the difference between undesired products of environmental activity, and those states of the environment which are acceptable, according to its own set of goals. The system may then exhibit what appears to be goal-directed behavior.

The extent to which regulation is possible, even in principle, is determined by what Ashby terms the "Law of Requisite Variety". (Ashby, 1956) Successful control of the environment implies the ability to reduce the variability of the environmental output to within some set of acceptable limits. This output variability can only be reduced if the organism has available an appropriate repertoire of outputs, sufficient in fact to assure that for whatever change in the environment, there is an adequate response. The variety in "O"'s outputs must equal that of "E"'s. Output information must equal, or be greater, than input information.

---

1. I.e., a "servo-mechanism".

### Plan of succeeding chapters

-----

In chapter 2, we turn to see how some of the ideas just outlined have begun to figure in the design and interpretation of experiments. We present an hypothesis of information overload, based on the Shannon model, and look at the experimental findings which have followed out of the use of the model. We then examine a second framework of explanation which incorporates other principles discussed in this chapter, and begin to look into the nature of the organization of information-related processes within the human (always at the level of the individual). Finally we turn to the question of practical implications of the experimental evidence, and begin the exploration of a framework for the analysis and design of real-world systems.

In chapter 3, the essential insights of Chapter 2 are expanded and extended to a new area of investigation, the social group. A communication theory of information transmission emerges: two types of information are specified: referential and relational. The link between communication and social relationship is developed. Finally, the question of network overload is discussed.

## CHAPTER II

### INFORMATION OVERLOAD: HYPOTHESES AND FINDINGS

### Organization of the chapter

The objective of this chapter is to consider the psychological bases for a theory of information overload at the level of the individual.

The intuitive idea of information overload is straight-forward: the processing of information is work, and as for other kinds of work activity, highly stressful conditions (pressure of time, fear, etc.) tend to affect the capability of the organism to carry out tasks, to a point indeed where he becomes confused and performance begins to deteriorate rapidly. When deterioration sets in, we refer to the condition in which the organism finds himself as "overloaded"; information overload occurs in communication systems and involves tasks which have an information processing basis.

The first part of the chapter is given over to an examination of one particular theory of overload - its theoretical bases, the evidence which has accumulated in favor of it, and some of the criticisms which have been directed at it. There are two reasons for choosing this particular starting point: a) first, because it has been, historically, around this particular theoretical and research tradition that most current expressions of concern about the danger of a surfeit of information in our society have centered (such as those alluded to in Chapter I), and hence it seems prudent to have clearly in mind on what foundation these recent statements are based, and b) secondly, because the theory in question makes relatively simple assumptions, and thus, from a presentational point of view seems a convenient point of entry to the more general discussion to follow.



At the conclusion of this introductory section, we will consider arguments to the effect that the original information-overload theory is inadequate to account for all the known evidence. We will attempt to show the theory fails on two counts: a) empirical - in that it fails to predict correctly certain phenomena, and b) heuristic - in that it limits excessively the type of experimental situation which can be investigated with its aid.

We will then, accordingly, offer an alternative explanation for the findings which constitute the major support for the information overload hypothesis. In so doing we will be forced to discard the concept of human channel capacity, as the term has generally been employed, without however abandoning the idea of a limited central capacity to which the concept of a "single channel" has been related. The explanation which is tendered attempts to avoid some of ambiguities and contradictions which have plagued some of the existing treatments of the subject.

The discussion of a limited-capacity central processing mechanism will lead in turn into a consideration of the organization of human behavior, and how it is affected by varying levels of information availability. At this point we will propose a concept of "program overload". We will then look in turn at theories of the reception of information, internal transformation processes, and the execution of responses.

We will, finally, in summarizing the contribution of communication theory to the understanding of why and how "information overload" occurs point out some of the areas where additional work is indicated.

## Reaction time experiments and the concept of channel capacity

---

Psychologists were attracted in the early 1950's to the Shannon-Wiener model of a communication channel primarily, it seems, because this new tool appeared to shed light on some old problems in psychology. In 1885, Merkel had found that the time taken by a subject to make a correct response to a single stimulus when a number of alternative stimulus-response patterns were available simultaneously was affected by the number of alternatives presented. That is to say, suppose the subject to be required to push a button under his index finger whenever a green light appears: how long it takes him to complete the response (i.e., push the button) has to do with how many other lights might have lit up, and how many other buttons he might have had to push. Commenting on Merkel's experiments, Blank (1934) and Woodworth (1938) observed further that the reaction times obtained in fact varied as a function of the logarithm of the number of alternatives. The coincidence of this finding with the measure developed by Shannon was immediately apparent, and in his 1951 book, Language and Communication, George A. Miller reviewed Merkel's experiments, making explicit the connection between reaction time and the information content of the stimulus display. Hick (1952), and subsequently Crossman (1953) and Hyman (1953), also beginning from a re-examination of the Merkel findings, then went on to conduct further experiments and to establish the foundation for a more comprehensive theory of the rôle of information in determining subject output. We turn to look at their experiments in the next section.

Before we can properly understand the original concept of information overload, we should first be clear about how the communication

channel model has been used in these choice reaction time experiments. Two main contributions of information theory are evident: a) the use of an information metric to measure features of the stimulus display, and b) the use of the Shannon paradigm of an ideal communication system to model the experimental situation.

a) Because of its great generality (see Chapter I), information theory provides an excellent instrument for describing stimulus-responses repertoires. By varying the rate of presentation of stimuli, the size of the stimulus ensemble, the probability of occurrence of the various stimuli, and contingent probabilities from one stimulus event to the next, the amount of information presented can be varied systematically, allowing for easy comparison across a wide variety of experimental situations.

b) In order to understand the utility of the Shannon paradigm as an interpretive tool, we require some familiarity with the notion of a "reaction time experiment".

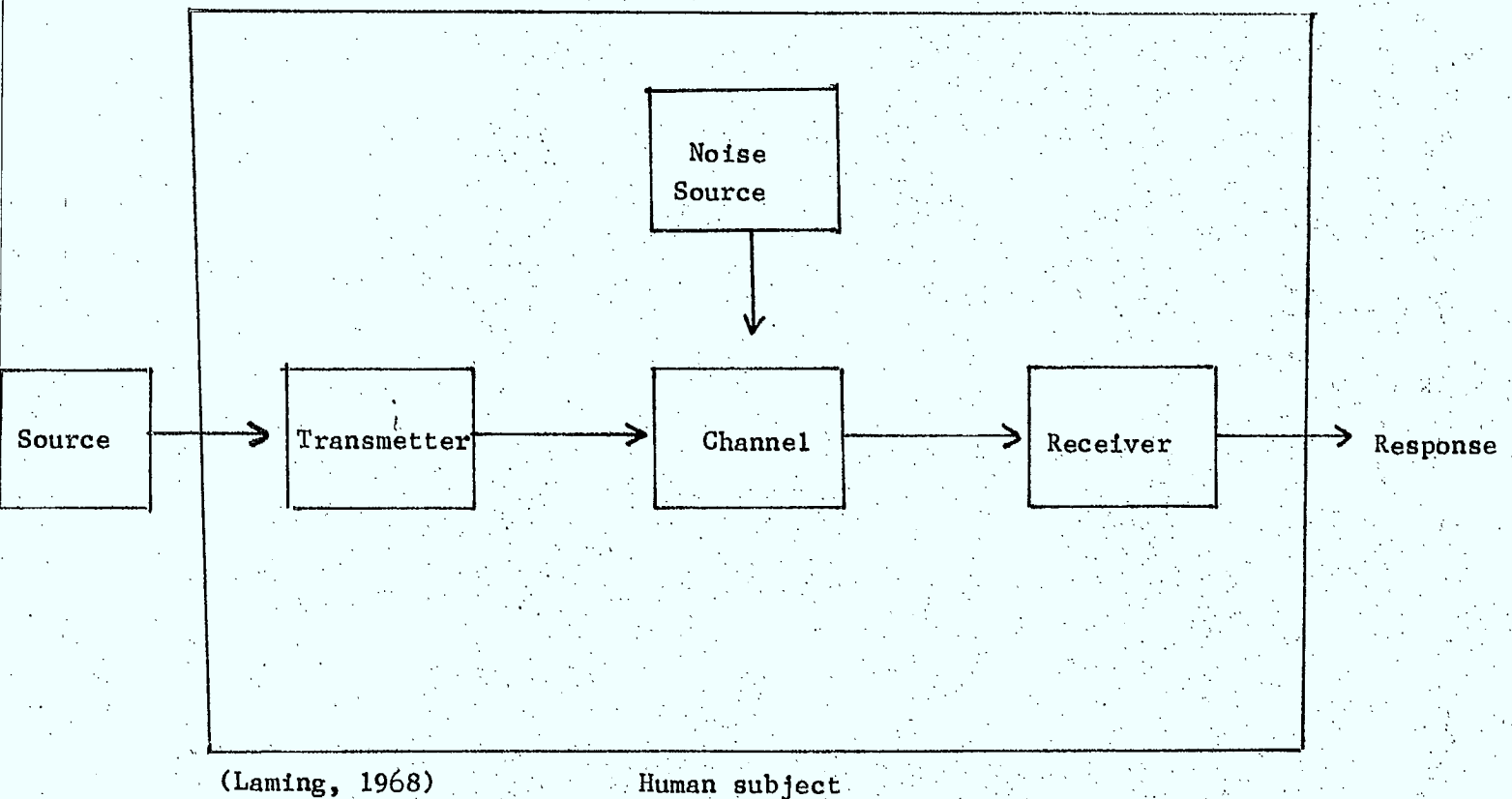
Reaction time (RT) is the lag between presentation of stimulus and the subject's response. Donders (1868) described three types of reactions, a-, b-, and c- reactions (and by derivation, three types of reaction-time experiment). In an a-reaction the subject makes a single response to a single stimulus. This is also referred to as a "simple" reaction. In a b-reaction there are at least 2 stimuli and an equal number of responses, and the subject makes a single response for each stimulus on the basis of a 1-1 mapping between stimuli and responses.

Since the latency of response of the subject in this situation is termed "choice reaction time" the b-reaction experiment provides the basic model for a "choice reaction time" (CRT) experiment. In a c-reaction there are two stimuli, and a single response which is to be produced only when a particular one of the two stimuli is present.

In a simple reaction time experiment, the only uncertainty concerns the time of appearance of the stimulus: the identity of the stimulus and of the response is known in advance. In a c-reaction experiment there is both uncertainty as to when the stimulus will appear, and which stimulus it will be. The nature of the response is pre-determined. In a choice reaction time experiment, time of stimulus occurrence may be either known in advance or left uncertain. In addition, there is both stimulus uncertainty and response uncertainty, and indeed the two tend to be confounded. In general, in the literature concerning information overload, it is the choice reaction paradigm which has generally been employed.

Now let us look at how the choice reaction time experiment has been interpreted with the aid of the Shannon model of an ideal communication system. One such interpretation due to Laming (1968) is shown in the following Figure.

FIGURE II - 1



An interpretation of the choice-reaction experiment in terms of the Shannon model of an ideal communication system.

In the CRT experiment, the sequence of stimulus signals can be seen as equivalent to the message output of the source; the responses of the subject are then interpreted as the message received after being transmitted through the human channel. The capacity of the "channel" is the number of bits which can be transmitted per second, which depends on the number of accurate responses made by the subject. The informational content of the messages emitted by the source is a joint function of the rate of emission, the size of ensemble from which the items were drawn, and of the probability of appearance of each.

Laming's interpretation is not the only one which has been used. Hyman (1953) provides a slightly different explanation.

"The choice reaction-time experiment can be looked upon as a model of a communication system. The display represents a transmitter of information. Each alternative stimulus or signal represents a message; more information can be transmitted the greater the number of messages from which one can be chosen. The channel over which the signal is transmitted can be considered as the air space between the light and S, and might also include part of S's visual afferent system. The S acts as a receiver or decoder in that at some point he decodes the signal into its message and reacts with the appropriate response (the destination of the information)".

The two interpretations differ in terms of how much of the transmission process is assigned to intra-individual phenomena: for Laming the presentation of the stimulus display constitutes a signal and everything which follows, including the encoding of the source message in the form of a signal (recoded in a form suitable for transmission through the human "channel"), occurs within the boundary of the organism; Hyman would limit that which takes place within the individual to part of the channel, and to reception, or decoding, of messages. Hyman thus assigns to non-organismic processes the encoding of messages.

Both Laming and Hyman interpret the subjects' responses as corresponding to that component of Shannon's model termed the "destination" of the information. The effect of this interpretation is to telescope into one event two separate phases of behavior: a) the reception of a message, and b) the execution of a response. This introduces the implicit assumption



that any novelty in the responses of the subject is due to events which intervene before or at the moment of reception of the message, due perhaps to characteristics of the encoding process, or effects of channel performance (which constitute in the Shannon model "noise", but which we might equally well interpret as more or less systematic disturbances which are associated with the peculiar characteristics of the human system as a channel or decoder).

One important consequence follows: the use and interpretation of Shannon's model which we have just discussed is appropriate only in those situations where the person's output (his behavior) is a direct linear function of the messages he receives. In this instance the individual is a simple transducer, that is, he is merely a relay in a communication network. As it happens, the conditions of the reaction time experiment are such as to make plausible such an interpretation; to what extent we can generalize to other situations will have to be seen.

The interpretation assigned by Hyman, Laming, and others in the same tradition, to the Shannon model is clearly not the same as that offered by Shannon himself (see Chapter I). Shannon had in mind the transmission of information between two nodes in a network; we have just been considering, on the other hand, the case of transmission of information through a node. The mere existence of such a difference is not something which in itself ought to disturb us: as we noted in the preceeding chapter, mathematical models, like that of Shannon, begin by being purely axiomatic, and can often, legitimately, be applied in several different ways to illuminate real-world phenomena. The criterion is a simple one of efficiency: each time we are led to ask what extra insight has been gained by the use of a particular model, whether the explanations are consistent with all the

available evidence, whether the model is productive in the sense of generating hypotheses, and on. We may also want to enquire whether in the choice of model we have so reduced our field of inquiry that we have begun to leave out of account many phenomena which, for other reasons, we are interested in studying.

In the following sections we shall accordingly look at some of the applications of the communication model to choice-reaction time experiments, and we will then go on to ask whether, on the whole, the model is efficient.

The information-theoretic interpretation of choice-reaction  
 -----  
 time experiments: early experiments  
 -----

In Hick's original experiments, ten pea lamps were arranged in a somewhat irregular circle and the subject was provided with ten corresponding Morse keys on which his fingers rested. Hick measured reaction times to several series of signal presentation (100 to 200 stimuli in length), varying the ensemble size from two to eight. Frequencies of each signal in a series were approximately equal and first-order autocorrelations were eliminated. His data confirmed the assumption of an underlying logarithmic relationship between choice reaction time and the number of alternatives available, when corrected to take into account the temporal uncertainty experienced by the subject as to when the response would be required.

The equation which best fit his data was

$$(1) \quad \overline{RT}(n) = b \log(n + 1)$$

where  $RT(n)$  is interpreted as the average reaction time to one of  $n$  equally probable stimuli,  $b$  is a constant, and  $(n + 1)$  includes a factor to account for temporal uncertainty as to when the signal will appear, which Hick took to be equal to the uncertainty due to an increase of one in the size of

(1)  
the stimulus ensemble. From his results, Hick estimated that subjects reach an average rate of transmission of from 5.5 to 6.0 bits per second.

Hick's motivation for the use of a logarithmic (rather than of, for example, a simple linear) function in his equation was as follows: there is, for each subject, following the presentation of the stimulus, a certain time which is required for him to identify (recognize) the signal. (2)  
There are several ways in which such a process of recognition could occur. Hick proposed as a likely candidate, in view of the fact that it would produce the type of results which can be fitted with equation (1), a "progressive classification" procedure. The method is similar to the coding technique of "binary fission" discussed in Chapter 1: in the first test, the subject places the stimulus in one of two equally probable classes. Depending on the result of the first test, a second cut is made which in turn subdivides the appropriate half of the original possibilities into two equiprobable halves. The process is repeated until the subject "zeroes in" on the correct identification of the stimulus (somewhat in the same fashion as one does in a game of "Twenty Questions"). Obviously, with this procedure, each time the number of elements in the ensemble of available stimuli

---

(1). In his original paper Hick used logarithms to the base "10": in this discussion we will continue to assume the base to be "2", as is the more general current practice.

(2). Hick's assumptions are discussed in greater detail later in the chapter.

is doubled, one dichotomizing step is also added. If it is assumed that each dichotomizing-scanning phase consumes an equal time, than the desired logarithmic relationship appears.

Crossman (1953) followed up these experiments of Hick by testing for the case of stimuli which have unequal probabilities of occurrence. Crossman assigned his subjects the task of sorting a deck of ordinary playing cards into various classes, eg., by suit, by number, by color, etc. His method had the advantage that it permitted an evaluation of the effect of varying the a priori classes into which stimuli must be assigned. His findings supported the earlier result that the response lag is in fact proportional to the uncertainty in the signal source, and that this relationship continues to hold up when signals are not equally probable. He discovered further, as we should expect if recognition were the mechanism primarily responsible for differences in reaction times, that the discriminability of signals was able to affect the outcomes significantly.

Crossman also introduced explicitly for the first time the concept of channel capacity, and hence the theoretical possibility of an upper limit constraint on the amount of information which subjects can transmit.

Hyman (1953) used as the stimulus in his experiment a visual display consisting of a matrix of small lights, with one to eight positions in which the light might be expected to appear. Subjects responded by uttering a matching nonsense syllable corresponding to each light position. Hyman varied not only the number of signals with probability held equal and the frequencies of signal occurrence (the number of alternatives remaining the same), but also first order sequential dependencies (while holding

constant the relative frequency of signal occurrence). He found the mean RT-to-information-content function to hold in all conditions.<sup>(1)</sup>

By 1955, Bricker found it possible to review the available choice reaction time studies and conclude that the evidence was sufficient to state with some confidence that average reaction time could be plotted as a function of the information in the input:

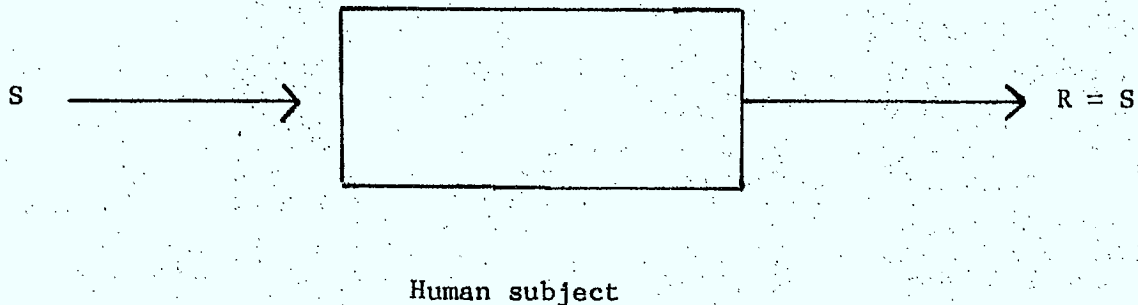
$$(2) \quad \overline{CRT} = a + b H$$

where  $\overline{CRT}$  is the average reaction time,  $H$  is the entropy of the stimulus ensemble, the intercept  $a$  is a factor to account for simple reaction time and the slope  $b$  reflects the time required to react to a stimulus in a two-choice situation.<sup>(2)</sup> The notion of simple reaction time is that of a threshold latency, determined, in part at least, by the time required to marshal any output activity. The interpretation of the slope  $b$  as reflecting an internal classification process is similar to that of Hick.

The experiments of Hick, Crossman and Hyman demonstrated empirically a relationship between latency of response in a choice reaction time experiment and the information-content of the stimulus display. They also introduced the principle that the human subject could be viewed as a channel through which information (presented in the form of stimulus signals) is transmitted (as the subject's responses). Hyman (1953) noted the underlying assumption: "These hypotheses assume that (a) the res-

- 
- (1) Hyman's results in certain respects raise difficulties with respect to a communication model interpretation which will be discussed in a later section.
- (2)  $H$  could more properly be interpreted as information transmitted,  $H_t$ , but in the case of a 1-1 S-R mapping, and low error values information transmitted approximates closely stimulus entropy,  $H_s$ .

ponses are completely determined by the stimulus series... (this) assumption demands a one-to-one correspondance between stimulus and response series". This assumption can be represented as follows (where effects due to noise, or errors, are disregarded).



The general resemblance of this model of the human subject to that of a McCulloch-Pitts neuron may be noted: in particular, responses occur (the subject "fires") only when certain threshold stimulus values are present; no other transformation of the information is assumed, and in particular no memory or learning is assumed (in fact, in the experiments cited, learning effects were treated as a source of contamination). The isomorphism between the model of the human subject and a neuron has encouraged, as we shall now see, an attempt to state a general hypothesis of information overload.

#### The Hypothesis of Information Overload

In none of the experiments mentioned above was a direct attempt made to measure channel capacity as such: in most cases, indeed, the information content was deliberately maintained at a level below that



which would result in any appreciable number of errors occurring.

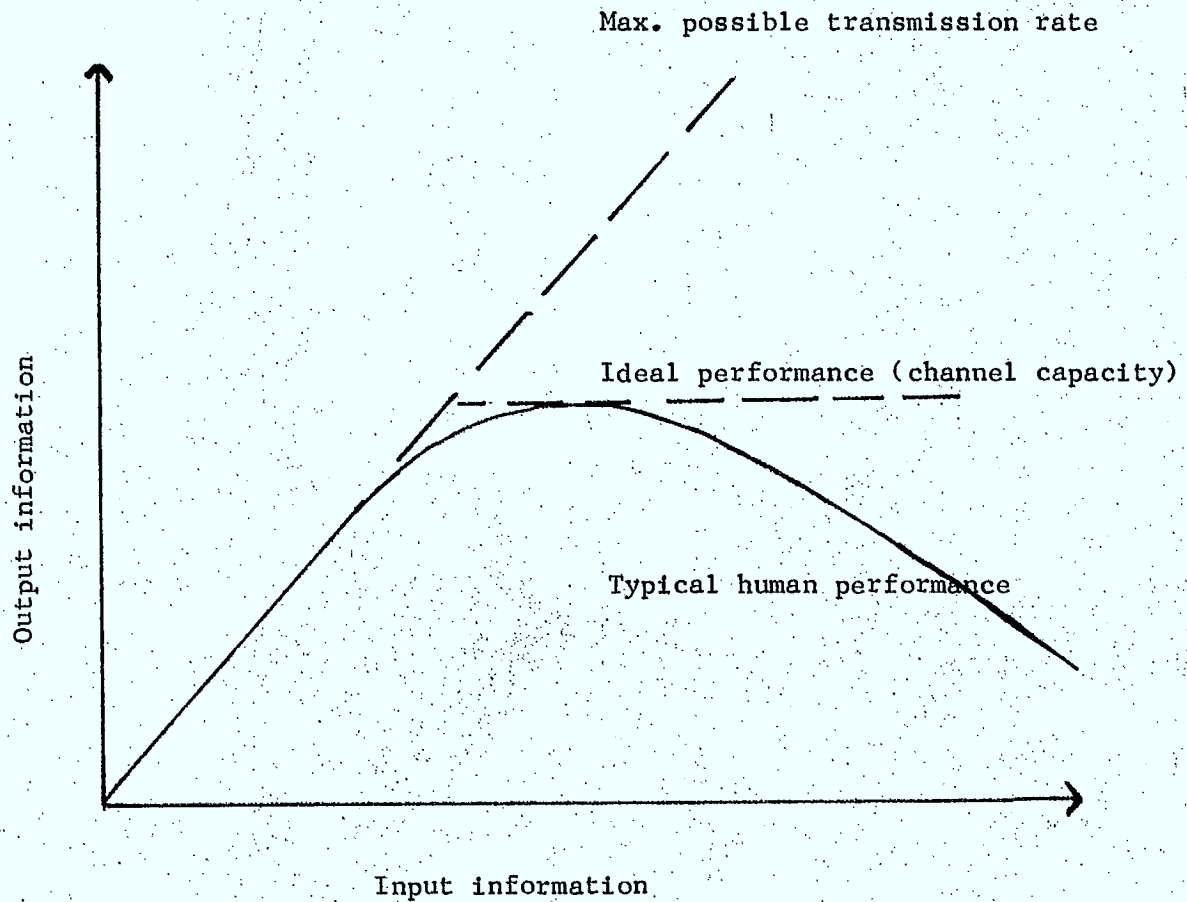
However, Alluisi, Muller and Fitts (1957) using as a stimulus numerals projected on a screen (varying the number of alternative numerals in the experimental ensemble and the rate of presentation) found that as the rate of presentation of the stimuli was increased, the subjects began to make increasing numbers of errors, and in fact that although the rate of information transmission first leveled off, it then began to decline.

How should we account for the decrement of performance after channel capacity has been, presumably, reached and then exceeded? In a series of articles in the early 1960's, James G. Miller (1960, 1962, 1963a, 1963b, 1964a, 1964b) set out to develop a general theory of information overload: (a) he stated a hypothesis that under certain circumstances the subject would perform at sub-optimal information-transmission capacities when compared with an ideal channel, and (b) he further attempted to show that the results obtained in choice reaction experiments constituted merely one example of a more general principle affecting organisms as simple as the single cell, and as complex as complete societies. He went on to identify some of the typical symptoms associated with communication situations where the organism becomes overloaded, ranging from elementary strategies of adaptation to full escape.

Miller's hypothesis of information overload can be stated as follows: when input information in bits per second is increased, the output at first follows the input more or less as a linear function,

then levels off at channel capacity and finally falls off toward zero.  
(Figure II-2)

The channel capacity of systems, Miller argued, should differ according to the level of complexity of structure of the organism's information-processing system.

FIGURE II - 2The Hypothesis of Information Overload:

As information input increases, output first reflects the increases information content of the input until an asymptote is reached, beyond which point information transmitted declines.

In developing Miller's argument for his hypothesis of information overload, we shall proceed in two steps: a) first we will look at some of the experimental evidence which has been adduced in support of the hypothesis; and b) we will look more closely at the theoretical foundations of Miller's position, which in turn will lead into a discussion of the general merits of the communication model of Shannon as a theoretical underpinning for the interpretation of choice-reaction time experiments. The advantage of proceeding in this fashion is that it corresponds in general to the two levels of Miller's reasoning which first attempts to show that certain very general laws of behavior apply across systems of very different levels of complexity of organization and secondly attempts to situate human behavior within the general framework of explanation we have been discussing in this chapter. There is the further advantage that in reporting the experimental evidence for organisms at differing levels of organisation the essential points of Miller's argument become quite clear; this, thus, leads quite naturally into an examination of the theoretical foundations of his position.

#### Overload at the level of the cell

-----

Much of the persuasiveness of Miller's argument derives from his assumption that the same channel capacity model which applies to individual cells of the nervous system is equally pertinent for more complex organisms. Miller has been able to accumulate an impressive volume of evidence to show parallelisms in response to varying information conditions between organisms at very different levels of organization, from the simple cell to large social organizations. Much the most convincing part of this evidence however refers to behavior of the cell.

For this reason, as well as to get a thorough grasp of how the hypothesis has been tested, it will be useful to look at some of the evidence which has been cited by Miller to substantiate his claim that the hypothesis is universally applicable.

In order to estimate the channel capacity of neurons, a preliminary difficulty must be cleared up. The simplest living system which is capable of processing information is the cell (and indeed the higher level information macro-processing activities of more complex organisms depend entirely on this cellular capability). The cells of the nervous system, or neurons, respond to several types of external stimulation such as light, sound, touch, etc... by emitting a sequence of output pulses or "spikes". Their behavior under different conditions of external stimulation can be investigated by varying the intensity or frequency of the stimulus. The number of pulses emitted per second is, however, not necessarily the same as the number of bits, since the code which the cell is using is not available to us by inspection. Suppose we determine that the stimulus will be varied along one dimension (eg., "loudness") on which we distinguish eight values. For us the stimulus ensemble has a maximum of three bits of uncertainty. If the cell discriminates between only four different values of the stimulus dimension, however, for it the maximum information contents of the ensemble is four bits of information. This problem is resolved by Miller by assuming that the number of bits transmitted is at least proportional to the number of pulses emitted, i.e., that input information can be estimated by output information. On this assumption, the validity of the overload hypothesis can be evaluated at the level of the cell, even though the actual channel capacity can be at best estimated approximately.

Given these constraints, there is extensive evidence to support the overload hypothesis at the level of the cell (1). Some examples will be sufficient to indicate the general nature of the experimental results on which support for the overload hypothesis is based.

Brock, Coombs, and Eccles (1953) found that when antidromic electrical pulses were input to the motor neuron of a cat at low frequencies (13, 20 and 28 pulses per second) there was a corresponding somadendritic spike output rate. When the input pulse rate was increased to 42 pulses per second, pulses occurred only at every second input. At 61 impulses per second an output occurred with every fourth input; and so on, with the output pulse rates falling gradually from a recorded maximum of 28 pulses per second. The theoretical curve of Figure II-2 thus gives a good general fit to these data.

Some adjustment processes can be at least indirectly inferred from research conducted by Granit and Phillips (1956), who found in their work with cells in the cerebellum that when the interval between input pulses was less than 3 milliseconds, every second input elicited an output pulse only 40% of the time, and when the inter-stimulus interval was less than 2.2 milliseconds, the intensity of the second output pulse also diminished. Other research appears to indicate similar alternation of strong and weak impulses, prior to more pronounced declines in the transmission rate (Wall, Lettvin, McCulloch and Pitts, 1956).

It has also been found that while the neuron appears to be able

---

(1) The evidence will be reviewed extensively in Miller's forthcoming book Living Systems, Chapter 5. Dr. Miller was kind enough to provide us with an advance copy of his text. We have greatly benefited from his generosity.



to discriminate between differences in the intensity of stimulation, and to indicate these differences by changes in the interpulse frequencies of its output sequences, there are limits to the resolving power of its discriminatory mechanisms. Mountcastle (1966) obtained results which show that the output appears to be able to reflect up to between 4 and 5 categories of input, but that beyond this limit, the cell does not respond separately to further differences in intensity. (It will be seen later that this limitation in discriminatory power has an analogue at the level of the human organism, and has provided an explanation for the phenomenon of the "span of absolute judgement", in the terms of George A. Miller).

Additional evidence for the assumption that the number of spikes per unit time is a coding mechanism which represents differences of intensity in the stimulus has been noted by De Valois (1958). More importantly, there is very good evidence to indicate that the information so encoded by receptor neurons is preserved intact at higher levels of the nervous system (Jung and Baumgartner, 1955 ; Tasaki and Davis, 1955). In such systems with numerous cells synaptically linked, the refractory period is however longer than for the individual cell (Mountcastle, Davies and Berman, 1957).

From these results the following conclusion can be drawn: the neuron is an effective information-transmitting channel within specifiable limits. Its ability to encode and transmit information with high fidelity is limited to stimulus ensembles containing not more than 2-3 bits of information and to rates of presentation of about 30 signals per second. When rates of presentation are increased to the point where the cell has insufficient time to recover its discharge

potential before a new stimulus is input, it exhibits signs of disorganization which increase until breakdown occurs and information transmission falls close to zero.

Hence the efficient functioning of the cell is particularly  
(1)  
sensitive to variations in the rate of arrival of new information.

---

(1) The explanation for the observed decrements in performance of the cell at the higher levels of stimulus information is as follows: in their normal resting state the membranes of neural cells are electrically polarized, the outside being positive (because of the presence of free sodium and potassium ions). The potential of the cell is positive. As the nerve impulse passes along the nerve fiber, sodium ions flow into the membrane, thus briefly altering the electrical balance. After the impulse passes, the original balance is restored, but the restoration period (during which the cell is refractory to further inputs) seems to take a minimum of about 30 milliseconds.

Direct evidence of overload at the level of the individual organism.

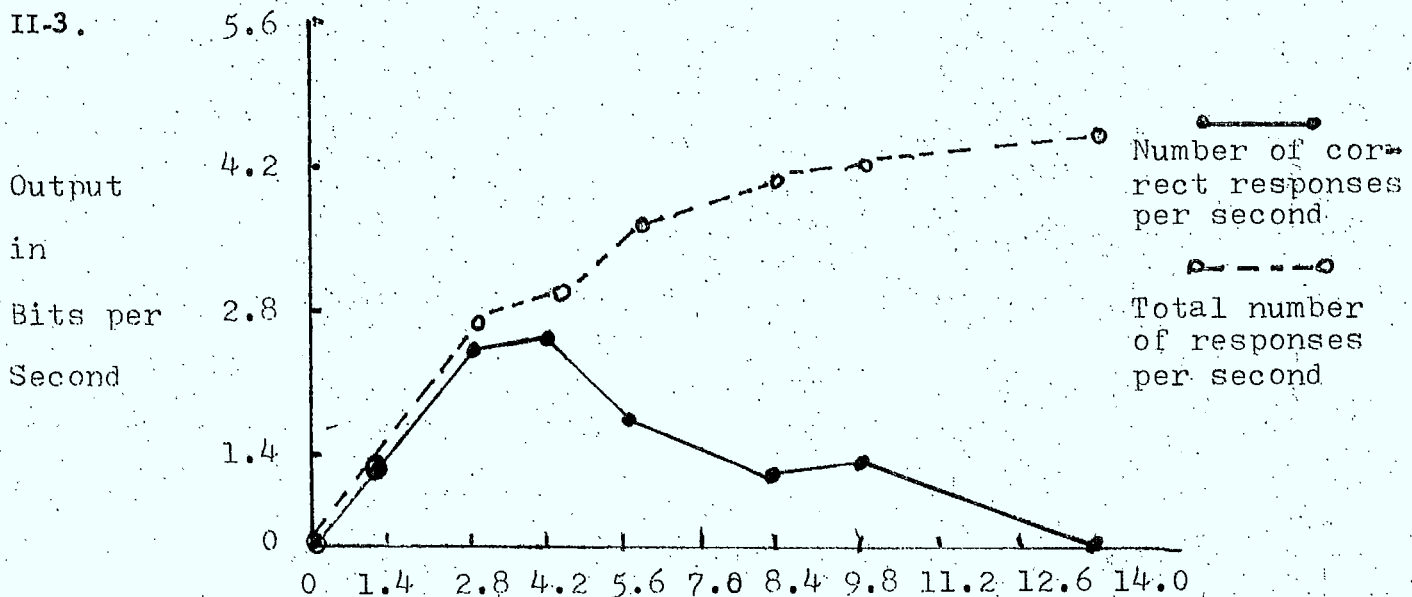
In this section we shall examine some evidence which shows that the same hypothesized decrements of performance observed at the level of the cell also appear, with certain restrictions, at the level of the organism.

The type of choice reaction time experiments reported earlier in this chapter turn out to be unsatisfactory as a source of direct demonstration of the overload phenomenon. The reason is simple: CRT experiments have not in general exemplified high levels of input information, certainly not of an order sufficient to lead to breakdown. Rate of presentation of signals is generally moderate (sometimes left to the subject in the form of self-pacing tasks). For the most part ensembles have been restricted to 32 elements and usually less (or a maximum of 5 bits input information). When this limit has been exceeded, as in Hilgendorf (1966), where the ensemble was increased to a maximum of 1000 (9.96 bits per stimulus), the error rate nevertheless remained very low (less than 1%) and, as we should suspect, there was no indication of breakdown. His subjects maintained, according to Hilgendorf's estimates, a roughly constant level of transmission of information of about 5.5 bits per second, for all levels of stimulus uncertainty. CRT experiments are in fact structured to keep error rates low (not more than 10% in any event), so that overload is, as it were, "designed out" of the experiment.

A better source of evidence are "tracking" experiments, in which the subject has to adjust his behavior to match features of long sequences of stimulus presentation and response. In this situation subjects can be pushed to the limit of their information-transmission capacity and beyond. We will begin accordingly by looking at some of the results which have been obtained in these experiments.

### Tracking experiments.

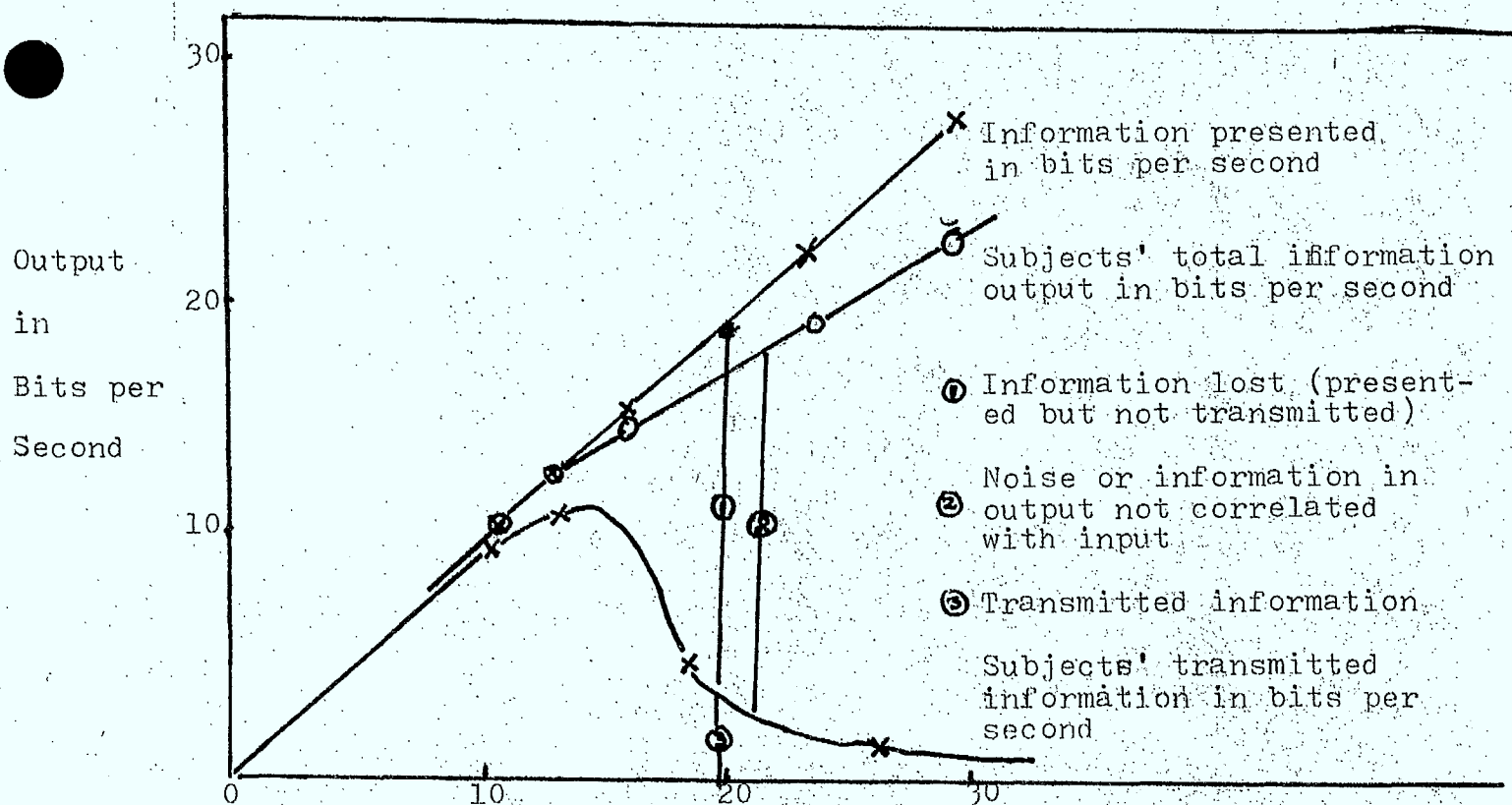
Vince (1949) reports a series of experiments in which input information was presented by passing a moving horizontal band of tape before the subject at various rates of speed. The information was coded in the form of clumps of dots, arranged horizontally, vertically, or in groups, and of numerals printed on the tape. Depending on the number of dots, or the value of the numeral, the subject was required to produce an equal number of taps on a telegraph key. The dots were spaced either equally or unevenly on the tape at intervals which resulted in the following rates of presentation: 1.0, 0.5, 0.3, 0.25, 0.17, 0.14, or 0.1 second interval between dots. The presence or absence of a dot was counted as one bit of information. The results are shown in Figure



At slower input rates, subjects are able to track successfully. As the rate of input of information is increased, the number of total responses continues to increase, but the number of correct responses first peaks and then falls, eventually to almost zero, a level at which chance alone would be sufficient to explain the results. Thus in general the results conform to the predicted pattern. At the faster rates, (i.e. 7 bits and 100 bits per second) the subject has to tap out individual responses very rapidly. The absolute upper rate of tapping, without regard to input, is between 5.5 and 8.0 taps per second. Hence at the faster input-output rates, the subject encounters simple physiological constraints, i.e. the capacity to organize effector activities efficiently.

Wagner, Fitts, and Noble (1954) used an experimental design similar to that of Vince, with the difference that additional channels or display windows on which input information could be displayed were employed (up to 3 channels). Subjects were asked to press a key when a moving dot came abreast of a fixed cross line. In general the results of this experiment showed also that errors increased and information transmitted declined as input rates increased past an asymptotic value. Klemmer and Muller (1953) employed a slightly different design using from one to five lights arranged in an arc as the input device. Subject was provided with five keys arranged in a similar pattern so that he could press one equivalent key for each light. The ensemble was varied from two to five lights per second, in sequences of 100 presentations. The findings are shown in Figure II-4.





The same general finding as that obtained by Vince, and by Wagner, Fitts, and Noble appears, with the difference that subjects are able to attain much higher rates of information transmission, a result which is explained by the greater complexity of the input signal.<sup>(1)</sup>

Conrad (1951, 1955, 1956), Conrad and Hille (1957) and Jackson (1958) conducted further experiments which confirmed that both total input information and the number of input sources affect independently

(1) This explanation raises another issue: since the information as presented by Klemmer and Muller is presumably closer to optimal encoding, and thus to the actual channel capacity of the individual, we might be led to inquire whether maximum rates of transmission in some of the other experiments cited above represent channel capacity rates, or some other constraint. This is a question to which we will return later.



the output rate of subjects. These experimenters found clear evidence of deterioration of output response at the higher rates.

In a series of experiments Quastler (1956), Quastler and Wulff (1956a), Quastler and Wulff (1956b), Quastler and Brabb (1955) studied the question of overload with respect to a variety of types of activity. In one experiment, young students of piano were required to play a score which consisted of a sequence of single notes arranged in random order. Ensemble size (the range of values of the notes) and rate of playing were varied. The results indicated that subjects were able to maintain an error free performance up to about 5 keys per second; from 5 to 10 keys per second there was a tradeoff between speed and accuracy with the consequence that information transmission remained high although the percentage of errors increased; while above 10 keys per second performance deteriorated. The amount of information transmitted was affected positively by the number of keys up to 25 keys after which increases in the range of keys used resulted in lower rates of information transmission. Maximum rates of information transmitted obtained in these experiments were about 23 bits per second. It was not determined whether this rate could have been increased by requiring subjects to play more than one note simultaneously, following normal practice in piano playing.

Similar experiments were run using skilled typists, who were presented with random sequences of equiprobable symbols, drawn from an alphabet of 4, 8, 16 and 32 symbols. Performance was paced by a metronome at 2, 3, 4, or 6 beats per second. With ensembles of 16 and less symbols, and at rates of presentation of slightly more than 3 symbols per second average, the typists made few errors. Beyond these limits,

errors increased proportionately to the increase in speed of typing. At higher levels performance deteriorated. The highest transmission rate achieved was about 16 bits per second. Other experiments using reading and mental arithmetic tasks produced comparable results.

### The theoretical basis of the information overload hypothesis

In this section we present a rationale for the overload hypothesis, couched only in terms of information and elementary network theory. We will avoid the introduction of any assumptions other than those already developed in earlier parts of this chapter. Our objective in this discussion is to present a plausible explanation, at the level of the organism, of why human information-transmission behavior should conform to the hypothetical pattern, as the findings quoted above indicate it does. We hope in the process to acquire the smallest amount of theoretical baggage possible, and reserve to the next section the task of assessing the utility of the model so developed.

Miller's reasoning may be summarized as follows: the rate of information transmission depends, as we have seen in Chapter 1, both on the number of signals per unit time, and on the average information conveyed by each of the signals. There are inherent limits associated with each of these factors, rate of presentation of signals and average information-content of the signals: we consider each in turn.

#### Signal rate of arrival as an information parameter -----

There is, it should be evident, an irreducible minimum time required to process any message whatever its content (simple reaction

time) and to get ready to accept a new message. The "getting ready" to accept a new message implies a period which must be added on to the actual processing time itself. It thus constitutes "dead" time (sometimes referred to as the "refractory period"). When rates of presentation of stimuli are set so high that they result in new stimuli appearing before the previous ones have been processed, an impaired performance will inevitably result. This principle, it has been seen, is clearly evident at the level of the cell.

How do we relate the decrements in neuronal performance to the leveling off and decline of transmission capacity of the organism as a whole? First let us recall that the nervous system of the organism is made up of very long concatenations of individual cells, and that such systems are particularly susceptible to accumulations of error. Wiener (1948) noted the possible analogue to a complicated telephone connection, with numerous switching points and relays:

The more stages which are involved, the more rapidly the service becomes extremely bad when a critical level of failure is exceeded, and extremely good when this critical level of failure is not quite reached. Thus a switching service involving many stages and designed for a certain level of failure shows no obvious signs of failure until the traffic comes up to the edge of the critical point, when it goes completely to pieces, and we have a catastrophic traffic jam.

Man, with the best-developed nervous system of all the animals, with behavior that probably depends on the longest chains of effectively operated neuronic chains, is then likely to perform a complicated type of behavior efficiently very close to the edge of an overload, when he will give away in a serious and catastrophic way... A point will come -- quite suddenly -- when the normal traffic will not have space enough allotted to it, and we shall have a form of mental breakdown, very possibly amounting to insanity.

Figure II-2 does not however indicate a sudden fall off: to explain the gradual character of the decrement in performance, Miller has posited the existence of what he terms adjustment processes, (filtering, omission, abstracting, etc.) which begin to come into play as channel capacity is neared. The analogy with a telephone system is far from satisfactory since, as we saw in Chapter 1 (p. I-44), the nervous system is constructed in such a way that it is able to combat relatively high levels of error (in the form of malfunctions of single neurons). The weapon which is employed is redundancy: continuing with the analogy of a telephone system it is as if every call made was duplicated exactly by thousands of other calls, conveying the same message at the same time. Assuming the operation of a simple statistical principle, then the fall-off in performance would in fact resemble the prediction of the overload hypothesis.

#### Size of ensemble as an information parameter

-----

There are, we posit further, intrinsic limitations on the fineness of discrimination an organism can make and hence constraints on the organism's encoding ability. There are two ways in which messages may be encoded (three if we consider the combination of the other two): (a) an amplitude, or pulse, modulation code, in which the transmission of information depends on whether a signal of a given pattern occurs, and (b) a frequency modulation or pulse interval code in which the transmission of information requires recognition of the length of the interval occurring between successive markers. In its most elementary form, this second code consists of the presence or absence of a signal during a succession

of intervals of equal length. Every coding system, however complex, is essentially an elaboration of one or the other of these methods.

Both types of code are subject to the following constraint: there is in practice an inevitable variability of performance which means that the actual signal emitted is at best an approximation of the ideal stipulated by the code. It is therefore proper to describe the obtained set of signals in terms of a set of parameters: eg., a mean and a variance. If the "jitter" in the transmission of signals is too great (i.e. if the channel is too noisy), or if the signals are too similar to each other for easy recognition, the resulting overlap of variance will inevitably result in numerous cases of misidentification, increasing equivocation (with attendant reduction in rates of information transmission). Hence there must be some upper limit to the number of gradations which can be discriminated along any dimension of the marker, and an according limit to the amount of information per marker which can be accommodated.<sup>(1)</sup>

Since there is an upper limit to the number of absolute judgments or discriminations which an organism can perform with accuracy along any dimension, it can be reasoned that the greater the number of stimuli which are to be recognized by the subject, the greater the likelihood that errors will occur. From this fact, and by making assumptions concerning the probability of inaccurate judgments occurring, a noise or confusion matrix can be computed. (Crossman, 1955; Quastler, 1956; Luce, 1959) One calculation based on a matrix as computed by Luce is shown in Figure II-5 it will be noted that information transmitted plotted as a function of

---

(1). For empirical verification of this principle, see pp.II-53-55.

(1)  
stimulus information follows the predicted pattern.

- 
- (1) For further discussion and criticisms of the utility of the confusion matrix, see Luce (1959, pp. 171-186). For one thing this model fails to predict the well established finding that absolute judgment is not noticeably improved by extending the range of the continuum on which the stimulus to be estimated is placed.



At this point we have completed our development of the argument in favor of the information overload hypothesis.

The proposition advanced as the information overload hypothesis states that as the rate of input of information increases the output does not level off at channel capacity, but rather falls. We might have, if we had wished, contented ourselves by thinking of this as a generalization based upon the examination of a variety of empirical findings at several levels of investigation. However we have been able to go further to show that there are a priori theoretical considerations which help to explain why performance should decline rather than level off at channel capacity, as we should expect if the human subject in fact conformed exactly to the model of a communication channel. We shall now turn to look at some of the evidence leading us to reject, in part, the Miller argument. We will at the same time examine some of the theoretical shortcomings of the line of reasoning which has been advanced in preceding sections of this chapter.

Modelling the human as a communication channel: an assessment.

In this section we turn to an examination of some of the objections which have been raised to the communication channel model. The objections are of two orders: those which have to do with problems of interpretation of the model itself, and those which concern certain empirically-obtained findings which are difficult to reconcile with predictions derived from the theory. At the end of the section an assessment of the utility of the model is offered.

Problems of interpretation  
-----

Laming (1968) has pointed out serious inconsistencies in the application of the Shannon model to choice reaction experiments. Shannon's concept of channel capacity was linked to that of coding. If the entropy of the source is less than or equal to the channel capacity of the transmitting system, then there must be a coding system which will permit transmission of the source message with an arbitrarily small error rate. To establish channel capacity, we would thus have to be assured that the original message had already been recoded in optimal form. But this in turn requires the examination of messages, if necessary, of infinite length, which in turn implies a (possibly infinite) coding delay. Clearly the analogy with most choice reaction experiments does not hold: whatever the reason for breakdown, or the confusional state, it can hardly be argued that the reason is that channel capacity has been exceeded, since the definition of capacity implies a condition of optimal coding, which cannot be shown to hold in the choice reaction experiment. Furthermore, Laming argues that the encoding system is embodied in the performance of the transmitter, which is to say the display system, and this is an invariant in the experiments we have been discussing, with the result that optimization of the transmission rate could not have occurred. Breakdown may be associated with varying levels of information, but cannot be a consequence of exceeding channel capacity. Hence the statement of the hypothesis itself is based on a misunderstanding.

"The simple manner in which Shannon's measure of entropy has usually been applied in psychology has already been criticized by Cronbach (1955). This measure applies only to ideal channels, in which messages are infinitely long, in which an infinite coding delay is acceptable

(though not always necessary), and where complete and accurate knowledge of the probability structure of the signal series is stored in the system. In a choice-reaction experiment the messages are, of necessity, very short. The very design of the experiment requires that each signal must be passed completely through the system, encoded, transmitted and decoded and the response registered, before the next signal is emitted from the source. Reaction time must therefore include not only transmission times but also the time required to encode and decode the message and to execute the message" (Laming, 1968). Support for Laming's view can be found in evidence from at least one experiment (Kirchner, 1958), where it is indicated that under conditions of enforced delay of response (of more than three signals), there is a decrement of performance, rather than an improvement as should be expected if in fact human beings could be compared to ideal communication channels. Presumably other constraints become operative at this point, eg., memory.

The restrictions noted by Cronbach and Laming are very serious; the concept of channel capacity has to do with only the transmission phase of the communication sequence, but all the "measure" of man's "channel capacity" which we have quoted lump together activities, or stages of encoding, decoding, and remission of message from source to destination (i.e. the execution of a response). Thus whatever it is which has been measured by investigators, it cannot be channel capacity in its pure form. And since the concept of an information overload has been explicated in terms of channel capacity, the foundations of our inquiry are put in danger.

How significant a role does encoding play in the process?

While the experiment quoted above (Kirchner, 1958) indicates that man is not an efficient machine in the classical engineering sense, we have against this result the following opinion, due to G.E. Miller: "The most glaring result (of the choice-reaction experiments) has been to highlight man's inadequacy as a communication channel... It is my own opinion that man's peculiar gift is his ability to discover new ways to transform, or to recode, the information which he receives. It seems to me that the very fact of our limited capacity for processing information has made it necessary for us to discover clever ways to abstract the essential features of our universe and to express these features in simple laws that we are capable of comprehending in a single act of thought. We are constantly taking information given in one form and translating it into alternative forms, searching for ways to map a strange, new phenomenon into simpler and more familiar ones. The search is something we call "thinking"; if we are successful, we call it 'understanding'" (Miller, 1956).

If Miller's observations are correct, i.e., if the human subject is in fact constantly trying to make sense out of - to read pattern into - sequences of stimuli, then all we have to this point are very noisy measures of human channel capacity indeed!

#### Experimental findings: Problems and Inconsistencies

One question which has excited some experimental interest is the following: are increases in information transmitted due to changes in size of ensemble equivalent in their effects to changes in the rate of presentation of stimuli (where it is assumed naturally that total information transmitted is always equalized)? Or to state the question

another way: since it is already well established that increases in the rate of presentation of stimuli beyond a certain point result in decrements of subject performance, i.e., overload (eg. Vince, 1949), can the same effect be obtained by increasing the size of the ensemble?

Alluisi and Muller (1956, 1958), Alluisi, Muller and Fitts (1957) used a random sequence of arabic numerals projected on a screen at a uniform rate. The experimental manipulation consisted in changes in the ensemble of numerals employed and rate of presentation of symbols. The effect predicted by an information theory interpretation appeared in the latter case, but was much less evident in the former. Mowbray (1960) presented numerals to subjects, varying the ensemble from which numerals were drawn from two to ten. Subjects were informed in advance which numerals might appear, and the size of the ensemble. No differences in reaction time were observed.

It may be objected that these experiments raise the question of the true ensemble, and as we observed earlier in our discussion of cellular transmission rates, unless the ensemble is known, the rate of information transmission cannot be estimated accurately. We might be led for example to distinguish between an "explicit" and "implicit" ensemble: the explicit ensemble is the one which the subject receives from the experimenter, the implicit ensemble he brings with him as a result of a lifetime of training. Thus in the case of numerals and letters there is probably little which can be done experimentally to overcome the subject's own personal implicit ensemble.

In Hilgendorf's (1966) experiment, types of stimuli were varied, all highly learned (principally letters and numbers). The following procedure was employed: equiprobable stimuli were presented singly, in random order, at the back of a box with a single window. Subjects held their hands on a palm key until they had identified the stimulus, then lifted their hand to press the appropriate response key. In this manner it was possible to discriminate between "recognition time" and "movement time". Ensembles of numerals of up to 1000 were used (constituting approximately 10 bits of information). Hilgendorf's findings show the effect predicted by Hick's equation:  $\overline{CRT} = a(n + 1)$ . They also indicate that a relatively large part of the overall effect is due to increases in movement time: the maximum rate of information transmission obtained with both phases considered together is 5.5 bits per second, but if we look only at the recognition phase, this rate rises to 27 bits per second. At no point does anything like an overload effect appear. Other experimental results, using a similar design, indicate that so significant number of errors occur in the ensembles of less than 20 bits of information.

From these experiments, it may be inferred that the determination of optimal rates of information transmission for the human system depends very strongly on how familiar the stimuli are for the subject, i.e. to what extent they are overlearned. Other research has shown that estimates of maximum transmission rates seem to vary widely from experiment to experiment depending on the degree of stimulus-response compatibility. In Leonard's (1959) experiment, subjects rested their fingers on a set of relay armatures, and were required to depress any armature which vibrated. No differences in reaction times whatever were found when the ensemble was varied to include two to eight alternatives.



In general, it seems clear from these and other findings that a) the overload effect is not clearly associated with increases in the size of ensemble, and indeed b) that the determination of channel capacity has not been achieved empirically beyond ambiguity. This is true even though, for reasons of experimental efficiency, the set of stimuli used have generally been very simple (on-off lights, moving dots, single letters and numbers, etc.). In ordinary life we are used to dealing with more complex stimuli: our eye takes in a glance a multitude of details. Unfortunately for the experimenter, there are no behavioral equivalences sufficient to represent directly the exact informational content of the message received.<sup>(1)</sup>

Hence for "destination" (conceived as behavioral output) to be the same as "receiver" (in the Shannon model), the experimenter is restricted to single stimuli.

The communication channel model has been found wanting on other grounds as well. Hyman (1953) found that, when the probabilities of stimuli were not equal, the equation  $CRT = a + bH$  made a good fit of the data only if one was considering the overall mean: when one turned to look at reac-

---

1 The reason is related to the distinction made by Jakobson and Halle, 1956; Jakobson, 1964) between selection and combination. Jakobson observes disorders of aphasia, associated with the successful decoding and encoding of verbal information, can be grouped under two headings: similarity disorders -- which occur during the input of information - and contiguity disorders -- which are associated with difficulties in combining elements into the appropriate output patterns. There are thus two types of relationship among stimuli which organisms must deal with in the process of transmitting information.

tion times to individual stimuli, the predictions of information theory were very wide of the mark. In one of his experiments, probabilities of stimuli (size of ensemble = 4) were set at 13/16, 1/16, 1/16 and 1/16, respectively. Mean reaction time as predicted by the overall regression line fitted to the reaction times of all 24 experimental conditions (for 0.99 bits of uncertainty) was 363 msec., and the obtained time was 361 msec., a very close fit. However, in looking at reaction times to the individual stimuli, the predicted times, following an information theory interpretation, would have been 258 and 824 msec. for, respectively, the frequently occurring and infrequently occurring stimuli. The observed mean reaction times for the most probable stimulus were 306 msec.; the observed reaction times for the least probable stimuli were 585 msec. Neither of these figures are close to the predicted figures.

It would be difficult to find outright support for the use of a communication channel model in these findings.

A somewhat different, if related, question is the following: are effects due to variability in the time of arrival of stimuli (i.e. variability in the inter-stimulus interval, or ISI) equivalent to those of variability of size of ensemble? (Uncertainty concerning the inter-stimulus interval does not imply changes in the overall rate of presentation, so that the two questions are distinct). Alegria and Bertelson (1970) varied size of ensemble (2, 4 and 8) and time uncertainty (for periods of 0.5 and 5.0 sec.), and found that, when amount of practice is equated, effects due to the two dimensions of uncertainty were independent. Thus, in terms of the equation:  $\overline{CRT} = a + bH$ , time uncertainty affected the intercept, ensemble size the slope. Hence time uncertainty

would seem to be associated with simple reaction time, and size of ensemble with choice reaction time. From this result we might draw the conclusion that if channel capacity is limited, it is limited in different ways at different stages intervening between stimulus presentation and response. Are there, therefore, different types of overload, and if so how can they be specified? It is to this question which we shall shortly turn. First we must attempt to summarize our discussion of the Miller information overload hypothesis.

The communication channel model in perspective

The late fifties, roughly the period when G.A. Miller's articles and Broadbent's book appeared, appear now to be the high-water mark of enthusiasm among psychologists for the information theory model. There was more than a slight tendency to state flatly that the human organism could be described as a communication channel. Even then there was an awareness of the limitations in the applicability of the communication channel model: as G.A. Miller (1956) observed ironically: "It is an act of charity to call man a channel at all. Compared to telephone or television channels, man is better characterized as a bottleneck". Subsequently, there has been if anything a further backing off from extreme positions on the subject.

What judgment should we now make? Basically, a scotch verdict: not proven. Within limits, the communication channel model works, and its use is accordingly justified. Both on intuitive and on experimental grounds, it seems reasonable justified to state a relationship between stimulus uncertainty and response latency (and ultimately efficiency of response).

Furthermore, these results are not limited to the somewhat artificial conditions of the laboratory. Richard Meier has observed similar conditions in organizations which are subject to periodic overloading, such as libraries, stock market exchanges. Meier (1962) describes the eventual breakdown of information processing that occurred in the American Stock Exchange, in 1959, following a sudden quadrupling of orders. The extent to which conditions of superabundance of available information have resulting effects on habitants of the (increasingly wired) city

has been discussed recently by Milgram (1970). We can hardly doubt therefore that overload occurs and that it has something to do with information-processing demands on the individual.

There are however difficulties. The communication channel model is too crude an instrument to serve as a tool of investigation of phenomena of human information-processing behavior. If interpreted literally it leads to the kind of extreme telescoping described in the last section, with an attendant difficulty in deciding how to explain apparently inconsistent results on the basis of an insufficient number of variables. To assign everything between stimulus and response to "channel" is too much. To classify responses as the "destination" of the message demands too great a distortion of our ordinary conception of the role of behavior to be easily accepted. As Craik (1943) long since pointed out, human beings do not ordinarily behave as a simple linear throughput system like a telephone network. On the other hand, if the channel model is used only as an analogy, it leads only to an empty heuristics.

In order to render useful the central insight of the communication channel theory, namely that somewhere between input and output there are mechanisms of limited capacity, we can no longer avoid facing the difficult question of attempting to specify the various stages of information-processing which intervene between the arrival of a stimulus and the execution of a response. In addition, we will attempt to discriminate between two kinds of limited capacity, a temporal limitation (associated with temporal uncertainty), and a spatial limitation (associated with size of ensemble).

It is to this problem that we now turn. In expending the set of assumptions with which we have been working, we will no longer find it necessary to adhere to the fiction that humans are restricted to behaving as simple linear input-output systems.

### Modelling the organism as a multi-stage adaptive system

We have seen certain limitations in the application of the communication channel model of the organism which have led us to reject it as an adequate theoretical basis for our inquiry. Our task now becomes to develop an alternative tool: one which is consistent, if possible, with the considerations outlined in the second part of Chapter 1. We will introduce two new principles: a) the principle of central intermittency in the organisation of behavior, and b) the principle of decomposability of reaction-time data. To develop these concepts more fully we will examine a certain amount of research material in greater detail. Finally we will attempt to pull together what we have found out into a preliminary synthesis, with particular emphasis on how the point of view developed in this chapter can serve as a guide to future research.

#### The principle of central intermittency -----

In two important articles published in 1947 and 1948, a young British psychologist, Kenneth Craik, advanced a view of the organism as an intermittent correction servo-mechanism, in the language of contemporary control theory. Craik noted that, in tracking tasks, as the difficulty of the task increased (i.e. as the information load increased) subjects become



incapable of tracking variations in the path of the stimulus object in a continuous fashion: instead they resorted to behavior which consists of a sequence of individually discrete, or discontinuous, adjustments to the moving stimulus. Each discrete correction requires about .5 seconds to complete. Hence, Craik reasoned, it cannot be correct to compare the nervous system to a vast telephone switchboard, which continues to receive input at increasing rates and to increase output continuously until the system jams. In Craik's model, the organism behaves rather more like a computer, accepting inputs, making transformations (both logical and computational involving both immediately available and stored information), and producing outputs. The output of the human organism in turn requires some monitoring time, during which feedback is required to determine that the desired response has been correctly executed. Craik suggested that decision time required .3 seconds, monitoring (or execution) time .2 seconds -- together accounting for the total .5 second response time. During monitoring the organism is not immediately responsive to new information or signals arriving during that period, thus giving rise to the notion of a psychological refractory period, analogous to the refractory period of the neuron.

The concept of a psychological refractory period had been advanced earlier by Telford (1931), following the analogy of neuronal refractoriness, and the concept was given further currency by some of Craik's colleagues in England. The present consensus among psychologists is that the term is a misnomer, and ought to be replaced by the more useful concept of intermittency (Bertelson, 1968; Smith, 1969), since it is clear that,

unlike the neuron, the organism is capable of accepting new information during the "refractory" period. For example, Vince (1949), using a design which required subjects to respond to dots on a rolling white paper tape by tapping a telegraph key, found that omission and errors occurred when the next input marker had arrived before the previous response was completed. However, subjects were able to maintain a high level of correct responses even under conditions of slight overlap. Mackworth and Mackworth (1956) similarly found evidence to indicate that omissions and errors were correlated highly with amount of overlap. However, in their experiment the task was rather complex, requiring an identification of six items on each of two cards, and an item by item matching of each of the pairs of items. Webster and Thompson (1953, 1954) also found that the amount of overlap was related to the efficiency of transmission, but concluded that the effect depended on the amount of information associated with each marker: where markers carried little information the effects of overlap were less serious. Broadbent (1958) concluded that two messages could be dealt with simultaneously if they conveyed little information.

The principle of central intermittency is consistent with, indeed implies, a hierarchical organisation of behavior: since one activity is to be terminated before another is undertaken, we must infer a master program responsible for the sequencing of activities.

It is also consistent with the decomposing of response-formation processes into functionally distinct subprocesses, to which we now turn.

### The principle of decomposability

A second assumption made by Craik, widely shared by other researchers in the field, is that the total reaction process is an amalgam of stages intervening between input and response, and, more importantly, that total reaction time is decomposable into its several individually separable components.<sup>(1)</sup>

The assumption of decomposability dates back to the work of a Dutch psychologist, Donders, in the nineteenth century. Donders reasoned that choice reaction time is made up of two components a) a simple reaction time, and b) choice time.<sup>2</sup> He argued from this assumption that it should be possible, experimentally, to isolate the effects due to each phase of the total choice reaction. (The equation cited above  $CRT = a + bH_t$ , exemplifies mathematically the Donders assumption: the intercept a is interpreted as a simple reaction time parameter, the slope b as a choice time parameter). The difference, Donders reasoned further, between a c- reaction and a b-reaction (choice reaction) is due to the additional time required in the latter for response selection. The difference between an a-reaction (simple reaction) and a c-reaction is explainable as an effect due to the necessity for stimulus recognition in the latter case.

Donders thus posits three distinct phases of reaction: recognition, choice and simple reaction. He assumes them to be additive: that is, he

---

(1) Stated with great clarity by Sternberg (1969) See also Smith (1968 ).

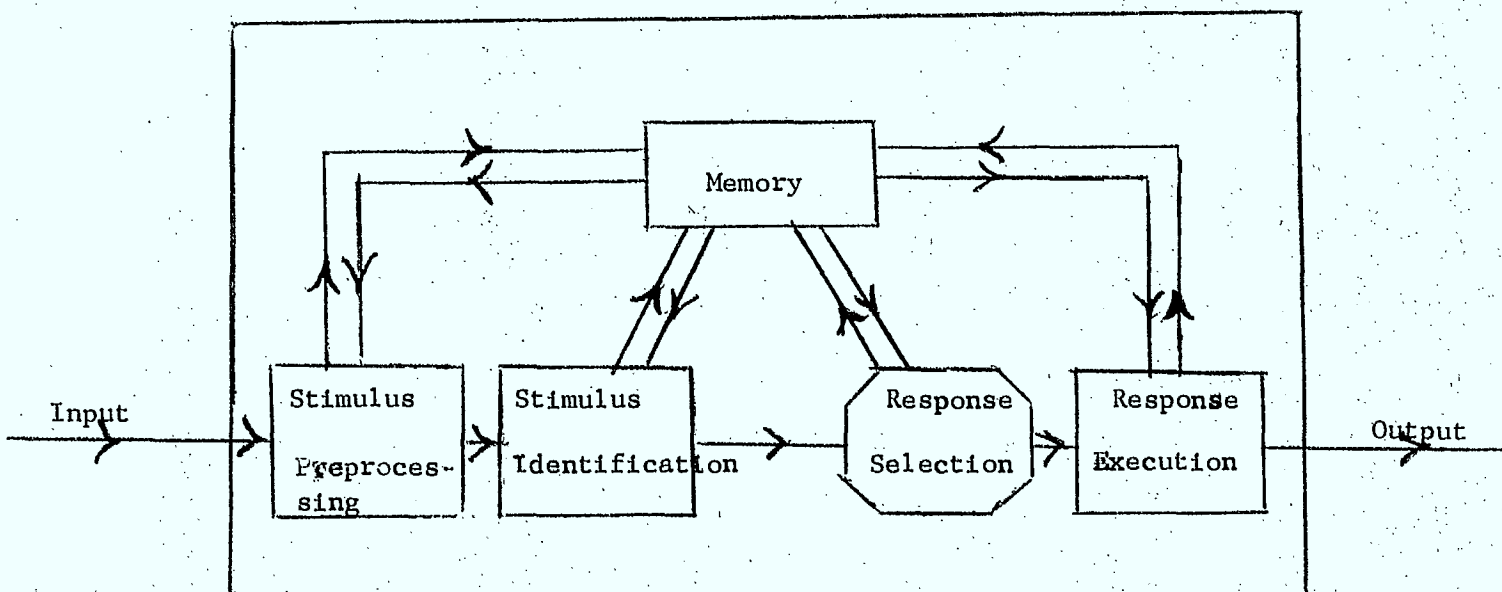
(2) See p.II-4 above.

assumes that one phase is terminated before the next begins. From this assumption, it follows that experimental means can be found to isolate the time required to complete each phase in turn, and hence to identify with greater specificity the effect of various types of environmental variability. This is the general procedure which we intend to follow.

The current tendency is to identify 4 stages which intervene between stimulus presentation and response execution:

- a) stimulus pre-processing: raw sensory data is organized to produce a preliminary impression of the stimulus object;
- b) stimulus identification: the stimulus is placed in an available category through the application of pre-existing representations held in memory;
- c) response selection: on the basis of the available information a response is chosen (utilizing logical and computational operations);
- d) response execution: the response is effected, and immediate feedback monitored.

The notion of stages is illustrated in the following diagram:



Model of stages of information processing on the individual  
(After Welford)

#### Attention

-----

In the last two sections we introduced the concepts of the division into stages of information flow within the individual, and of central intermittency. The latter notion implies that while the organism is preoccupied with one phase, for example the execution of an action, his capacity to perform another phase is reduced. He has what we would call a limited attention to give to any phase of the activity cycle at any given instant. "Where is the channel limited", it has been asked?<sup>(1)</sup>

(1) Moray (1967).

The answer, it appears, is in the mechanism of attention. What then is attention? It does not figure as a stage in the model outlined above, and we assume it to be a mechanism of a different order: that which regulates the organization of all the other activities. The importance of this mechanism has become increasingly clear.

During the 1950's Colin Cherry initiated the investigation of a phenomenon which he termed the "cocktail party effect". He asked how, in the midst of a noisy room with many competing conversations audible to the listener, the listener is able to select out one source from the others and disregard all other irrelevant speech which is available. Cherry's investigations were pursued further by Donald Broadbent (1958), who rejected the hypothesis of the operation of peripheral processes in favor of an explanation based on a central selector mechanism, which acted as a filter to reduce the "load" on higher-level cerebral components by utilizing largely physical cues (directionality of sound, timbre of voice, etc.) to screen out irrelevant material. (1)

Broadbent's assumption that physical cues alone were the operative factor was shown in turn to be inadequate by Treisman (1964), who demonstrated that the meaning of signal also served as an important selective cue. If, however, even rejected information, which could not subsequently be recalled by subjects, had been processed to the point of assigning a

---

(1) These research findings are reported in greater detail below.

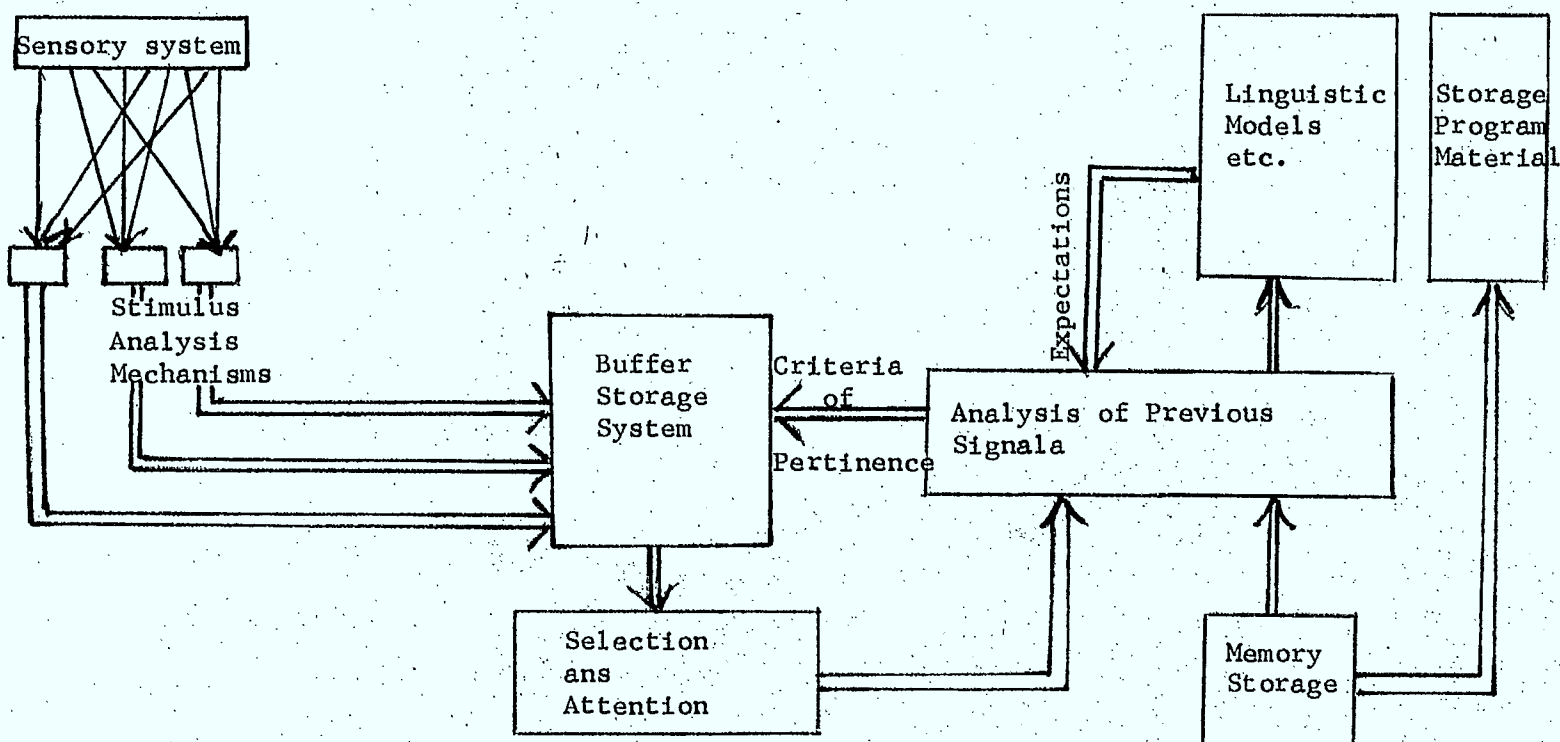


semantic interpretation, then a good deal of central processing capacity must have been already used in the operation of the filtering mechanism itself.

It has recently been proposed (eg., Norman 19 69) that the

selection process is constructed somewhat as follows:

Physical signals



Incoming sensory material is analyzed and excites a representation in a buffer storage system, or transient memory. Concurrently, the analysis of previous signals is going on, and this information also excites a representation in the transient memory. This analysis of previous signals establishes for the individual a scale of pertinence, since it makes it likely that out of the available information, what is selected is related to what has gone before. In this respect, expectations based on prior experience, on linguistic models, etc. may in turn influence the determination

of what is pertinent. Through the interaction of new sensory material and analysis of previous signals, the strongest representations in transient memory are selected out for attention and in turn are fed into the analytic process, to affect, in turn subsequent selection processes.

A mechanism of this kind, supplemented by adequate memory capacity, is sufficient to provide a dynamic for the multi-stage model proposed earlier. Suppose for example the individual to be occupied by the effectuation of a response. Responses usually consist of step-by-step execution of instructions to individual effectors and the monitoring of immediate feedback. Information which is now pertinent consists of feedback cues and other information tends to be disregarded until that part of the program is completed. Attention then shifts back to the recognition of incoming information, and subsequently to choice of response, until the cycle is complete. Because of the buffer storage system, a certain degree of overlap of signals is not incompatible with responsiveness to the total available stimulus information.

#### The operation of the buffer storage

-----

The presence of a very short term buffer storage capability is strongly suggested by experiments of Sperling (1960), Averbach & Coriell (1961), and others. Their experimental evidence seems to be best explained by a theory of what Neisser (1967) terms transient iconic storage. When a visual stimulus is presented to a subject, the sensation of the stimulus may outlive the presentation of the stimulus. This is explained by the well-known psychological finding that each sensory modality is associated with

a projection area in the brain. The concept of an iconic memory supposes that "the persistence of visual impressions makes them briefly available for processing even after the stimulus has terminated" (Neisser, 1967).

The experimental method employed involves the use of tachistoscopic display of stimulus material. The principle of tachistoscopic experiments is that the subject is shown extremely brief presentations of material, which is then re-presented in successive exposures of gradually increasing duration. Sperling (1960) used rectangular arrays of letters such as the following:

T D R

S R N

F Z R

which were displayed for periods of 50 milliseconds, too brief for the eye to respond actively by changes of fixation. In general, subjects were not able to read more than 4 or 5 letters (consistent with Miller's notion of a "span of attention" or "span of apprehension"). Sperling then instructed subjects to read only a single row of the display. Subjects were cued by a different tone for each row, sounded after the tachistoscopic presentation. The result was near perfect accuracy for the selected row. Averbach and Coriell (1961) showed similar results, substituting a single letter for a row, and a visual pointer rather than a tone.

From these results, it can be concluded that a visual input can be stored briefly, that it decays rapidly, but that while still present

in memory, information can be read from it. Subjects reported that the letters were visually present and legible, even though the stimulus had not any longer been present for 150 milliseconds.

A similar mechanism associated with auditory input has been posited by Neisser, who terms this auditory storage "echoic memory".

#### The operation of the selector -----

In considering mechanisms of information reduction, it is well to bear in mind that the total of all information provided by the sensorium of the human organism far exceeds the amount which can be usefully utilised. The retina of the eye alone contains in the order of one hundred million cells, and the optic channel to the brain carries about one hundred thousand nerve fibers. It has been estimated that the ear is able to transmit 8,000 bits per second, the eye perhaps 3.4 million bits per second (Jacobson, 1950, 1951). The effective difference between the information provided to the brain, and the amount it is able to use is considerable.

It has been suggested that information reduction occurs in several stages; here we consider two: a) those associated with primary sensory analyzers, and b) those associated with the retrieval of information from buffer storage.

#### a) Primary sensory analyzers

To understand the limits on selection of input, we must first establish the capacity of the organism to discriminate differences.

Humans are very good at relative discrimination, which simply implies a comparison along a dimension or dimensions (essentially more or less of some attribute(s); many tasks however require absolute discrimination, or judgment in the absence of any external reference. George A. Miller (1956), in a famous article "The Magic Number Seven, Plus or Minus Two:

Some limits on our Capacity for Processing Information", made an extensive review of the literature on this subject, comparing the experimental results across sensory modalities, and arrived at a principle of the span of absolute judgment, which say that "there is a clear and definite limit to the accuracy with which we can identify absolutely the magnitude of a unidimensional stimulus variable". (Miller, G.A.; 1956). The subject's ability to discriminate in absolute terms has been evaluated by Miller in terms of information theory, and he has estimated approximate maxima of 2.5 bits in the judgment of tones (Pollack, 1952; Pollack, 1953), 2.3 bits for judgment loudness (Garner, 1953), 1.9 for judgments of the concentration of salt solutions (Beebe-Center, Rogers, and O'Connell, 1955), 3.25 bits for judgments of visual position (Hake and Garner, 1951). The limits are in all cases approximate: subjects begin to make occasional errors as the number of discriminations required reaches three or four, and increases steadily as the number of discriminations also increases. Such upper limits of discrimination do not depend on the range chosen: Pollack discovered that the same subject who could accurately discriminate 5 high-pitched tones, presented in one series, and 5 low-pitched tones

---

in a second series, could still only distinguish 5 tones when the ensemble included both high-and-low-pitched tones. (1)

The importance of the perceptual constraints imposed by the span of absolute judgment may be better evaluated when placed in the context of the complete perceptual system. First, they apply only to tasks which require absolute judgments; where the problem is one of relative judgment, the same limits do not hold. Secondly, the results reported by G.A. Miller refer only to unidimensional judgments; most sense organs are capable of simultaneous discriminations along more than one dimension simultaneously (leading Miller to propose a second principle of the span of perceptual dimensionality). Experiments which utilize a two dimensional variation in stimuli have demonstrated increases in information transmitted varying from 2.3 bits for saltiness and sweetness combined (Beebe-Center, Rogers and O'Connell, 1955) to 4.4 bits for dots in a square (Klemmer & Frick, 1953). The addition of further variables increases the judgment capacity, but the additional information transmitted is less than additive.

#### b) Selection of information from buffer storage

Much of the available sensory stimulation provided by primary exteroception mechanisms is discarded at a second stage of processing. The internal filtering mechanisms, which are utilized by the organism as a normal part of his perceptual process, operate both within and between

---

(1) A comparison of these results with those mentioned earlier for the single cell, where Mountcastle, Davies and Berman (1957) found that input intensity is coded in 7 discrete steps for thalamic cells, suggests that the limits reported by Miller have a physiological basis. A different explanation not yet studied in detail might see these limits on the organism's "resolving power" as a function of short-term memory constraints. We have not as yet found a discussion of this point.



sensory modalities;<sup>(1)</sup> the effect of the filtering is a reduced and more manageable picture of the environment, to which the organism can more capably respond.

With respect to the two most important channels (for the human), the visual and the auditory, the process differs somewhat. The visual field normally contains a diversity of objects, to which we attend only in part. Everything which is going on in the field of view is not of equal importance, and we become conscious mostly of events which are relevant to our activities, so that other things in the periphery do not really exist for us. The eyes operate by making sequences of saccadic movements, or jumps, from one fixation point to another, remaining fixated about 85 per cent of the time. In this way, if important events occur in different parts of the visual field, they can be scanned in a succession of fixations. Such scanning processes serve to permit the eye to select a point of attention and to ignore other information; they also, incidentally, illustrate well the principle of intermittency alluded to above.

The ear works on another principle from the eye, and external scanning is accordingly more difficult. Spatial relationship between events in the audioscope are difficult to determine with precision, while the perception of temporal relationships can be rather easily affected by overlap of messages, irrelevant atmospheric noise and the like.

---

(1) As contrasted with the "adjustment" processes mentioned by J.G. Miller, this customary resort to omission, filtering and abstraction seems to form an essential element of the perceptual system. Even the phenomenon of information overload finds a useful application within the perceptual system as a whole: an example is "critical flicker fusion", where the overloading of some cells permit certain specialized kinds of perception, necessary for example to film-viewing.

If two messages arrive at once, and if both are relevant (to be attended to) the resulting overlap produces a considerable loss of information (Webster and Thompson, 1954; Poulton, 1956). Broadbent (1954) read lists of digits to subjects over separate channels into each ear. Under conditions of overlap, subjects appeared to be able to deal with both sources at rather slow rates of presentation, but as the rate of presentation increased, they increasingly showed a tendency to pay attention to stimuli reaching only one ear. At certain speeds, he found that although inputs to one ear were dealt with before those to the other ear (Titchener's phenomenon of "prior entry"), subjects were able to retain additional information in memory for a relatively short period. These results, as noted earlier, depended on the amount of information associated with each message source: messages which convey little information can be dealt with simultaneously, while with more information, overlap produces correspondingly greater decrements in transmission.

In general, if some information is to be retained, and some discarded, the ear is better able to select out the wanted portions of the message if the sources can be isolated, e.g., by being fed into different ears. Hirsh (1950) and Kock (1950) showed that noise has less effect on intelligibility if, for example, two loud speakers are employed separated physically, one for the relevant signal, the other for noise.

Different levels of explanation have been offered for these findings: it was first proposed that peripheral sensory masking alone was sufficient to explain the difficulty of the organism in paying attention to two high-information sources. Broadbent (1958) rejected this interpretation in favor of a second explanation which supposes a rôle for mechanisms originating in the central nervous system. His assumption was based on an examination of expe-

rimental evidence concerning the effect of instructions on subsequent performance. When the subject is asked two questions simultaneously, if the experimenter announces which voice is to be answered, the subject is generally able to respond as instructed. In the absence of such instructions, or if the instruction is issued after presentation of the stimulus, the performance of the subject shows serious deterioration. Such results do not support a theory of peripheral sensory masking; they do however support the model presented above.

Broadbent's supposition that the selection process consisted of a choice of "channel", using largely physical cues, was in turn shown to be inadequate by Treisman (1964), who found that between messages read by the same person in the same language, selection was based on transitional probabilities between words, although there was considerable interference from one passage to the other. In this case it is apparent that "selection" is delayed until the moment of readout from buffer storage. It does not depend, for example, on which ear receives the message (Grey and Wedderburn, 1960); hence it will not do to identify "channel" with a particular organ of reception.

In order to investigate the Broadbent theory of alternating attention between channels, Moray and Jordan (1966) investigated a highly compatible two-channel task. They provided subjects with a means of parallel output matched to parallel inputs, simultaneously presented to the two ears of the subject. They found that this procedure increased quite considerably the overall "channel capacity" of the organism (from 1 pair every  $1\frac{1}{2}$  seconds to 2 pairs per second). The explanation given for these differences notes that the earlier experiments require parallel to serial conversion of data, and hence allow for the intervention of memory constraints.

The findings reported have applied only to information-reduction processes within single sensory modalities; there are further losses between multiple sensory inputs. In general, it appears to be possible to carry out simultaneously two redundant tasks involving more than one sensory modality: many people find it possible to drive a car and at the same time engage in animated discussion. This capability is in turn limited by the information-processing requirements of the separate tasks involved. However, when subjects are required to perform simultaneously visual and auditory scanning tasks, and if difficult material is presented to one channel and easy material to the other, the easy material is disregarded (Harris, 1950). When there is overlap involving symbolic material, presented rapidly, then one sensory input is disregarded completely (Mowbray, 1954). The presence of noise intensifies these effects (Broadbent, 1953).

Where the inputs arriving via different sensory channels are non-competing, there is some evidence that they interact (Bernstein, 1970). The effects, however, although frequently are not necessarily additive; it appears that where there is ambiguity in the interpretation of a stimulus received on one channel, the information provided by another sense is used as a means of verification<sup>(1)</sup>. Thus there is a tendency for observers to interpret the directionality of sounds by means of an apparently related visual event (Thomas, 1941). When interpretations clash, it is the visual information which is given priority.

---

(1) Birdwhistell (1970) has proposed that our use of non-verbal cues has a similar function in communication.

A general result in the experiments discussed here is that where there is competition among signals, in the sense that alternative stimuli are present to be attended to, then the organism shows a tendency to suppress one source in the interest of continued successful reception of another source.

# Pre-attentive processes: some conclusions

We have been led to the conclusion that the act of perception includes a process of "read-out" from the available ensemble of sensory stimuli, or a re-codification of input into another form which can be stored more easily. The translation or recoding must often imply going from a visual to a verbal medium. Neisser argues from this that "perception is not a passive taking-in of stimuli, but an active process of synthesizing or constructing a visual figure. Such a complex constructive act must take a certain amount of time" (Neisser, 1967). Such "analysis-by-synthesis" requires a concept of organization of activity based on what has been termed a contiguity relationship: a capacity to combine acts in sequence.<sup>(1)</sup> Hence, included in the act of perception, is a performance which follows the type of organization we usually associate with motor activity. Perception includes an activity in which successive steps occur, and as Neisser notes, such an activity requires time to perform.

Whithin the analysis-by-synthesis model of perception, the role of attention is critical. Attention determines what region of the sensory field is to read out, or re-coded.<sup>(2)</sup> The processes of focal attention cannot operate on the whole field simultaneously. Such processes pre-suppose some degree of prior "setting": they can come into play after preliminary operations have already segregated the figural units involved (Neisser, 1967). Neisser terms such preliminary operations "preattentive processes". They "produce the objects which later mechanisms are to flesh out and interpret". They are in essence crude

---

(1) See note p. II-38 above.

(2) "A 'perceptual set' operates by affecting what the subject does during the brief period of iconic storage. This does not mean, however, that the set affects only "response" and not 'perception'... There are no instantaneous perceptions, no unmediated glances into reality. The only way to use the term 'perception' sensibly is in relation to the extended processes that can go as long as the icon continues". (Neisser, 1967)



approximative images of the world, often as much determined by prior expectations as by the nature of the stimulus itself.

A point to note is that, in addition to providing the raw material for more refined perceptions of the external world, these crude, half-processed images also serve to set in motion response tendencies which in the model presented above would seem to be reached only after intermediate phases of recognition and choice of response had been passed. It is somewhat like the batter who begins his swing as soon as the ball is thrown, before he has any idea if it is likely to be a strike or a ball.

Dixon (1972) has argued recently on the basis of findings from experiments in subliminal perception that "at an early pre-conscious stage in cerebral processing, incoming information actually makes contact with memory systems, thereby activating conceptual associates to the applied stimulus". The evidence for this conclusion is often intriguing: Dixon for example quotes the case of one subject who guessed "Valparaiso" when presented with a sub-threshold representation of the word "LINE". Later, in another test he gave the word "Line" as a first association to "Valparaiso". The reason, it emerged, was because he had once voyaged on a liner called Valparaiso.

Other evidence indicates that material presented subliminally is capable of producing subsequent effects on dream imagery, and thus, while not perceived consciously, seems to be held in storage for a considerable time.

A more relevant finding is that conscious recognition of a signal is found to depend on its emotional significance to the individual. Dixon

concludes that information can be received, classified and even responded to without ever becoming conscious: "All in all, these data suggest that at some preconscious stage of the perceptual process the brain detects the meaning of the incoming stimulus and so initiates an appropriate change in its level of sensitivity for their conscious representation". This assumption is consistent with the idea that a first response is initiated, and then modified by the control system, when the full meaning has been extracted. The only physiological requirement this makes is that the "regulator" channel should conduct faster than the "information" channel.

This capability for rapid response based on early alerting is important where rapid responses are required. Fehrer and Raab (1962) measured latency of response to stimulus alone and stimulus followed by mask (where only an impression of movement was possible) and found no difference. Fehrer and Biederman (1962), Schiller and Smith (1966) obtained similar results. This finding illustrates the point that while, at one level, the process of read-out or recoding is still underway, the organism has already begun his response on the basis of a first signal that an event has occurred; "The mechanisms which register this onset are different, simpler, and faster than those which identify the letters" (Neisser, 1967).

This theory serves to explain two phenomena: (a) the association of error with increasing rates of signal presentation, and (b) Kornblum's (1967, 1968, 1969) finding that reaction times in sequences of equiprobable stimuli are significantly faster for repetitions than for non-repetitions.

In the first instance, error is naturally associated with increasing reliance on crude pre-attentive processes.<sup>(1)</sup> In the second case, the "setting" of attention should be expected to follow the previous behavior, in the absence of other clues.

The distinction between pre-attentive processes and secondary recoding suggests in turn a reason for the relationship between information presented and latency of response. The concept of information is equivalent to that of variety, and is associated with the idea of "surprisingness". The image formed from the operation of pre-attentive processes is approximative and indistinct: it permits the organism to begin an (approximative) response while more refined verification procedures can be accomplished (and the response adjusted accordingly in the light of fuller information). The less informative the stimulus, it should follow, the more likely the image due to pre-attentive processes is to be accurate, and the more cursory can be the verification process. The more informative the stimulus, however, the greater the extent of verification required, and the slower the response.

- 
- (1) But, as argued earlier, it is precisely output activities which are most affected by the rate of presentation of markers, independent of the rate of information transmission.

The rate of presentation of markers has two possible effects: (a) Since increases in rate of presentation necessarily must eventually reduce the duration of the stimulus, there must come a point where the brevity of display itself causes non-recognition. Mackworth (1963) showed that for values of less than 50 milliseconds, recognition declines sharply. (b) Increases in the rate of presentation of markers result in increased "crowding" of signals into a given temporal period. In this case, we should expect to find increasing confusion of one stimulus with the next, or "masking". There must in a word be a limit to the temporal resolving power of the visual system.

A great number of experiments indicate the correctness of this assumption; an interesting result associated with the theory of an iconic memory is the phenomenon of backward masking, in which a stimulus presented later masks or obscures an earlier one, which is still present as an icon (Sperling, 1960; Eriksen & Lappin, 1964; Eriksen and Collins, 1964).

### The identification of stimuli

In explaining the direct relationship between stimulus entropy and reaction time, most theorists have focussed on the stage at which identification of the stimulus occurs as one critical determinant. (Smith, 1968) In Bricker's equation,  $\overline{CRT} = a + bH$ , the intercept,  $a$ , is taken to be equivalent to simple reaction time; the slope,  $b$ , represents processes of signal recognition and response selection. In this section the first of these phases is considered.

It is generally agreed that stimulus recognition is accomplished when a representation of the incoming stimulus has been compared with pre-existing representations stored in memory (to which names are mapped) and a match has been found. We can distinguish two major questions concerning the process, both of which have excited a good deal of interest and controversy: a) first, what is the nature of the "fitting" operation between new stimulus and memorial representations; and b) secondly, what is the nature of the search process among available memorial representations?

#### Fitting the stimulus to the memorial representation -----

Two major theories of how the matching process takes place can be discerned: a) "template" theories, and b) "feature-testing" theories. Template theories assume that stimulus information is presented in a central display area, where a generalized image, or template, of the concept concerned is then compared, and, depending on the closeness of the fit, an identification is made, or rejected. If positive, the stimulus has been "recognized". For example, if the stimulus to be recognized is a letter of the alphabet, the actual stimulus, say an "A", is tried on for fit with the templates of each letter in turn. Since the raw stimulus information may be quite "noisy", in that there is a good deal

of variability in the form in which characters are presented, it is assumed that a certain amount of "cleaning up" of the image may intervene between stimulus presentation and identification.

Feature-testing models assume a more active transformation of stimulus material: essentially a mapping into the semantic space of the individual. Identification consists of the application of a series of tests, utilizing as many dimensions as necessary: "Is it red?" "Is it square?" "Does it move?" etc. Recognition occurs when the number of tests giving positive results is great enough to exceed some critical limit.

Neither explanation is fully satisfactory, but acceptance of the feature-testing model, in some modified forms, has been gaining, principally because of two factors: a) evidence that the neural analyzing mechanisms of the human do appear to operate on a feature-extracting principle, and b) attempts to develop artificial character-recognition systems based on template fitting which have not proven particularly feasible.

#### The memory search process

-----

Hick (1952) outlined four possible search procedures:

##### 1. Replication with simultaneous trial

When a sufficient representation of the stimulus has been formed, it is then to be compared with, or matched against, pre-established memorial representations. Suppose the comparison trials are conducted simultaneously, suppose each trial to take an equal time, then comparison time does not depend on the size of the stimulus ensemble, contrary to the obtained results.

However, we could explain effects due to ensemble size as an effect due to the time taken to produce replicas of the original representation. There are three processes by which replicas could be produced:

- a) simultaneously, in which case replication time would be independent of ensemble size, contrary to findings;  
(1)
- b) serially, in which case replication time would be a linear function of ensemble size and;
- c) geometrically, i.e., by successive expansions or "doublings" of the original representation, which produces a logarithmic function of ensemble size.

## 2. Random searching

On this theory, no replication is required, since each memorial representation is compared singly. There are no simultaneous trials. The memorial representations are chosen one at a time and in random order, and having been tried are replaced, so that each may, in principle, be tried more than once. On this assumption, the average number of trials before the stimulus is identified is equal to the size of the ensemble, not a logarithmic function of it. However, the expected variance predicted from this theory is greater than experimentally obtained results would justify.

- 
- (1) With simultaneous replication, we can produce a satisfactory result by dropping the assumption that all comparison trials take equal time. If we assume that comparison times are distributed around a mean and if we assume a certain kind of distribution (the exponential), we can obtain an appropriate theoretical basis for the obtained results (Rapaport, 1959).



### 3. Systematic searching

This process is similar to the previous except that the ensemble of templates is scanned systematically (either in a random or a given order) without replacement. Assuming a stop rule, average number of trials required to find a stimulus - template match is a linear, rather than a logarithmic, function of the ensemble,  $S + 1/2$ . Assuming no stop rule reaction time is a linear function of size of ensemble.

### 4. Progressive or serial classification

In Hick's fourth category, the notion of template matching is abandoned in favor of a procedure of classification by features. In this view, identification consists of the application of a series of dichotomous tests: Red? (Yes, No) Square? (Yes, No) etc., similar in kind to the parlour game of "Twenty Questions", the answers to which narrow down possibilities until recognition is attained. This process will produce, given the appropriate set of tests and a reasonably unskewed distribution of probabilities, a logarithmic relationship between recognition time and size of ensemble (Hick, 1952).

None of the theories proposed by Hick have proved to be entirely satisfactory. In the original Bricker formulation,  $CRT = a + bH$ , the intercept  $a$ , as we have seen, has usually been taken to describe simple reaction time, including elementary pre-processing operations, and the slope,  $b$ , to refer to secondary recognition and choice processes. None of the template-matching processes suggested by Hick will produce this effect: with simultaneous comparison trials, the slope is independent of ensemble size, with serial scanning, the function is linear. If we shift to an explanation based on replication times, we are compelled to identify the intercept with the recognition phase and the slope with preprocessing, which

is an unconvincing interpretation, on other grounds (Smith, 1967).

The concept of serial classification appears initially more promising, for two reasons: a) because the "Zerging -in -feature testing procedure is capable of producing the desired logarithmic function, and (b) because recent physiological evidence makes it plausible to believe that the perceptual system is essentially a feature-testing procedure, even at primary levels of sensory reception. However at least one experiment designed to provide a direct test of the hypothesized classification process has produced negative confirmation. Leonard (1958) argued that if recognition consisted of performing a finite set of (dichotomous) tests, then providing advance information about one of the tests should have a result equivalent to that of reducing the ensemble by half. Leonard's prediction was not supported.

The serial classification model discussed here is, conveniently, on error-free one, in the sense that no provision is made for classification decisions made on the basis of less than perfect information. This deficiency can be easily remedied by building in some of the aspects of modern (Bayesian) statistical decision-making theory. On this view, the individual goes on collecting information until he has enough to make a recognition test by comparing it with a memorial representation, presumably a list of features. This has the advantage that it seems to explain certain well-established findings, such as:

- a. the speed-accuracy pay-off - several experiments have shown that subjects working under time pressure make more errors than those under little pressure; there is a direct trade-off in information transmitted between speed and accuracy;

b. the effect of value - the Bayesian model is well equipped to explain the tendency of subjects to respond faster to highly valued stimuli than to lower-valued stimuli, since optimal decisions in the Bayesian sense weight both probability and value;

c. the effect of expectation - experiments have repeatedly shown that reaction times are influenced by a priori expectations, which figure in Bayesian computation;

d. effects due to decreases in discriminability - either degradation of the stimulus or greater similarity between stimuli may result in an increase of the number of tests to be performed;

e. effects due to practice - the feature-testing model is compatible with an increasing "automatisation" concept equivalent to practice effects.

Recently, Sternberg (1971) has argued that an information-theoretical view does not provide any real understanding of underlying processes. In some of his research he has attempted to concentrate on the matching from memory part of the total identification process. His method consists of requiring subjects to memorize lists of numerals, drawn from the set of primary digits, of varying size from 1 to 6. The subject is then required to say whether a given stimulus is drawn from the "target" set. If it is he gives a positive, if not a negative, response. The measure obtained is the time of reaction.

From his results Sternberg has concluded that recognition is accomplished by high-speed serial scanning of the target set (Sternberg, 1966). Reaction time, in his experiments, is consistently a linear, rather than a logarithmic, function of the size of the target set.

We can explain the difference in functional relationship between size of ensemble and reaction times in these experiments from those found earlier in different ways. First, it should be noted that, regardless of the size of the target set, the response set consists of only two possibilities, a positive or a negative. Studies by Rabbitt (1959) and Pollack (1963) indicate that CRT increases with the number of responses, and that the number of stimuli associated with a given response affects CRT only when the number of responses is large (in the vicinity of six, for example). A more interesting interpretation has been proposed to explain the apparent incompatibility between the obtained linear and logarithmic functions by Cavangh,<sup>(1)</sup> who suggests that the key mechanism is the span of immediate memory.

In the Sternberg experiments, the subject has to hold in the short-term memory a maximum of six items against which stimuli are to be compared, plus two possible responses. This is still a feasible memory load. In a typical choice reaction-time experiment, however, the subject may have to keep in mind different numbers of items. If the size of ensemble is 1 (simple reaction), then he has only to remember the appropriate response. If size of ensemble is 2, he has two stimulus alternatives, and two responses to bear in mind, still an easy load. For an ensemble of 4, the memory load has risen to 8, and this is the extreme upper limit of the short term memory capacity.

If the ensemble S-R ensemble is 8, the memory load is 16, which is well above the capacity of short term memory, and hence implies access to long term memory, where search rates are known to be slower than those for immediate memory (Sternberg, Knoll & Nast, 1969). While this explanation is still speculative, as a line of explanation it looks promising:

---

(1). Personal communication.

it has the advantage of pinpointing limitations of "channel" capacity in terms of a specific function, memory.

The reaction function obtained by Sternberg, since it was linear, could also be described as a sum of two terms, an "intercept"  $a$  and a "slope"  $b$ . The intercept, as before, is taken to be a term which includes pre-processing, or "cleaning-up" operations on the raw image; the slope reflects memory search operations. Some recent investigations have centered on the question of how much "cleaning up" occurs before recognition processes are started, in other words how the matching occurs. Sternberg (1967) argued ingeniously that if the serial-comparison process occurs relatively late, then effects due to degradation of the image would largely be reflected in changes of the intercept value (since it is known that overall reaction times are lengthened by diminished stimulus discriminability); if the image arrives relatively unprocessed at the point of memory search, then effects due to degradation should be reflected in an alteration of the slope. The results of this work showed in fact that serial comparison does not take place until the image (or list of features) reaches a relatively high point of refinement.

Chase and Posner (1965) showed that increasing the similarity of stimuli, and hence the potential difficulty of determining a definite match, had the effect of increasing the steepness of the slope, as a two-step theory of signal identification would predict.

What can now be said concerning the overall capacity of the organism to identify stimuli, and hence about potential overload? The answer

is complicated by the fact that, in a character-recognition task for example, if a stimulus is presented singly, and if the size of the target set is 5, response time is about 500 msec. If however the same stimulus is presented in an array of other characters, all to be compared with the target set simultaneously, average response time per stimulus is reduced to 100 msec., and even this figure can be improved with practice (Neisser, 1963).

To explain this Sternberg and Scarborough (1971) have proposed that serial-comparison processes can be conducted in parallel. What limits there are on how many such processes can go on simultaneously, and indeed the whole question of the identification of complex images remains a field for more research.

#### The stage of response selection

-----

As Smith (1967) notes, by far the greatest attention has been paid in the experimental literature to processes of stimulus preprocessing and identification. There are few models available to suggest how response selection occurs. Luce (1959) has proposed a model which resembles the serial comparison model of recognition just discussed. He suggests that the list of available alternatives is scanned exhaustively until the "correct" response is identified. Falmagne (1965) has attempted to incorporate in his model the idea of effects of previously performed actions on the present state of readiness of the organism. His model is thus a version of the progressive classification procedure discussed above, with response times to individual stimuli showing carryover effects from previous responses. There is in fact clear evidence of such "repetition" effects (Bertelson, 1961, 1963; Landauer, 1964).



The relative neglect of the response selection phase is unfortunate. There has existed a persistent suspicion that the empirical logarithmic relationship between size of stimulus-response ensemble and time of reaction was more an effect of response selection than of any other stage. There are two main reasons to argue in this way. First, a number of experiments have shown that where stimulus-response compatibility is sufficiently high (as in the case of the Leonard experiment discussed on p. II-37 above), the relationship disappears (Fitts and Seeger, 1953; Fitts and Deininger, 1954; Kay, 1955; Conrad, 1962; Alluisi, Strain, and Thurmond, 1964). The slope of function relating CRT to stimulus uncertainty is inversely related to the degree of compatibility. Since stimulus recognition is still required, it appears to be the stimulus-to-response stage which is affected.

A second important source of evidence concerns the effects of overlearning. It has been known for some time that practice also has the effect of reducing the slope of the reaction time - size of ensemble function, (Mowbray and Rhoades, 1959; Davis, Moray and Treisman, 1961; Neisser, Novack and Lazar, 1963; Egeth and Smith, 1965). Again practice, or learning, affects seem to have more to do with response selection than with stimulus identification.

Choice reaction experiments are one kind of a larger class of experimental types, which can be differentiated accordingly to the nature of the S-R connection, or the type of task. The following non-exhaustive list serves to illustrate:

- |                            |    |  |
|----------------------------|----|--|
| a. Overlearned tasks       | -- | S - R connection "automatic"   |
| b. CRT experiments         | -- | S - R connection known, but not yet perhaps "wired-in"   |
| c. Learning tasks          | -- | i) S - R connection known in the abstract, but not yet integrated                                  |
| d. Recognition tasks       | -- | ii) S - R connection unknown; learned by a conditioning schedule based on a one-one S - R mapping. |
| e. Concept formation tasks | -- | S - R connection unknown; principle of many-one mapping to be discovered and then learned.         |
| f. Gambling tasks          | -- | S - R connection is unstable (i.e., based on a principle of randomness) and hence unlearnable.     |

In each category the cognitive operation involved is different; the amount of central computing time required is different; and the possibility of an interaction between reaction time and number of stimuli is apparent.

This is an area in need of more intensive study.

The stage of response execution  
-----

Earlier a distinction was made between the principle of "selection" and "contiguity". Another way to think of the problem is to bear in mind that, regardless of the information in the original display, by the time the response execution stage is reached the alternatives have narrowed down to one. The function of the input phases is thus to narrow the information content of the incoming messages down to zero.

Response execution, on the other hand, is time-consuming: output consists of a concatenation of mini-activities, organized serially

to produce a coherent pattern of behavior. Each individual sub-action may have to be fitted into an organized sequence, in which subsequent elements of behavior cannot be initiated until previous steps have been completed. The whole sequence has to be monitored, which means that part of the input system has to be co-opted for purposes of obtaining feedback.

Output activities are particularly susceptible to disruptions associated with time uncertainty, and with the rate of presentation of signals.

Evidence of output limits having a physical basis can be found in an experiment by Quastler and Wulff (1955) which involved playing piano notes arranged in a random pattern. Here physical limits imposed by the necessity to move the hand in order to strike seem to have been an important factor in rate of information transmission. Up to about twenty keys (roughly two octaves), considerable gains in channel capacity occurred; for a range of 65 keys there were errors even at slow speeds of performance. The same experimenters found similar results using a typewriter. Up to about 16 keys, subjects could achieve about the same speed of output with comparable accuracy, but with 32 keys their performance was strikingly poorer. This result conforms roughly with the ordinary situation faced by a practiced typist, who is accustomed to using about twenty symbols with maximum frequency.

Nevertheless, Quastler and Wulff noted that their results could not have been due to either output constraints alone, and indeed theoretical interest has consistently centered not on physical constraints but rather on the role of central processing mechanisms. Craik (1948), for example,

in attempting to explain the intermittent-correction pattern of subjects, speculated that "if... the time-lag is caused by the building up of some single 'computing' process which then discharges down the motor nerves, we might expect that new sensory impulses entering the brain while this central computing process was going on would either disturb it or be hindered from disturbing it by some 'switching' system." Since the early experiments of Craik, numerous studies have confirmed the effect of prior signals on reaction times of signals which follow (Bertelson, 1960; Davis, 1967; Welford, 1967), and most explanations have posited a limited capacity central "channel" and some sort of queueing procedure.

Posner and Keele (1970) distinguish between the time required for an operation and the space required within some limited capacity central processing system. At one time it was considered (Hick, 1952; Fitts, 1954; Broadbent, 1958) that the processing of a signal pre-empted channel capacity for a time; more recently it has become evident that not all operations use up equal amounts of the limited capacity: as a result some operations can be carried on simultaneously without interference.

"Our current notions suggest that many mental operations which involve the access of external stimuli to long-term memory are orderly in terms of their time relations, but that these same mental operations do not seem to require space, that is, they need not interfere with other mental operations which must be performed at the same time unless the two have some specific incompatibility. Other task components require both time and space in the sense that they will interfere with virtually any other task which must be performed simultaneously." (Posner & Keele, 1970)

Encoding processes, for example, may be time-shared. In general

it is the processes which form the initial stages of a task which require little space. The processes which do require space are those performed on the retrieved products of memory search, in other words those occurring late in the tasks.

In somewhat similar vein, Moray (1967) has proposed a model of the brain, not as a limited capacity channel, but as a limited capacity central processor "whose organization can be flexibly altered by internal self-programming" (p. 85). The total capacity of the brain can be divided, and allocated in different ways, according to the task, or the phase of the task being performed: preprocessing, stimulus categorisation, response selection and execution. Parallel processing of certain activities may occur (the idea of a "slave" computer). In addition, Moray argues that the functions performed on the message themselves take up the capacity of the transmission system:

"The best analogy of which I can think is the relation between the data storage and program storage in a digital computer. If we want to perform a complex function on the data, one which requires elaborate programming, then we seem to have a smaller computer as far as the data is concerned. If you have a lot of data, then you are restricted to relatively small programs. What the analogy fails to bring out is that I think there can be a transfer of hardware, as it were, from the store registers to the arithmetic register and vice versa. This is unusual in computers, but the flexibility of the more or less 'universal neurones' in the brain may allow it".

If the model of Moray is valid, then we should think not only of information overload but also of program overload. We shall return to this question in the next chapter.

Summary

We began this chapter by asking whether the Shannon model of an ideal communication system was a useful tool for explaining human information-processing behavior. We discovered that the main utility of the model has been to isolate an interesting functional relationship between output and input, where the main input variable is amount of stimulus information, and the main output variable is the correlation between output responses and signals received. The function has the form of an inverted U, and has provided a means of defining a condition called "overload".

We discovered that the evidence was too chancy, and the model too crude, to remain satisfied with this level of explanation. We then considered a multi-stage serial processing model, and discussed problems of information selection and the allocation of attention. We discovered a number of things, such as the following: first, in one sense information overload is a universal condition, and hence uninteresting.

As Simon (1968) has observed: "Saturation with information is no new thing... the world is constantly drenching us with information through eyes and ears -- millions of bits per second, of which, according to the best evidence, we can handle only about 50". The critical process is selection and organization.

Secondly we accepted provisionally an "analysis-by-synthesis" model of information reception, in which processes of active stimulus - identification and memory-search constitute main phases. We found that memory restrictions are a key factor in limiting information processing, and probably explain some of the original findings on which the information overload hypothesis was based.



Finally, we discriminated between information reception and program organization, and proposed that much of what has been called "information overload" is more properly "program overload". Interference with the temporal sequencing of activities can produce serious disorganization of the organism's behavior, which justifies fully the term "overload".

#### Implications for further research

-----

Although this chapter has been devoted to a review of experimental literature, our overall point of view in this book is essentially constructivist. That is to say, we are interested in the application of knowledge and in particular in designing effective systems in a future society where exchanges of information figure larger in the picture as a whole, and where individuals function increasingly as information processors. Within this perspective we may ask what has been learned from our survey of literature.

Essentially, we have been working towards a framework for describing and analyzing information-processing behavior. By looking at experimental theory and research, we have begun to form ideas about what questions to ask when we turn to more naturalistic conditions. We have been alerted to the potential significance of certain factors. Let us see what some of these are. In order to do so, we should imagine that we are in the normal living and working conditions of an individual who, we may have reason to suspect, is a candidate for information, or program overload. Here are some of the questions we might be led to ask.

1. How many communications are received?

We start by making a time-budget for the individual. We ask what messages arrive, recognizing that what constitutes a "message" (a letter? phone call? TV program?) is already somewhat arbitrary. The arrival-rate of messages is specified.

2. What ensemble are the messages drawn from?

Now we start to look at the relative informativeness of messages received. Some individuals may receive numerous messages but on closer inspection it turns out that those which have to be attended to are highly redundant, and hence uninformative.

3. How predictable is the arrival of messages?

We saw that the inter-stimulus interval of messages is in itself an important source of uncertainty, whose effects we may want to study.

4. What is the overlap of messages?

A point of major importance in the study of program overload is the extent to which a second message "interferes" with an earlier message. Either the second message causes an interruption of the first, or has to be held in storage, in which case we ask:

5. What are the queuing procedures employed?

Particularly if a number of messages arrive during the "refractory" period (while he is still occupied), we want to know something about how these new messages are in turn accessed.

6. What is the complexity of the message?

By this we mean input complexity: how easily can the message be apprehended? We also wish to ask:

7. How complex is the task associated with the message ?

Every task can be described as a certain number of operations to be performed, taking more or less time, requiring more or less organization, implying more or less communication, etc. How well learned is the task?

8. What are the memory requirements?

Messages require the use of stored material, either in the brain of the receiver, or more likely, in files. We want to describe memory access, or retrieval, procedures.

9. What modalities of transmission are employed?

More than likely, messages are received in several forms: written, spoken, graphic. What are the prevailing modes of reception.

10. What are the deadlines?

Deadlines, unlike inter-message intervals, represent the imposition of external demand factors, which lend urgency to activities.

11. What is the value, or emotional loading, of messages?

We saw in the earlier review that time of response is a function of value. Can messages be ordered on this dimension?

12. How compatible are responses to messages?

Some messages can be reacted to on the spot; others require the initiation of complex responses.

13. How are messages selected for attention?

This question relates to the problem of agenda setting, and the allocation of conscious attention to problems.

14.. How are unwanted messages filtered out?

The business of filtering is similar to the previous process of selection, but logically distinct. How much can be effectively delegated, for example?

15. What are the adaptation strategies?

As message load is increased, do letters become shorter? or lunch-breaks? How does the individual cope?

16. What are the signs of stress?

What is the pathology of overload?

These questions are only some of the ones we might want to ask, arising out of our consideration of the literature. Our objective is to develop a set of profiles of information-processors in a variety of occupational fields. From this set, problem areas will emerge, leading eventually to the design of techniques to reduce the overload on the individual.

There is equally a need to continue, and expand, fundamental research. Let us identify one or two major areas which have not yet been fully probed.

First, we need to know a great deal more about human memory. No matter how elaborate the filing system, ordinary memory is still the line of first defense in information overload. We saw earlier that memory is also a major limiting factor in the reception of information. In spite of some very interesting recent work, the underlying principles of memory mechanisms are still very imperfectly understood.

Secondly, we need much more research on recoding procedures.

The peculiar genius of the human, and his ultimate weapon of adaptation, is his ability to reshape, re-interpret, re-conceptualize, find patterns in the flux of events that eventually permit him to live in environments to which he may not have seemed initially adapted. Hostile physical environments will require the same kind of assiduous research and planning that was needed to develop life systems for outer space.

Thirdly, we need to know more about the organization of activities: how decisions are taken, how programs are articulated, what kind of sequencing and takeout procedures apply in the human system, as contrasted with the computer. The program model is appealing intuitively; what is needed is now more experimental attention. In this respect, we need a better inventory of tasks and acts. We lack a language for coding output behavior in terms compatible with a cybernetic model.

Fourthly, we need very much to investigate how complex stimuli are received, in particular visual and iconic material. Most of the stimuli reported were rather simple in character; they said little about the potentialities for multi-model reception of messages.

One could add further questions. This is sufficient to indicate at least part of the road ahead.

CHAPTER III

COMMUNICATION OVERLOAD



In chapter 2, the empirical basis for two different models of organismic functioning was reviewed -- one a simple "channel" model, one a "program-execution" model. In this chapter, while some evidence of network overload will be considered, the major purpose is to extend the program-execution view of an organism and to show how it leads to a method of examining communication in networks. With this step accomplished, we will then turn to the question of designing networks which minimize communication overload.

In the first section of this chapter, we re-state some of the arguments already developed in the earlier discussion, but in a rather more formalized way. This will in turn lead us to a consideration in depth of the program model, particularly as propounded by Miller, Galanter, and Pribram. The implications of this model, and some neurological evidence for the model are examined. This is followed by a consideration of the role of symbolic representations of reality, which in turn leads into a discussion of communication. Two levels of communication, referential and relational, are specified, and the implications in terms of program load considered. Finally models of network design are considered, and the whole problem of overload re-stated in terms of the model developed.

### Limitations of the Concept of Transfer Function

Much of the literature of information overload has been couched, as we have seen, in terms of a model of the individual as a communication channel, which in input exhibits the irritability common to all living forms, but is otherwise the passive transmitting instrument of information-carrying events which impinge from the external environment. Its output reflects in appropriately transformed manner only these input events. Information overload in this perspective is uniquely a consequence of (1) features of the stimulating properties of the environment. When overloaded, the organism adopts a strategy of defense, shutting off the information stream at its source, leaking it, filtering it, and so on. The "strategies" are merely defense mechanisms and in no way affect the essential concept of the organism as a pipeline or channel between a given input stimulus and an output response, but a pipeline which carries information, rather than matter.

This model of the organism is exemplified in the concept of "transfer function". Experimentally realised, the transfer function implies the presentation of a stimulus (S) at time (t) and the observation of a response (R) at time (t+1). It is a transfer function if the response is a monotonic function of the stimulus, and of no other variable, i.e.,  $R_{t+1} = f(S_t)$ . In this section we will examine the utility of the concept of

---

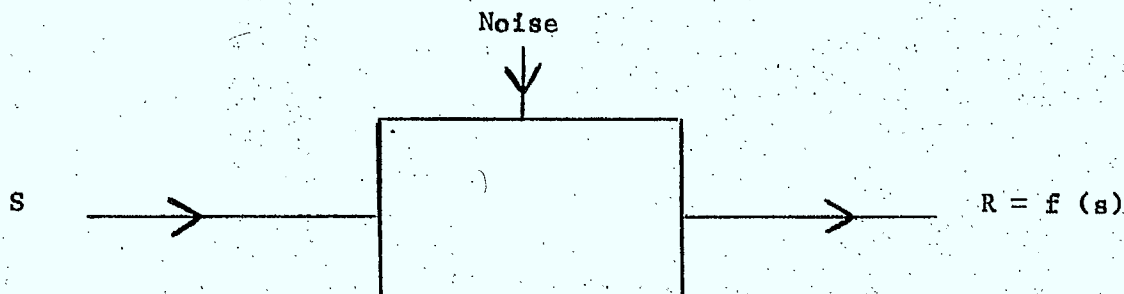
(1). Cf. Lipowski (1971): "By attractive stimuli are meant those which arouse appetitive and approach tendencies in people on whom they impinge. Overload implies excess or surfeit of such stimuli, in that they exceed the individual's capacity to process, choose, approach and consummate".

transfer function, and, in so doing, provide the basis for a different approach to the modeling of organismic functioning and information overload.

Biederman (1966), Bair (1971) have recently distinguished three types of input-output function, which correspond to three models of information-processing machines. "The functions performed by man on received information may be divided neatly into three categories of tasks, information, conservation, reduction, and creation, which subsume more specific functions labelled transforms".

#### Information Conservation

In an information conservation model, not only is the output assumed to be a direct function of the input, i.e.,  $R_{t+1} = f(S_t)$  but also the function is assumed to be reversible. That is, given an output, and knowledge of the transformation performed by the organism, the original stimulus input could be determined. The machine which corresponds to this model is subject to two major difficulties: random error (noise) and dropouts (omissions, in Miller's terminology). This model views man as essentially a re-coding machine, ultimately as a transducer. Graphically, the transfer function can be represented in the following way:



Experimental conditions which exemplify the information conversation task are described by J.G. Miller (1962):

"To test our individual subjects, we designed and built an Information Overload Testing Aid apparatus, which we refer to as an 'IOTA'. This is arranged to present stimuli to the subject on a ground-glass screen which is on a table in front of him. He responds by pushing the proper buttons. Stimuli are thrown on the back of a screen by a projector, a perceptoscope, which shows movie film at rates of from one to 24 frames per second. Our film presents black arrows on a white background, appearing in from one to eight of the eight two-inch vertical slate which run down the screen. There are 8 possible angular positions, like those of clock hands, which the arrows can assume. There are 8 corresponding buttons for each of the buttons being used...If an arrow in Position b appears in Slot 3, the only correct response is to push button b of the set for Slot3".

The transfer function, in much of the literature on information overload, is effectively the criterion function, since less than complete conservation of information is regarded as an unsuccessful response: the next class of functions to be discussed tends constitute, in this perspective, "defenses" adopted by the individual in the face of overstimulation.

#### Information Reduction

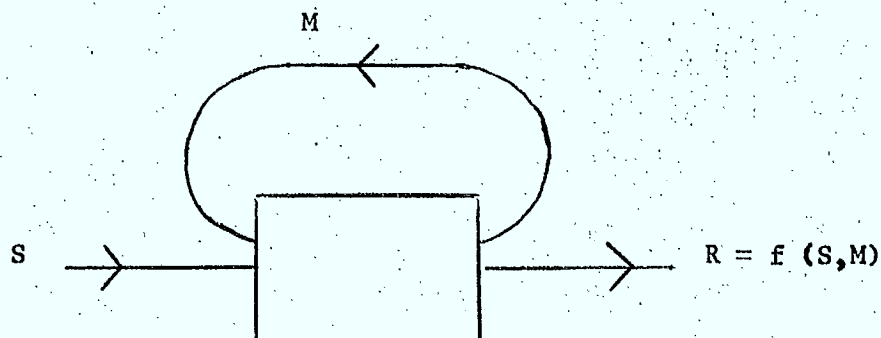
Information reduction machines produce a systematic loss of information, while at the same time maintaining essential features of the input. The associated functions are thus not in general reversible. Reduction models may be further classified according to the type of reduction involved:

## i) Filtering

The concept of filtering is that some types of input are systematically given higher priority than others, or that some inputs are consistently screened out, ignored, lost.

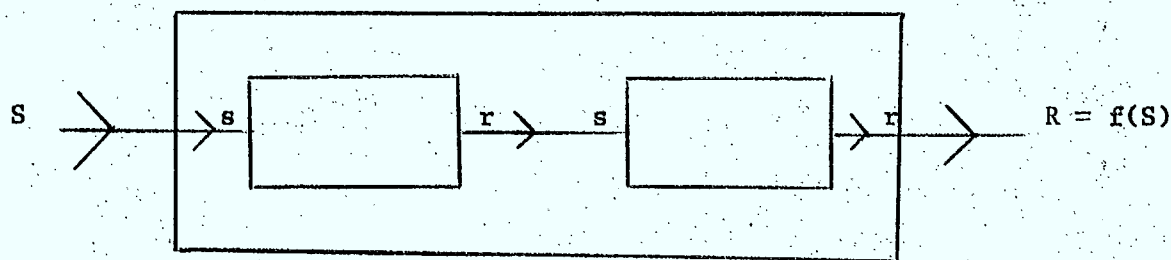
## ii) Condensation (Abstraction)

In condensation, none of the input information is ignored, as is the case for filtering. Rather the signal is processed to produce an output which represents the input but in reduced form. Arithmetical operations such as addition, subtraction, multiplication, division, are examples of condensation. Machines capable of condensing information require a memory capacity. First, the machine must store an algorithm or program which is capable of providing instructions concerning the steps involved in the reduction. Secondly, the input information constitutes data which frequently must be held in short term storage while awaiting processing. Graphically, we represent memory capacity by a self-terminating loop, as follows:



### iii) Contingent transformation (Sequential Processing)

The concept of contingent, or sequential, processing of information is similar to that of condensation; however, in contingent processing, it is further required that the output of one program serve as the input of a second. The relevance of this model for human concept formation has been discussed by Hunt (1962) and Biederman (1966). In addition to memory, such a machine includes a selector, or executive, component, which guarantees that the operations or subroutines are performed in the appropriate order. The analogy of a computer program has been widely employed to illuminate the processes involved in contingent transformation of information. Graphically, such multi-staged algorithmic transformations are represented as programs with programs contained within them:



### Information Creation

Bair (1971) discusses a transformation which he relates to information creation tasks and which he refers to as a "one to many mapping of stimuli resulting in a greater output than input".



A one-to-many "mapping" implies the existence of further unspecified variables. From Bair's discussion, it appears that two fundamentally different models are involved:

i) Information retrieval

The example given by Bair is the task of multiple word association, in which one stimulus word produces a chain of output responses. To explain this phenomenon, we require, in addition to the notion of a program, a long-term memory. Within the memory, data in the form of words are organized by a principle of association, so that the stimulus word serves as an entry point to the list. The output is then dictated by the program instructions (presumably including a "stop" rule), and by the organization of the stored list of words. The output is therefore no longer a function only of the input, but also of memory. Since memory in turn implies a previous process of learning, the new output is in effect a function of the present input, previous inputs, and (for reasons explained in the next section), previous outputs. This cannot be considered a "transfer" function.

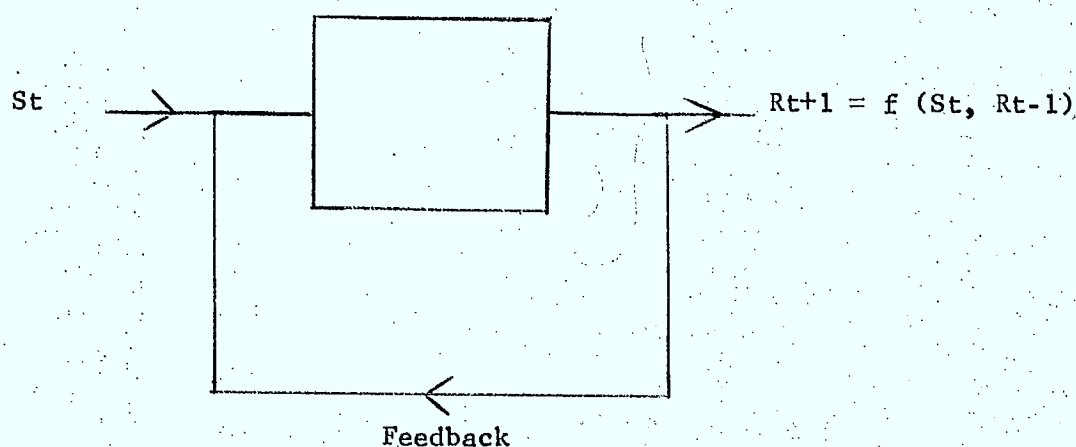
ii) Match-mismatch feedback

In a probabilistic learning task, the subject is required to decide for each trial in a continuous sequence of trials which of a set of events will occur. If the subject receives knowledge of the results (KR),

he will tend to use this information, which provides him with clues about the probabilities of the events, in order to guide his subsequent choice behavior.

(Hilgard and Bower, 1966; Posner, 1965; Schipper, 1967).

(i) and (ii) above differ principally in the memory requirements, and the immediate role of feedback in the process. The latter notion is illustrated thus:



This discussion reveals a fundamental difficulty inherent in much of the discussion on information overload. Whether one likes or dislikes the term "information creation" as a label for information retrieval and feedback, it is apparent that memory and learning are very nearly universal elements of human behavior, and a theory which is restricted to explaining behavior in which these fundamental processes are absent, is excessively limited in its application. Yet, as we saw earlier, much of the literature based on the channel model has pre-supposed that behavior could be viewed within the frame of reference of a transfer function. We have seen that this assumption fails to hold for information creation tasks. In these latter instances, since the output is a function

of variables other than the immediate input, the range of variation of information content of the stimulus input should be found to have less importance in determining the output, and this is in fact what Bair concludes from his (admittedly cursory) review of the literature:

"The generalization that task difficulty increases with increasing transmitted information has been shown to be not entirely applicable to information creation tasks. In information creation tasks, reaction time is more closely correlated with response uncertainty (variance) rather than transmitted information". (Bair, 1971)

The difficulty appears to be twofold: (a) insufficient attention has been paid to the function of stimuli for the organism, and (b) sources of information have been too narrowly defined with respect to the organism.

#### The purposive aspects of behavior

The notion of the human as an information channel seems to have blinded workers to the purposive aspects of behavior: for the most part, organisms do not passively receive stimuli; rather they actively seek out sensory information.

The point can be illustrated by reference to a simple example drawn from everyday experience. We can imagine a man walking down a long flight of stairs. For part of the way, the steps are quite broad, so that the walker has to take two steps to cross each step and one pace down to the next. At a certain point the stairs curve to the right, then straighten out, but at this point they become narrower, while the "lift" between each becomes higher. Eventually the walker reaches the street and turns left.

Suppose the man is buried in thought, so that most of the task occurs "automatically". If the steps are long, the walker quickly falls into a rhythm of two across and one down. New information is represented by a stimulus which indicates that he has to change from a two-and-one straight-down program to a two-and-one right-turning to a one-and-one straight down to a flat-surface left-turning to a flat-surface straight-ahead program. How are we to evaluate stimulus input information in this case? It would seem most useful to perceive visual information as constituting signals by which the full complex program is organized. Any explanation which emphasizes that the individual appears on observation to actively organize his head movements in order to "scan" his environment for the needed cues to change his behavior and hence suggests selective attention and information-seeking behavior is likely to be a more adequate explanation than one which begins by a description of the uncertainty in the environment. We could imagine of course that the set of stairs is a stimulus presented to the walker and his walking is merely the appropriate response (there are in fact certain advantages in conceptualizing it in this way), but in order to do so we should have to take account of the role of purpose in the resulting "experiment." There is no outside experimenter to set the criteria for successful performance of the experience; it is the walker's own objectives which establish the order of presentation of stimuli.

A re-examination of the question of information overload in the light of this discussion leads to the abandonment of a theory of "transfer function", and in turn to the abandonment of the communication channel concept. What the experiments using reaction time as a criterion effectively

demonstrate is essentially that a person can only run so fast down a set of unfamiliar steps in the fog. A fuller account of information overload, in our view, must take into account in addition the purposes of the individual, the programs he is executing, the importance of feedback and the role of perception in providing appropriate signals for the carrying out of his programs.

There is, as suggested earlier, a second difficulty related to the definition of environment (and hence source of stimulation) generally employed. The experimenter has generally been able to manipulate the conditions of the external environment. It tends to be overlooked however that there is a second environment, the internal environment. (Deese, 1967, p. 47; Miller, Ratliff and Hartline, 1961). Such internal environments have recently been shown to be in turn subject to experimental manipulation (Schacter, 1964). By the internal environment is meant the functioning of the organism itself, and the numerous internal processes by which the individual accomplishes homeostasis. The same information processing system which receives information concerning external events also receives messages from the internal environment. These latter messages serve in a certain way as an index of our success in responding to changes in the external world.

In this context, we may view the individual as being the recipient of two orders of feedback, one direct, one mediated by the dependence of the states of the internal environment on those of the external environment.

The conclusion to be drawn from this is that we may not safely in all

cases ignore the effect of variables other than those of the immediate stimulus configuration when evaluating the effects of rates of information on the performance of the individual. We must thus be sure how the subjects situates the stimulus within his image of the world.

### Plans and behavior

At this point we will outline a somewhat different approach from that exemplified by J.G. Miller and others. In this revised perspective, the role of plans in behavior will take on a greater importance.

Rather than visualize the organism as a channel for the transmission of information, we conceive him as exhibiting behavior which is guided or produced by a program, consisting perhaps of a collection of subroutines, which are stored, which can be occasionally modified and supplemented, and which are available to be called up by appropriate instruction from the main program. Such subroutines are in fact conceived as being triggered by specific stimuli.

Such a model of behavior has been shown to fit a variety of patterns of behavior observed among organisms less developed than man.

The larvae of barnacles will swim upwards towards the surface or downwards towards the bottom of the sea depending on its relative warmth or cold. This program could be represented as the following:

```

SUBROUTINE DEPTH:
  IF TEMP > K + A THEN CALL SUBROUTINE SWIMDOWN;
ELSE IF TEMP < K - A THEN CALL SUBROUTINE SWIMUP;
  RETURN
  STOP

```



Additional examples can easily be added: Ethologists have for example made detailed analyses of the behavior of the small fish called the stickleback which show the explicit cues required in order to set into motion an entire sequence of behavior. There is a term for such stimuli: "releasers".

(1)

A channel is a one way conductor . In this respect it shares something in common with the old notion of a reflex arc: stimulus → receptor → afferent nerve → connective fibers → efferent nerve → effector → response. The concept of an information processor as a collection of sub-routines organised by an executive routine is of a decidedly different order:

The neural mechanism involved in reflex action cannot be diagrammed as a simple reflex arc or even as a chain of stimulus-response connections. A much more complex kind of monitoring, or testing, is involved in reflex action than the classical reflex arc makes any provision for. The only conditions imposed upon the stimulus by the classical chain of elements are the criteria implicit in the thresholds of each element; if the distal stimulus is strong enough to surmount the thresholds all along the arc, then the response must occur... The threshold, however, is only one of many different ways that the input can be tested. Moreover, the response of the effector depends upon the outcome of the test and is most conveniently conceived as an effort to modify the outcome of the test. The action is initiated by an "incongruity" between the state of the organism and the state that is being tested for, and the action persists until the incongruity (i.e. the proximal stimulus) is removed.

- 
- (1) Cf. Miller (1962): Where the following set of stages of the channel are identifier: Boundary → input transducer → internal transducer → channels and nets → decoder → learner → memory → decider → encoder → motor or output transducer.

The general pattern of reflex action, therefore, is to test the input energies against some criteria established in the organism, to respond if the result of the test is to show an incongruity, and to continue to respond until the incongruity vanishes, at which time the reflex is terminated... Consequently the traditional concepts of stimulus and response must be redefined and reinterpreted to suit their new concept. Stimulus and response must be seen as phases of the organized, coordinated act... Because stimulus and response are correlative and contemporaneous, the stimulus processes must be thought of not as preceding the response but rather guiding it to a successful elimination of the incongruity. That is to say, stimulus and response must be considered as aspects of a feedback loop.

Miller, Galanter, Pribram  
(1960)

The organism, in such a model, is conceived to be continuously directed by a program in which control passes from one subroutine to another, and one from stage to stage within the subroutine. The organism performs operations (output) and makes tests (input). It utilizes sensory data in order to accomplish certain outcomes; it is stretching a point to consider this as the transmission of information in the sense of a channel, although acting like a channel is one possible task which the organism can undertake.

#### The efferent control of sensory input

The real disadvantage of a communication or transfer function model is that it takes insufficient account of the fact that man is a general purpose machine, capable of performing many kinds of activity in a variety of environments (including, when called upon, acting as a subject in a choice reaction experiment). If man is merely a signal transmission system, then he is activated when signals are presented to him. His behavior depends on the nature of the stimulus field in which he finds himself;

we would not suppose him to look for stimulation. The advantage of a "program" model is that it lends itself to the explanation of adaptive behavior. However, it also implies some re-orientation of perspective with respect to the inputting of information. This topic is considered in this section.

If the assumption is made that all behavior is guided by a program, or plan, than it should follow that there is a set of choice points, where the presence or absence of a certain indicator determines the choice of the next sequence of behavior. The thesis of central direction of behavior, when contrasted with that of man as a communication channel, requires a different interpretation of the choice reaction experiment. In the latter, the subject faces a display panel on which an event is to occur (perhaps the appearance of a light, varying, let us say from red to yellow to green) and a control panel with parts he is required to operate (suppose three buttons worked "R", "Y" and "G"). He is informed that each button matches up with a single light, and that any time a light appears, he must hit the correct button, in the shortest possible time. Other possible stimuli are now irrelevant to the assigned program: the sound of a distant siren, the color of the experimenter's tie, the clock on the wall. Attention becomes riveted on the display panel, and the motor mechanism is arranged for optimal response. The accomplishment of the task (or completion of the program) requires information. When a light flashes on, the test is performed: "Red? Yellow? Green?" and depending on the result execution follows. After verification to determine that the program has in fact been completed, control passes back to stage one, ready for the next run of the program.

The concept of stimulus is defined by the functional role of information within the program: it is not objectively determinable by reference to external criteria. The experimenter is reduced to inferring the nature of the stimulus from his observations of events in the environment of the organism and the organism's subsequent behavior; whatever stimulation is potentially available, the "stimulus" is what the organism responds to.

From this we should be led to suspect that stimulus input mechanisms, like motor output, are subject to central control. What evidence is there for this assumption?

In addition to the evidence presented earlier in the discussion of iconic memory and pre-attentive processes, some physiological evidence can be adduced which supports the theory of a measure of efferent control over input. Jung, Creutzfeldt, and Grüsser (1957), Creutzfeldt and Grüsser (1959), Jung (1958) have demonstrated that stimulation of the thalamic region of the brain stem may alter the critical flicker fusion of cortical neurones. The nonspecific thalamic nuclei and the reticular formation are usually considered to be the mechanisms which control arousal and attention (French, 1957) and are the most likely candidates for the role of the control mechanism which was hypothesized in the preceding discussion (Wooldridge, 1963). French (1957) performed an experiment which demonstrated the control of reflex motor reactions by the reticular formation. The degree of a response of an anesthetized monkey to knee taps was recorded: it was shown that activation of the reticular formation affected the intensity of the response.

French was led to conclude from a review of the evidence that "these centers can enhance or inhibit sensory as well as motor impulses. In short, the RAS (Reticular Activating System) acts as a kind of traffic control center, facilitating or inhibiting the flow of signals in the nervous system."

"The astonishing generality of the RAS gives us a new outlook on the nervous system. Neurologists have tended to think of the nervous system as a collection of more or less separate circuits, each doing a particular job. It now appears that the system is much more closely integrated than had been thought. This should hardly surprise us. A simple organism such as the amoeba reacts with totality toward stimuli: the whole cell is occupied in the act of finding, engulfing and digesting food. Man, even with his 10 billion nerve cells, is not radically different. He must focus his sensory and motor systems on the problem in hand, and for this he obviously must be equipped with some integrating machine.

"The RAS seems to be such a machine. It awakens the brain to consciousness and keeps it alert; it directs the traffic of messages in the nervous system; it monitors the myriads of stimuli that beat upon our senses, accepting what we need to perceive and rejecting what is irrelevant; it tempers and refines our muscular activity and bodily movements. We can go even further and say that it contributes in an important way to the highest mental processes - the focusing of attention, introspection and doubtless  
(1)  
all forms of reasoning".

---

(1) Cf. also Dixon (1972) who identifies the RAS as the system which provides for conscious representation of information.

Granit (1955) found evidence that activation of the reticular formation was capable of causing potentiation or inhibition of photocally induced activity in retinal cells, indicating centrifugal effects. Hernandez-Peon, Scherrer and Velasco (1956), in a famous experiment, determined that activation of the brain stem area was able to depress afferent conduction at the lateral geniculate body in the visual pathways, and hence to produce a reduction in sensory impulses to the visual cortical receiving area. Their findings indicated that the effect was due to a true inhibitory influence from the brain stem reticular formation. This influence was modality-specific. They concluded: "It appears that this effect is exerted by inhibitory centrifugal fibers to the retina, and that their functional role, therefore, is to block sensory impulses during attention, preventing them from entering the brain, and from interfering with the neural mechanism of integration occurring during that physiological situation."

While these experiments support a theory based on the concept of a central control mechanism which directs the attention of the organism, selects response programs, and supervises stimulus input, it does not state on what basis reticular activation is brought into play, or how perception occurs.

#### Pre-attentive processes

Neisser, it was seen earlier, proposes a two-step model of perception: in a first step, the organism is alerted and a rough general picture formed; in a second, the information is read out in greater detail. Sokolov (1960) has discussed in some detail the experimental evidence for the mechanism termed by Pavlov an "orienting reflex".



This reflex is non-specific; it occurs as a result of any increase, decrease, or qualitative change of a stimulus; and it produces as well a primary non-specific response. The orienting reflex "is evoked when the neuronal model set up in the brain does not coincide with all the parameters of the stimulus." The stimulus might be a sound, cold, a shock: "The orienting reflex is produced not only by the stimulation itself, but by impulses arising as a result of non coincidence between a certain cortical pattern (the model) and the applied stimulation."

In addition to the generalized orienting reflex, Sokolov identifies a localized orienting reflex, which is modality-specific. The function of this mechanism is to increase the discriminatory power of analysers, as a result of direct stimulation through descending pathways to receptors from the reticular formation and the cortex.

A theory having similar elements to that of Sokolov has been advanced by Melzack and Wall (1965) to explain the findings concerning the experience of the sensation of pain. While the full details of their theory are not relevant here, the following conclusion is pertinent: "It is now firmly established that stimulation of the brain activates descending efferent fibers which can influence afferent conduction at the earliest synaptic levels of the somesthetic system. Thus it is possible for central nervous system activities subserving attention, emotion, and memories of prior experience to exert control over the sensory input. There is evidence to suggest that these central influences are mediated through a gate control system."<sup>(1)</sup>

---

(1) The authors propose a model of such a system based on rather complex feedback mechanisms.

"The manner in which the appropriate central activities are triggered into action presents a problem. While some central activities, such as anxiety or excitement, may open or close the gate for all inputs at any site on the body, others obviously involve selective, localized gate activity. Men wounded in battle may feel little pain from the wound but may complain bitterly about an inept vein puncture...The signals, then, must be identified, evaluated in terms of prior conditioning, localized, and inhibited before the action system is activated. We propose, therefore, that there exists in the nervous system a mechanism, which we call the central trigger, that activates the particular, selective brain processes that exert control over the sensory input."

The authors then note that certain pathways projecting in the brain stem and thalamus are extremely fast, and that messages arriving on these pathways could activate selective brain processes to receive subsequent afferent volleys arriving over more slowly conducting fibers.

Each of these theories are built around the role of match-mismatch error signals indicating variety in the environment which must be attended to. We must now explain how the subsequent information is analyzed.

#### Perceptual analysers

Perception is not determined simply by the stimulus patterns; rather it is a dynamic searching for the best interpretation of the available data... It seems clear that perception involves going beyond the immediately given data of the senses; this evidence is assessed on many grounds and generally we make the best bet, and see things more or less correctly. But the senses do not give us a picture of the world directly; rather they provide evidence for checking hypotheses about what lies before us. Indeed, we may say that a perceived object is a hypothesis, suggested and tested by sensory data.

The process of perception will be illustrated here only with respect to the visual channel. While the detailed processes of perception are very different for other channels, we assume here that at higher levels, much the same principles apply.

The retina is composed of rods and cones, the rods being connected in large groups to secondary nerve fiber conductors, the cones being connected to fewer individual nerve fibers. The experiments of Hübner and Wiesel (1962) conclusively demonstrated that the retina functions essentially as a pattern recognizing device. Single cells in the visual area of a cat's brain proved to respond only to certain patterns of stimulus on the retina. A bar of light would stimulate a given cell only when presented at a certain angle; for other angles the cell remained silent. Different cells respond to different angles. The general principle which these results illustrate had been stated as early as 1942 by Lashley: "The principle involved is that the reaction is determined by relations subsisting within the stimulus complex and not by association of a reaction with any definite group of receptor cells." This accounts for the fact that we see the same object even though its image happens to fall on a different part of the retina.

Cells which are deeper in the brain in turn respond only to more generalised characteristics. (Hübner and Wiesel, 1962). We are led to view perception, from this evidence, as a process of **identification** of dimensions of stimuli, of increasing generality as higher order mental processes are involved, and information is integrated from additional sensory channels.

Visual impressions, finally, "consist of organized objects, seen against a less coherent background. Discriminative reactions, when analyzed, are found to be based upon certain generalized features of the stimulus." (Lashley, 1942).

The primary task of perception in this view is identification and classification. This suggests an explanation for our earlier observation that although the senses provide an overwhelming ensemble of information, the organism as a whole seems to transmit little. The point is made by Morrell (1967): "...Information is processed in parallel in thousands of cells so that the organism need not depend on the reliability of any single element for identification of an experience. These parallel chains need not all carry exactly the same information and, strictly speaking, therefore may not necessarily be redundant. It is only necessary that the nervous system receive enough information about an experience to identify it even if some aspects are left out or distorted. Furthermore, it is likely that on first exposure to a stimulus, the nervous system specifies it less precisely than after many exposures. Ultimately, the code must be transformed from one based upon a discharge pattern through time to one that is more stable, i.e. immune to electrical interference, more disseminated, and susceptible of very much faster read-out."

The point we are making is nowhere better illustrated than by reference to the classical question of why the world remains apparently stable when we move our eyes - why we do not experience the "swish-pan" effect of a film or television camera system which also depends on an optic system similar in some ways to that of the eye.

Evidence now appears to clearly support an outflow theory of Helmholtz, which states that command signals flowing outward to the eye/head effector system are monitored by an internal loop in the brain and fed into the analysis process in order to correct for head movements and thus retain stability of image (Gregory, 1966).

The importance of active analyzing processes becomes even more salient when we turn to the question of recoding of perceptual information into symbolic forms.

#### Symbolic representation of stimuli

We have proposed that the process of perception includes a recoding, or "read-out", component; the necessity to suppose the existence of a read-out mechanism becomes peculiarly evident when we turn to the question of how perceptual data gets to be represented in symbolic (above all, linguistic) form. The simplest explanation - that each symbol of the language becomes, through conditioning, associated with certain stimuli and thereafter functions as a sign of the original significate - was effectively shown to be in certain respects grossly inadequate by Chomsky in his 1957 review of Skinner's Verbal Behavior. The theoretical difficulty is that the grammar of a natural language is capable of generating an infinite number of syntactically well-formed sentences, and indeed in ordinary discourse "new" sentences are constantly being produced. If meaning were the result only of conditioning, we ought to experience great difficulty in understanding new sentences. The fact is however that we are often presented with novel sentences, and may have no trouble in comprehending them.

Furthermore, the process of human thinking, by universal experience, involves the manipulation of symbols to arrive at conclusions which are not derived directly from empirical evidence, although they may subsequently be so tested. S - R theories of language provide relatively poor explanations for such a process.

A "read-out" model assumes that incoming sensory data are recognized as comprising one or another pertinent pattern. Thinking about the world requires first that sensory impressions be mapped into a domain of symbolic forms or images, such that, ideally, a one-one correspondance can be supposed to hold between the symbolic image of the world and the world itself. All inputs originating as non-symbolic events may be thought of as transformable into symbolic equivalents. If the image, or symbolic model of the world, is explicitly linguistic, then it consists of an ensemble of sentences.

The problem which the individual faces is to keep his model up to date. How is this accomplished? First, we assume that he receives indication of a mismatch between his existing image and the real state of the world (perhaps via the orienting reflex discussed earlier). To alter the model, a new sentence, or sentences, must be generated. These may in turn be verified by comparing them with available sensory information. We should have to suppose that sentences are generated sequentially, tested, and depending on the sign of the test, either the model is adapted, new sentences are generated and/or additional sensory information is sought. Another way to express this process is to describe it as hypothesis - testing.



Since the number of possible sentences is infinite, the process is in principle nonterminal. Presumably however other mechanisms come into play to limit the process.

How are hypotheses confirmed and disconfirmed? The sentence which is generated will have two logically distinct components: a referent or referents and a predicate which attributes something about the referent (s) or states a relationship between referents. The referent may be an explicit object (s) or class (es) of objects. In the sentence "That rose is red", the referent is "that rose" and the predicate is "red". In order to verify such a statement, two steps are required: first, a set of measures will have to be decided on a priori for the concepts rose and red. In their simplest form, such measures are simply categories; in other cases, a scale may be implicated. The consequence for the organism is that before encountering sensory stimuli, he has available a set of classes into which we can order observations. Such an assumption is consistent with physiological evidence indicating that brain analyzers are activated before other stimuli appear. Second, upon encountering sensory impressions, a decision must then be made concerning how to class the sensory impressions. Objects are discriminated. There are thus two phases:-application of measures and classification of observations.

When the measuring and classifying task has been completed, one further step remains, to compare the data against the original hypothesis: the observations obtained are not data until they are made to serve the function of mismatch or error signals (in the sense of hypothesis-confirmation and disconfirmation).

The complete process from hypothesis generation to data analysis is regarded as a feedforward process: perception, in this model, consists not of passive reception of stimuli, but active obtaining of feedback signals serving to modify a pre-existing image.

This approach assumes that perception resembles motor activity, with the difference that to effect motor events, instructions rather than hypotheses are generated ("Cut the red rose" rather than "That is a red rose"). The instructions are broken down into separate actions, the changes of state in the effectors are signaled by means of proprioceptive channels and evaluated as mismatch or error signals. The process is illustrated in Figure - 111-1

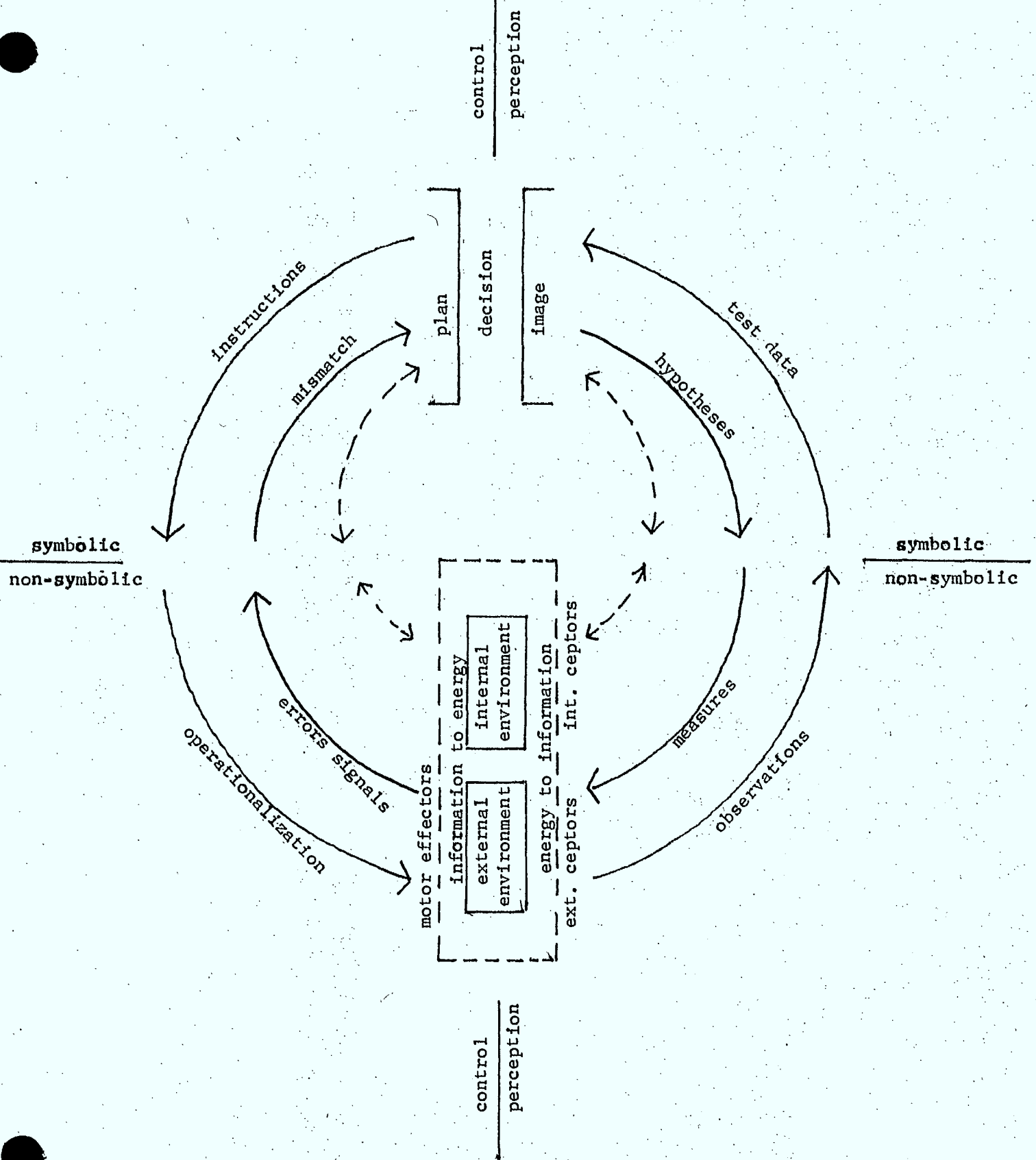
#### Feedback and information

Within the approach outlined above, the role of feedback has been given a central place; it is seen to be essential to perception, to the effecting of motor activity, and to the control of behavior itself. At this point, it is pertinent to enquire what revisions in our conceptualization of information are required within this altered perspective.

The subject in a task situation is in a state of uncertainty, first as to the nature of the stimulus which is to be presented to him, secondly as to the nature of his response, thirdly as to whether his response was "correct", in the sense that it reduced mismatch signals to within acceptable limits. In the latter case, the task has been accomplished.

It was seen earlier that the amount of information contained in the input signal is identical with the reduction of uncertainty (and hence depends on the ensemble from which the input signal was drawn).

## Program model of the Human Information Processing System



Feedback provides knowledge of the results (KR) of his response. It reduces uncertainty concerning the outcome of his response. The information value of the KR, or feedback, depends on the number of kinds of KR which could have been sent to the subject, i.e. on the variety in the KR message ensemble.

We may define two types of KR: (a) intrinsic KR, which is feedback that is either the result of proprioceptively available information (muscle stretch cues are an example), or is usually present to the individual in the performance of a particular task (as for example in steering tasks, where visual information can be used to supplement kinesthetic cues); and (b) extrinsic, or augmented, KR, which is present only when an additional feedback loop to those usually present is found. This latter type of KR is peculiarly susceptible to experimenter manipulation: it may in many cases simply consist of the experimenter (or his stooge) informing the subject how he has done. Augmented KR provides a means to "train" a subject to perform a task according to certain criteria: the subject should be expected to continue to modify his behavior until he has eliminated or reduced the mismatch signals - in this sense the application of extrinsic KR is equivalent to that a reinforcement schedule.

This distinction permits a further. We may discriminate between two types of experiment, according to the role played in each by augmented KR: (a) skilled performance, and (b) concept formation experiments.

In skilled performance experiments, the criterion for satisfactory performance of the experimental task is unambiguous. In an experiment of Trewbridge and Cason (1932), four groups of subjects were required to draw 100 lines all of a specified length.

After each attempt, the experimenter provided augmented KR of three kinds: (1) spoken nonsense syllables, (2) right/wrong messages, and (3) magnitude of error messages. In the control condition, no KR was provided. The results indicated the importance of information to the subjects: in condition (1) and the control condition, there was no improvement of performance; in conditions (2) and (3) improvement occurred, with maximum learning in the case of (3).

The choice reaction experiments to which the hypothesis of information overload has been most frequently related are based upon, for the most part, skilled performance tasks. The effects of augmented feedback do not in these researches appear to have received a great deal of attention. The application of extrinsic feedback by the experimenter, in order to "teach" the subject a standard of performance, violates the concept of a transfer function. The subject must learn to discriminate the appropriate stimuli, understand which response can be acceptably associated with each stimulus, and finally learn to perform the correct response. In concept formation experiments the learning of the rules governing the appropriate correlation between stimulus and response are made much more complex. With respect to this type of experiment, it will be seen, it is more appropriate to ask how much information is used, or must be used, to complete a single correct response.

#### Hypothesis testing and the formation of concepts

"One might speculate that, in the adult human subject, any task that leads to the search for a rule, for example one that relates the subject's response and the experimenter's 'reinforcement' must first produce the specification of the rule before the rule can be applied... There is now adequate evidence that the adult human organism will usually generate rules or hypotheses whenever the environment demands some consistency in behavior".

Mandler, 1964.

"The number of ways in which an array of events can be differentiated into classes will vary with the ability of an organism to abstract features which some of the events share and others do not ... Categorization at the perceptual level consists of the process of identification, literally an act of placing a stimulus input by virtue of its defining attributes into a certain class ... By categorizing as equivalent discriminably different things, the organism reduces the complexity of its environment ... To know by virtue of discriminable defining attributes and without need for further direct test ... is to know in advance about appropriate and inappropriate actions to be taken.

Bruner et al. 1956

Experiments in concept-learning are concerned with human being's use of information to learn or recognize "patterns." In a series of experiments conducted during the fifties, Bruner, Goodnow and Austin (1956) investigated some of the conditions under which subjects were able to attain new concepts. In a concept attainment experiment, the subject is required to discover a principle of grouping stimuli into equivalence classes: he must determine the intrinsic attribute properties which serve to characterize members of a given class. Experimentally two main elements are required for the performance of the task: a) a sequence of instances, consisting of pictures of objects, geometric patterns or even words, each instance being characterized by a set of attributes - - geometrical figures varying in size, shape, color, number, orientation, etc, which in another experimental situation would be simply referred to as stimuli. Subjects are then required to produce a single response to a stimulus set according to the desired attribute dimensions. (For example, "sort instances of all red forms regardless of shapes as opposed to other-colored forms."). b) The second element consists



of knowledge of the results of his attempts at identification (KR), or validation, for each presentation of an instance. The subject must be able to assume an underlying pattern in the sample of stimuli presented to him, and accurate feedback on the results of his guesses. Each attempt at identification followed by validation provides the subject with information, since it constitutes a test which limits the number of attributes the subject has to take into account in attaining the concept.

There are two main types of concept formation experiment, depending on the method of presentation of stimuli, and the freedom of the subject to control his feedback. Those two types of experiment may be termed (a) array sorting, and (b) serial sorting experiments.

In array sorting, subjects are presented with any array of instances and are required to sort them into groups.

In serial sorting, instances are presented in sequence, either pre-determined or random, and subjects are asked to make a placement of them in groups according to desired stimulus attributes.

In array sorting, the subject receives feedback about choices he makes as to whether he is right or wrong, but he is also free to select the next instances because the array is displayed before him. In principle this allows him to choose instances so that he can maximize the information in the KR.

In serial sorting, the subject is shown one instance of the concept and must choose whether or not it is part of the concept; then he is told whether

he is right or wrong. Here, the experimenter can control the next instance in the sequence and therefore the subject cannot maximize information received from knowledge of results.

The importance of a "program" model of information processing is more immediately evident when we turn to look at concept formation experiments than for skilled performance experiments, in part because the individual is required by the design of the experiment to adopt a more active exploratory role. The description of the "strategies" employed by subjects given by Bruner and his associates conforms very well to the model of subroutines: subjects tend to follow well-defined strategies, "conservative focusing", "focus gambling", etc. The amount of information to be gained from a stimulus depends, in part, on the subjects previous choice. This considerably extends the concept of the choice reaction experiment, and allows us to place it in a different perspective; as a subclass of sequential choice experiments which do not encourage hypothesis-formation.

Bruner and his associates found, among other things, that the choice of strategies varied systematically as a function of the informational, strain, and risk characteristics of the problems. Information was varied by both amount provided and by the form (positive versus negative instances). Cognitive strain was varied in several ways, of which the one most salient to our earlier discussion was a stepping up of the pace of presentation. Under conditions of moderate information and low cognitive stress, subjects followed

a program certain to provide a certain amount of information per choice. Under no time pressure, this "strategy" guarantees eventual success. When time and other limitations were imposed, it was found that subjects tended to shift to a "strategy" in which, if they were lucky, they could obtain much information quickly, but which involved considerable risk. It was also found that under pressure of time subjects tended to fall back on cues that seemed in the past to have been useful, cues that were most easily available or most easily discriminable. Finally, they noted that under accelerated time pressure, subjects began to cast about in search of almost any available piece of information, even though the result was to overwhelm their limited information-carrying capacity. It appears that with increasing stress, the strategies show a tendency to degenerate from orderly systematic search towards random search.

A further limitation discovered by Bruner, Goodnow and Austin can be traced to limitations of memory. The subject is required to store the results of previous positive and negative choices. It is for this reason that Bruner et al. hypothesize that "negative" instances tend to be underutilized by subjects: a greater memory strain is involved.

In other research, Schroder, Driver and Street (1967) used a war game simulation which required subjects to integrate available information in order to issue a series of commands resulting in the deployment of their forces on an imaginary island against a similar enemy force. The experimenters varied the amount of information presented, input rate, and

the proportion of results called positive and negative. They found that increases in the rate of presentation of information at first improved the information processing performance of subjects, but that beyond a certain point severe decrements occurred. They also found, like Bruner et al, that negative instances depress performance. Under conditions of low stress, subjects used more dimensions of information than under high stress, when judgments tended to take on a black and white cast, stereotyped thinking become evident, and complexity of integrations declined.

The results reported by Schroder et al. support the view that under conditions of stress, it is the nature of the program-utilized by the information processor which is affected.

#### The internal environment

In the previous discussion it was shown that KR, knowledge of the results of one's choices, may be a source of information which is as important as stimulus information. In this section, we look at a further neglected domain, that of interoceptive stimulation.

The role of the reticular formation, and in particular of the thalamus, in the control of behavior in general, and the execution of many simple program of behavior, has already been noted. Adjacent to the thalamus in the brain stem is the hypothalamus, which is an important center for the control and regulation of visceral processes of the body, body temperature, and the glandular system. Hess (1957) found that stimulation of cells in the hypothalamus affected rate and depth of breathing, blood pressure, heart rate, and caused vomiting and body elimination. Appetite is apparently con-

trolled by the hypothalamus: destruction of a part of this region will prevent an animal from eating no matter how hungry, or will cause it to keep on eating, no matter how satiated. In addition, emotion can be aroused in an animal by stimulation of hypothalamus sites, fear, hostility, rage. Additional work by Olds (1956), Brody, (1958) and others demonstrated the existence of pleasure and punishment centers in the brain, stimulation of which produced evidence of hunger reward, sexual reward, intense pain - in the absence of other external stimulation.

Since the body for its continued functioning, requires the maintenance of homeostasis within a great variety of subsystems (Cannon, 1932), it is not surprising that the system responsible for monitoring and controlling activities of the body should be closely associated with the program-directing mechanism of the body responsible for activity upon the external environment. The organism remains informed about states of the environment in two ways: (1) directly from the exteroceptive system, and (2) indirectly, as a result of changes occurring within the internal environment which can be traced to changes in the external environment. We are becoming accustomed, for example, to measuring pollution as much by its effects on our internal good health, as by smell and tastes which often does not provide good information.<sup>1</sup> The state of the internal environment is of primary importance, since if its continued efficient functioning cannot be assured, the existence of the orga-

---

1. An ingenious use of this fact is illustrated in the experiments of Schacter (1964). Schacter induced subjects under another pretext to take epinephrine, which produces symptoms of palpitation, tremors, accelerated breathing. Schacter found that subjects labeled their emotion by reference to conditions in their external environment.

nism is threatened. The state of the internal environment is a criterion by which we sometimes evaluate states of the external environment. Beside beliefs stand feelings, beside intentions, wants.

Most choice reaction experiments have assumed the states of the internal environment to be constant; however, there is at least some evidence in the Schroder et al. experiments to indicate that effects of emotion cannot properly be disregarded. These experimenters found effects due to what they termed "noxiety" and "eucity", roughly unpleasant and pleasant environmental reward. This suggests that the monitoring of changes of state, resulting in adoption of program change, and the evaluation of possible effects, is not restricted to those occurring in the external environment; the internal environment is also implicated.



The role of communication in the information processing system.<sup>1</sup>

The effect of communication, the consequence of symbolic interchange, is to link information processing systems. We ask people to tell us what they see, and act on what they tell us. We request opinions and advice. We issue instructions and others carry them out; we commit ourselves to certain types of program because of our relationships with others. Someone has access to direct experience of the external environment, but it may not be us. We live, in effect, a vicarious experience.

Let us restate this point slightly.

Every human needs information and advice on what program to carry out. Every individual is susceptible to information management and direct methods of control, and hence to the domination, or manipulation by others, and ultimately to the accomplishing of the objectives of others, and the neglect of those of self. Each individual has to measure the implications of others' messages in terms of his own self-interests. Social roles can be defined by the messages appropriate to them: it follows that every message either reinforces or changes a role relationship. There is in all communication relations this built-in tension. The use of the word "tension" in the sentence above is equivalent in part to saying that there

---

1 The importance of communication has been seriously neglected in the literature on information overload with unfortunate results in limiting the generalization of laboratory results. Within the "wired nation" the increasing importance of symbol exchange may be confidently predicted. The individuals who will be subject to information overload in that society more likely than not will have had their information passed to them by someone else. The effects of this interdependence do not appear to have been examined experimentally. One index of this dependence seems the wirespread and apparently increasing phenomenon of viewer suspicion of the news services.

is a further element of uncertainty (and hence another source of information) which is a function of being in a linked set of information-processing systems. There are in fact two new sources of information of which we must take account.

Every linguistically-encoded message contains information on two different levels: (a) referential (b) relational information. First, a sentence has some propositional content which refers to an external world. The content of sentences may be described as falling in the category either of (a) reports, (b) commands. Reports are due to the operation of processes of perception; commands are intended to result in the effecting of a plan. Reports refer to a world which already exists; commands to a world which does not yet exist, except in imagination. Images model perceived worlds, and projected, or possible worlds.

The external environment is something we assume to be directly given to all of us through our respective sensory systems: one main function of language is to refer, or point to, things in the common environment. By contrast, our internal environments are our own, and only we have immediate access to them. For each of us our basic knowledge consists not only of beliefs about the state of the external environment, including our explanations of its dynamics, but also of our feelings which represent the states of our internal environment. Not only "There are roses in my garden" but also "I love the smell of roses". The common information system which results from communication in a dyad has one general external environment, (with however two images of it) but two internal environments. Let us represent the latter as a set of ordered couples. (Newcomb, 1953)

Then this set defines a relationship. Every communication potentially carries relational information: it may specify whether first the two believe the same things, and second have the same objectives, want, intend, feel the same things. Such information is never fully attainable by direct sensory experience: it can be arrived at only by inference.

#### Problems of the determination of referential information content

The fact that messages can convey information (about events in an external environment) has to do with the existence of a mapping between the symbols of the message and the elements, states or events of that world. There must exist some equivalence relation between the two domains of message and world. In the discussion of chapters 1 and 2 however the measure of information was a function only of relations between the symbols themselves, that is to say of internal constraints (associated with contingent probabilities, for example).

Garner (1962) has distinguished between the significance of a message, and its structure. The significance of a word is "the particular specifying or indicative relation for any single event or symbol". Significance may be either external or internal. The external significance of a word is what it refers to: the significance of the word "tree" is the object it stands for, or some other symbol of the object (such as a picture, or a word in another language meaning the same thing). The internal significance of a symbol is its set of associations, i.e., the other symbols which it brings to mind.

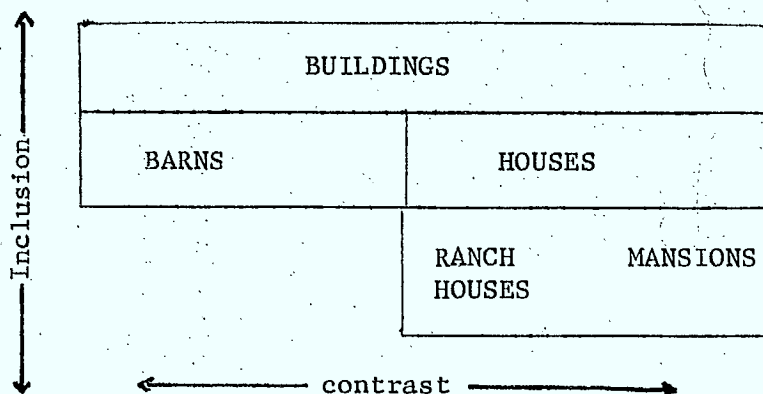
Structure may equally be partitioned into internal structure and external structure. By the word "structure" is meant "the totality of the relations between events". As relations between events or symbols, structure is amenable to quantification, using an information-theoretic metric. Internal structure measures the amount of relationship between the elements of the message itself, the pattern or formal constraint. External structure is measured by the totality of the relations between the elements of the symbol event and events of the external environment, "as long as there is a high correlation between the two sets of events". (The high correlation is necessary to assure that a specifying, or indicating relation exists, that is to say that the information in the message continues to reflect uncertainty in the environment, rather than in the symbol-events themselves independent of their external referents). For example, the meaning of a radarscope comes from what it reveals about the structure of events in the outside world, for example the movement of aircraft.

#### The internal structure of the symbol system

-----

The most important symbol system is language. The basis of language is a set of names, which in turn refer to classes. Underlying all of our linguistic behavior, accordingly, is a classification system, which constitutes the internal structure of language. The classes themselves are ordered into larger groupings, hierarchically, called "taxonomies". Two principles are employed: inclusion, and contrast. Examples of taxonomies are easy to find: for example we include both "houses" and "barns" under the heading of buildings, but in turn we distinguish

between "ranch houses" and "mansions".



The principle of inclusion works vertically, from the most fine discriminated level of categorization, to the most inclusive. The principle of contrast determines how many discriminations are made at each level.

The principles of contrast and inclusion depend on the operation of semantic features, or components of meaning. (Goodenough, 1958 ). To give only one brief example, we discriminate between a "bull" and a "cow", a "steer", a "bullock" and a "heifer". To make these discriminations we employ two basic semantic dimensions, sex and age. "Cows", "bulls" and "bullocks" are adult, "steers" and "heifers" are not fully grown. "Cows" and "heifers" are female, "bulls" male, "bullocks" and "steers" neuter.

#### SEX

	MALE	FEMALE	NEUTER
ADULT	"Bull"	"Cow"	"Bullock"
IMMATURE		"Heifer"	"Steer"

AGE

Such an arrangement is termed a paradigm.

The complexity of a taxonomic system (and of coding systems generally) depends on two factors: a) the number of features employed (or the number of types of contrasts used), and b) the "fineness" of discrimination at each contrast level. In the example above sex and age are different semantic features; three categories of sex are discriminated, and (in the partial scheme above) two categories of age. MacKay (1969) refers to the first as the structural information content of a representation, or its logon-content ("the number of definably independent respects in which it could vary - its dimensionality or number of degrees of freedom") and to the second as metrical information-content, or metron-content ("the number of logical elements in a given group or in the total pattern").

There is nothing predetermined about the number of features or the fineness of a system of representation. As Tyler (1969) notes: "It is through naming and classification that the whole rich world of infinite variability shrinks to manipulable size and becomes bearable. Our methods of classification are entirely arbitrary and subjective. There is nothing in the external world which demands that certain things go together and others do not. It is our perception of similarities and differences together with a set of hierarchical cues that determine which things go together. We not only react to certain discriminable stimuli as if they were the same, we name them and organize them into groupings".

A classificatory, or representational system is still an "ensemble" in the language of information theory, but one in which there exists already an internal constraint. If such internal structure did not exist, the system could not function to represent the external environment. "Within any fixed system of symbols or events, structural meaning is prerequisite to signification meaning. Unless correlation exists between a symbol system and another symbol system or a system of real events, there can be no



external signification. Unless the symbols themselves are correlated, there can be no specific rules by which the internal signification can be learned". (Garner, 1962).

The measurement of total information content of the ensemble or system of representation must take into account both the internal structure of the system itself, and, since the symbols are correlated with external events, external structure. How frequently a given symbol appears has to do with the frequency of the event it stands for in the world of phenomena, and idiosyncrasies of the coding system. That the Shannon information theory measure can be adapted to handle semantic information in systems of representation is well-known, and requires no particular extra assumptions (McGill, 1954; Garner & McGill, 1956; Garner, 1958, 1962; and Watanabe, 1954, 1959, 1960).

The existence of a common ensemble of symbols, with common referents, is a pre-requisite to communication. Verbal communication between individuals does not however consist of the transmission of single symbols. Messages consist of concatenations of symbols (sentences) which are constructed according to generative and transformational rules (Chomsky, 1957). The number of such sentences which can be formed from a finite vocabulary in a language system which permits of recursiveness and imbedding is however infinite (Chomsky, 1957). This assumption of an infinitely

---

extensible ensemble of possible messages introduces a serious complication.

One solution which has been proposed is based on the procedure by which sentences are produced. At the core of all linguistic systems are a set of elementary propositions (or kernel sentences in Chomsky's original formulation). An elementary proposition contains one or more points of reference ("subjects" and "objects" of the sentence) and posits something about it or them. The positing takes in general two forms: a) it assigns an attribute to the point of reference ("Roses are red", "Children run"), or b) it states a relationship between two, or more, points of reference ("Horses eat grass", "Ottawa is between Toronto and Montreal").

Each elementary proposition stands for a state of some part of the external environment. Now the set of all elementary propositions is finite, and in fact the total set of such propositions constitutes a state-description of the environment. Furthermore, to each state-description can be assigned an a priori probability. Hence for the set of all elementary propositions, when weighted by their probabilities, an information content value can be defined. The transmission of a message, hence, conveys information about the environment.

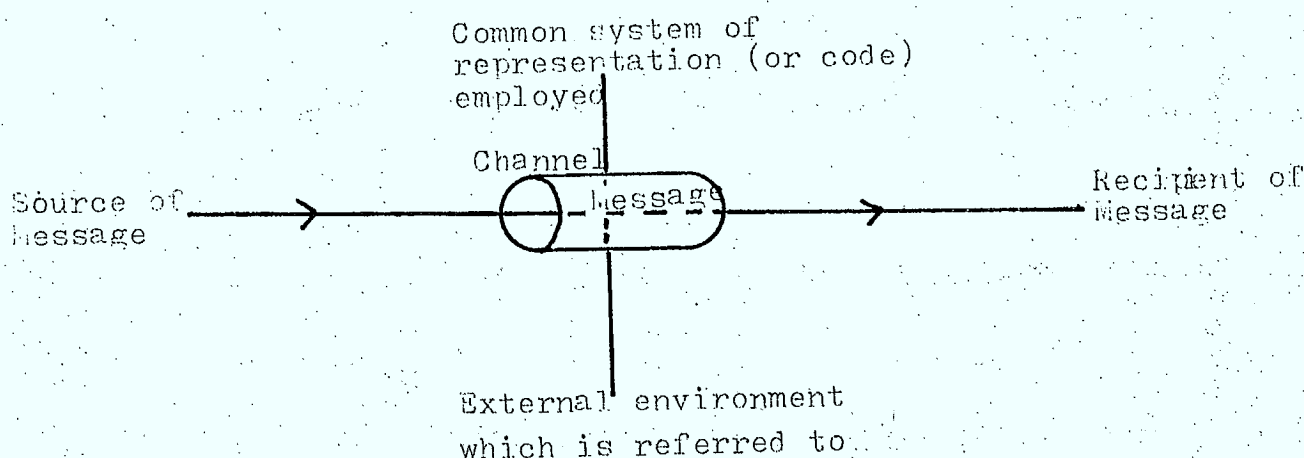
It should certainly be recalled at this point, however, that the information content of messages (as we saw in Chapter 1) is a measure of the amount of uncertainty associated with the ensemble as a whole, and varies with the average reduction in uncertainty per message. It is not intended to measure the information carried by any particular message. Hence, if we wish to know, in a real-life situation, how much information overload an individual is subjected to, we should have to know how frequently he

receives messages, and, secondly, what is the total uncertainty in his environment.

Problems of the determination of relational information content

The discussion which preceded suggests the need for a revision of the Shannon model to incorporate ideas of a common code, or system of representation, and reference to an external environment.

FIG. III-2



It has been proposed (Jakobson, 1950) that parallel to the several relations implicit in this model (source-message, message-recipient, message-code, message-referent, message-channel, message-message), communication has several discriminable functions (and hence several points of reference):

a) Expressive: many messages (verbal, non-verbal, iconic) report the state of mind of the sender. This function is enshrined in but not limited to phrases such as "I feel", "he believes", "she is angry", "we wish", etc. The total structure of the message, including external structure, may have to account for the total uncertainty of the sender's internal states.

b) Referential: the relation between message and external environment has already been discussed.

c) Metalinguistic: some messages concern the code itself "when I say X, what I mean is...". There is always some uncertainty about the code, and hence messages can inform the receiver about the code itself.

d) Phatic: the term "phatic" has been employed to refer to messages which are about the channel itself (keeping it open, terminating the connection, etc.) In ordinary conversation it sometimes appears that an inordinate number of messages have a mainly phatic function.

e) Poetic: as Garner (1962) has noted, there is internal as well as external significance. Some messages refer to the association within language, and these are appropriately termed the "poetic" function of communication.

f) Conative: sentences phrased in the imperative are clear examples of the conative, or "effective", function of communication. This is what MacKay (1969) has in mind when he writes: "The meaning of a message can be defined very simply as its selective function on the range of the recipient's states of conditional readiness for goal-directed activity; so that the meaning of a message to you is its selective function on the range of your states of conditional readiness".<sup>1</sup>

Let us develop this idea slightly further.

When a message reaches its recipient (intended or otherwise), it may produce a contingency between the subsequent output of the recipient

---

<sup>1</sup> Cf. Searle (1970), p. 48.

and the message. MacKay calls this the "meaning" of the message. Austin (1962) termed it the illocutionary force of the message. He argued for a distinction between the performance of an act of saying something, and the performance of an act in saying something. The latter constitutes the performance of an illocutionary act. He was thus led to distinguish between the force of a message, and its meaning. What MacKay called "meaning", Austin meant as "force". Thus one further point of reference of a message is the domain of acts produced by the recipient upon its reception.

A similar point has been made by Newcomb (1952). Newcomb argued that a communicative act links three elements: a communicator, a person being communicated to, an environment. This triad he termed an "A-B-X" system. A communicative act must assert something about an environment (what Searle, 1969, terms its propositional content). At the same time, and at a different level, it constitutes a state of the A-B-X system. "It is presumed that a given state of the system exists when a given instance of AtoBrex occurs, and that as a result of this occurrence the system undergoes some change (even though the change may be regarded as only a reinforcement of the pre-existing state)". Thus all communicative acts at once describe a state (on one level), and are a state (at another level, that of the A-B-X system). They report an experience, and they impose a response.

Another way to make the distinction is to consider the difference between "information-transmission" and "signalling". The result of information-transmission, as of direct experience, is to increase the recipient's certainty about something where there was, previously, an absence of knowledge. Signals, however, customarily set off, or trigger, sequences of activity, as for example when a hockey referee drops the puck. Communication

serves both to transmit information and to signal.

Mahl (1959) has referred to the distinction between a representational model and an instrumental model of communication behavior. It "A" wishes "B" to choose a certain line of behavior, he chooses those messages which his past experience has led him to believe are most likely to produce the responses he wishes. He may not represent directly either his own state, or the response he wishes. He may not say: "I'm cold; close the door", but rather "it gets cool early these days, doesn't it?" He chooses a message having the illocutionary force appropriate to the situation.

What is the nature of the "force" involved? Searle (1970) has noted a distinction between "brute" facts, and "institutional" facts. Acts which owe their force to brute facts have their basis in the physical realm. Institutional facts owe their reality to the operation of what he terms "constitutive" rules, of the type: "X counts as Y in context C". Saying "I will" in a church, before a minister, to someone of the opposite sex, following the phrase "Do you take this man (woman) in holy wedlock?" constitutes marrying that person. Even though the expenditure of physical energy is feeble, the effects may well be momentous!

In the interpretation of messages, there appear to be two decoding stages: one based on the system of representation underlying the communicative act, and one based on a system of interaction. It is this latter coding system which determines how we assess the expressive and conative content of the messages.

The concept of "interaction" will lead in turn to that of "relationship".



First, let us develop slightly the notion of interaction.

FIG. III-3

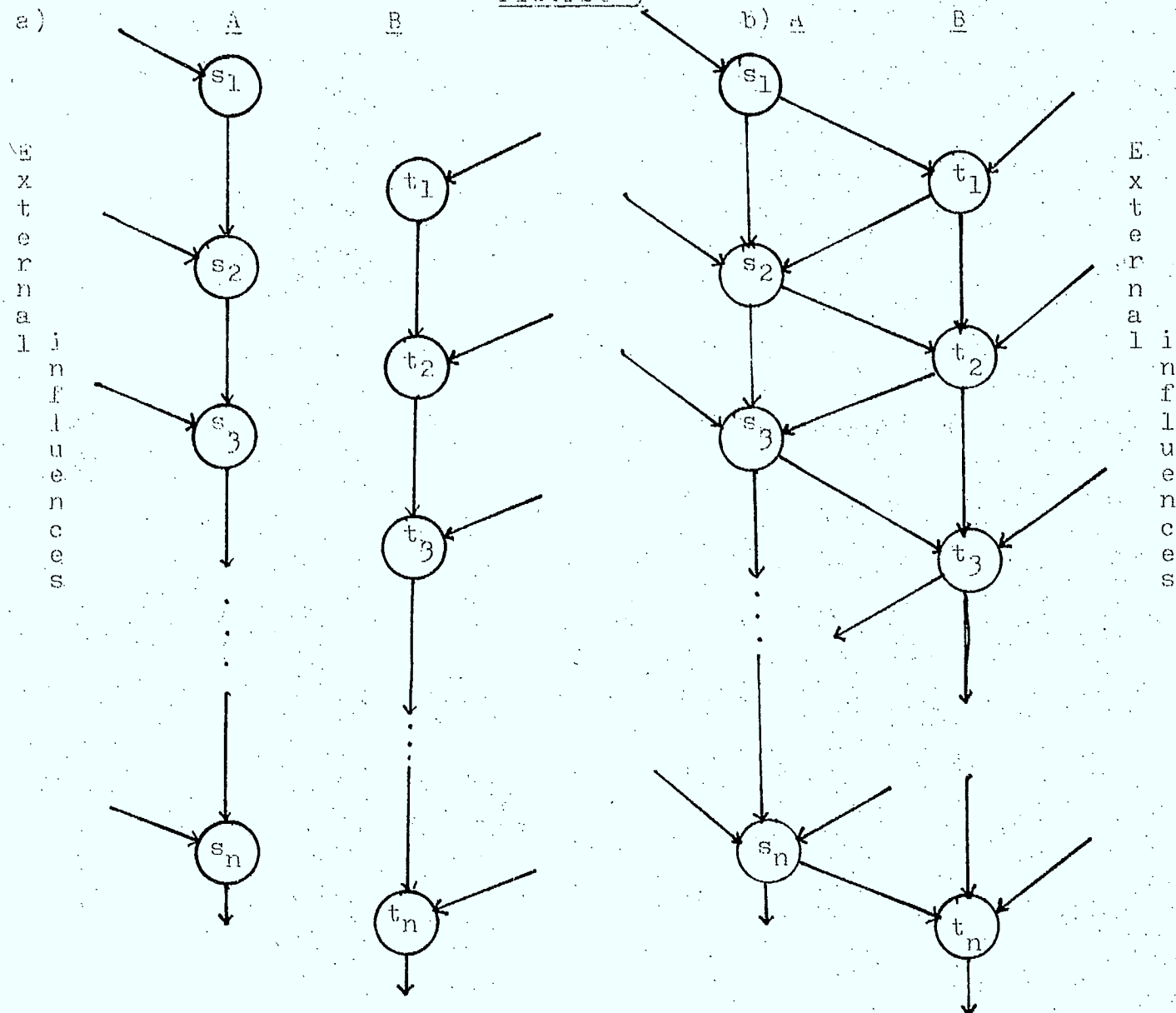
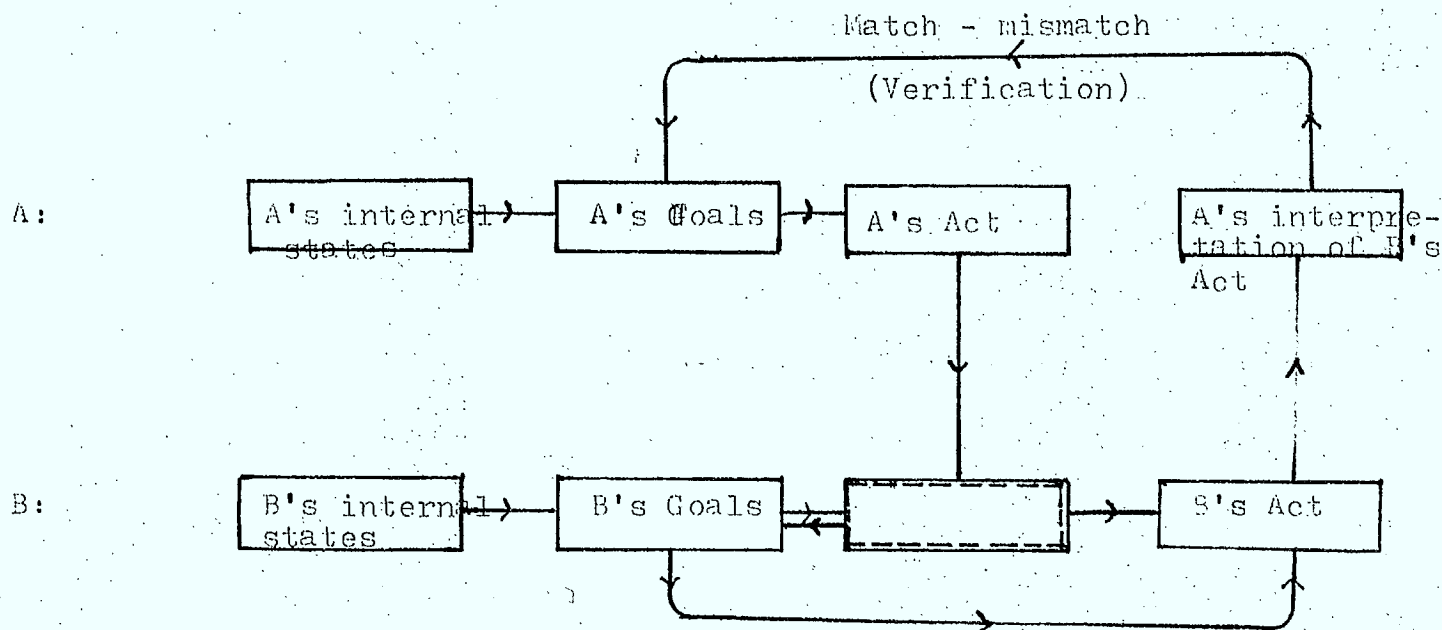


Illustration of the concept of interaction

In Fig. III-3 -a there is no interaction between A and B. Each state ( $s$  and  $t$ ) of the two individuals can be explained as a function of previous states of the individual, and of external influences. With interaction arises the possibility of mutual influence. It is however an influence of two goal-directed systems. (III-3 -b)

Hulett (1966) has shown how a completed interaction sequence might be analysed:

FIG.III- 4



In general these are the stages:

a) goal identification: "To the extent that A's orientation either toward X or toward B is contingent upon B's orientation toward X, A is motivated to influence and/or to inform himself about B's orientation toward X". (Newcomb, 1952)

b) choice of instrumentality. Given the need to do something: to undertake a particular act, express opinion, reveal his feelings, the means to accomplish this end is the delivery of a particular "AtoBrex" which, it is hoped, will elicit the desired response on the part of B.

c) B's response ("BtoAreX").

d) Validation of the response. The communicator now finds his original objective (one particular BtoArex) either confirmed or disconfirmed; if disconfirmed, presumably, the need is intensified and the likelihood of a new (probably "stronger") AtoBreX increases. The cycle is repeated, but with the difference that A's pre-existing image is now changed, with

consequent implications for the selection of the next instrumental act.

The choice of a certain illocutionary act, in a specific situation, has to do with the communicator's perception of the relationship between himself and the object of his message, in other words with the constitutive rules he considers appropriate for this context ("X is to count as Y in context C").

The interpretation phase of the sequence raises a different point: The issuing of a communication is invariably informative about the state of the organism who executed the act.<sup>1</sup>

One function of communication is thus, in Newcomb's words, "to maintain simultaneous orientation toward one another". The information contained in the message concerning the communicator's states, and his choice of act, are critical to the determination of the other's attitudes towards self. Hence one motive which may underlie a communicative act is the attempt to either assert or re-assert something about the essential

---

1 This follows logically from the mutual presenting of two persons to each other: in one sense all acts are communicative (cf. Watzlawick et al's (1967) axiom "One cannot not communicate").

A similar point of view is expressed by Leach, 1970: "When an individual acts as an individual, operating upon the world outside himself -- e.g. if he uses a spade to dig a hole in the ground -- he is not concerned with symbolisation, but the moment some other individual comes onto the scene every action, however trivial, serves to communicate information about the actor to the observer -- the observed details are interpreted as signs, because observer and actor are in relation".

relationship between the two communicators. For the recipient of such a relationship-related communication, the process of interpretation may be complex, and based on elaborate procedures on inference.

Harold Kelley (1966) has made a careful analysis of the structure of the communicative act which we call a "threat". The illocutionary act, in Austin's terms, may be formalized as "A cause Y, unless B cause X" or "not (B cause X) implies (A cause Y)". From this, we may infer that: (a) A wants X to occur, (b) is indifferent or negative to the occurrence of Y, (c) B does not want X to occur (at least by his agency), (d) B strongly does not want to occur, A intends X to occur, A intends Y to occur given not X, B does not expect Y to occur, (e) A believes X is feasible, (f) A believes Y is feasible, (g) B believes X is feasible, (h) B believes Y is feasible, (i) A is not strongly attracted to B, (j) B is not strongly attracted to A, (k) A assumes B's interests to differ from his, (l) A is asking that his interest take priority over B's, (m) A is relatively stronger than B. These inferences appear to be related to the concept of threatening itself; Kelley argues however that in the interpretation of the threat B may also have to consider what might be termed contextual factors: in particular, the strength of A's need for X, A's perception of the cost to B, and the relative pre-existing statuses of A and B. If A is perceived to be motivated by a very strong desire, then "his threat is more in the nature of a frantic plea for help than an attempt at intimidation". That is to say, the intended threat fails to meet the requirements of the communicative act of threatening, even if it succeeds as a plea, and hence accomplishes the purposes of the communicator. Thus, the intended threatener must take care not to express an emotional state, e.g. extreme anger, which

is consistent with the assumption of very strong underlying need, else "he reveals a weakness which enables others to comply with his threats without loss of face". His weakness consists in his dependence on the threatened person for satisfaction of a very important desire -- he opens up the possibility of blackmail. Conversely, argues Kelley, "the lesser the indicated need ... the greater the extent to which the latent message reads, I don't really care about this thing I'm asking. This is simply an occasion for setting you straight about who is on top in our relationship".

Thomas Schelling (1960) has demonstrated that in order to communicate effectively a threat, the threatener must also effectively express an appropriate attitude towards Y. The most effective means to communicate such an attitude is to show that in the event of the non-occurrence of X, A's motivation to cause Y becomes very strong. In this case, the establishment of a causal connection between the two events means that "the threat is no more than a communication of one's own incentives, designed to impress on the other the automatic consequences of his act". In this event, the "threat" appears to degenerate into a "warning"; not uncommonly, communicators specify: "No, I'm not threatening you, I'm just warning you" in such circumstances. To establish a threat, the communicator must communicate, Schelling proposes, his relative indifference to the occurrence of Y! should it become clear that he actively desires the non-occurrence of Y, then the threat is unlikely to have the intended effect.

Every statement of a point of view by an individual is part of his presentation of self: the two dimensions of transmission of information and assertion of a determinable and consistent self-image cannot be disentangled. "Since most persons have a positive conception of self, a very pervasive tendency in social interaction is to maintain a presentation

of self consistent with the favorable conception... In any social interaction a person is attempting to validate his occupancy of several important social positions". (Allen, 1968)

Every interaction may be weighed in terms of its potential effect on raising or lowering a chosen social value. The value cannot be measured against an absolute social scale : the individual is forever in the situation of attempting to assess the underlying scale on the basis of a series of paired comparisons. He therefore requires sources of information, consisting of (a) the choice of communicative acts on the part of others, directed towards him, and (b) the feedback to his own communicative acts directed towards others. Potentially, therefore, no interchange is irrelevant to his attempts to maintain social value.

#### network overload

Through out most of the discussions on overload, both in this report and in the literature generally, the question of effects has been posed at the level of the individual. Our discussion of information and communication has opened the possibility of proceeding to a higher level of system, at which we may ask what is known concerning network overload. The importance of this aspect of the problem is this: just as it is true that overload at the individual level may have extremely serious by-products (in the form for example of nervous disorders, illness, shortening of life-span), we should expect network overload to have similar dysfunctional consequences, in the form of reduced efficiency of group functioning, lowered organizational efficiency, and eventually the breakdown of social organization itself.

---



Since, contrasted with the field of individual information processing, no well-developed body of literature related to network information-processing exists, our approach will be slightly different. We will first suggest a certain number of ways the question could be addressed. We will then consider some of the available experimental evidence, in order to develop, as far as possible, the empirical basis for our assumptions. We proceed in two steps: first we examine networks in which referential information is transmitted; second, we turn to the question of the effects of varying amounts of relational information.

#### The transmission of information in networks

---

A first question we may ask is what is the effect of increasing the size of the network. First we note that for any fully-connected network, i.e., where everyone can communicate with everyone, the potential number of messages which can be received per position, " $x$ ", is a function of the number of nodes in the network:

$$x_1^{\max} = n - 1$$

where " $x_1$ " represents the number of possible simultaneous received messages for a node, and " $n$ " the number of nodes in the network. For networks with two elements, only one message per instant may be received. For networks with three elements, two messages may arrive simultaneously. For four nodes, 3. Etc.

FIG. III-5



Illustration of number of possible simultaneous message for a given position

For totally connected networks, the total number of possible simultaneous messages, for all positions, is:

$$Y_{\max} = \sum_{i=1}^n x_i = n(n-1)$$

Even though, since neither processing nor transmission time are accounted for in this formula, the maximum would rarely be attained, still it is obvious that even minimal increases in the size of network are likely to increase materially the load on individuals in the network.

Let us assume that each node has some average maximum output, which we may want to represent in terms of information theory as a certain number "a" of bits per period of time "t". Then the actual information input of a node is

$$z_i = a_t x_i$$

and the maximum input per position is

$$z_i = \max a_t x_i$$

The maximum acceptable input is not necessarily the same as maximum output: the individual may select pertinent information, may "chunk" it in different forms, may store it, etc. Hence increases in available information will not lead in themselves to overload. However it is reasonable to assume that the maximum output is some (possibly variable) function of the input, which we might therefore represent by a weight "w". Maximum output is then

$$\frac{\max}{w \cdot z}.$$

Suppose the value of  $w$  is set so that the output of one mode is sufficient to occupy totally the attention of another, in a given time period, and overload occurs as soon as one individual has to pay attention to the production of more than one other person. The system as a whole may still not overload because communication is not continuous. Nevertheless, increases in size cause increased pressure on the system. There are two ways adjustment can occur; first, the total time spent in communicating may increase and second, the frequency with which one communicates with any other given individual may decline.

To illustrate, let A be in a network with two other individuals, B and C. Each spends  $\frac{1}{2}$  of his time communicating. Each spends at a maximum (in the fully-connected network) 50% of his time receiving communications (leaving him 25% for communication and 25% "free"). Now A moves into a 5-man network. Since he now has 4 input channels, each 25% occupying of his time, he has no time left for communicating. His solution is (a) to increase his total time spent in communication (both sending

and receiving), (b) to reduce the frequency with which he communicates to any given individual, by restrictions, for example, on the all-channel characteristics of the network.

As soon as network restrictions occur, another problem arises: saturation. If for one reason or another, because of physical, social or other constraints, nodal capacity or actual nodal performance is not standardized, then one position may become "saturated", in the sense of being unable to handle all the traffic directed through it.

Finally, it should be observed that the uncertainty of the environment is the other critical factor. How quickly a network loads is presumably a function of how complex the information-transformation requirements of the task are.

#### Experimental results

-----

A number of the propositions advanced here have been investigated experimentally. Walker (1954) discovered for example that, as the size of network increased from three to five, group efficiency (in terms of problem solution times, and errors) decreased, group morale declined, the number of messages increased and unanimous selection of a leader decreased. This result tends to support the assumption that increases in size place additional strain on groups and increase overall levels of communication.

The concept of "saturation" was proposed by Gilchrist et al (1954) to describe the condition associated with a super-optimal load of

number of messages for a given position. Two types of saturation were identified: "channel saturation" and "message unit saturation". The former refers to the number of channels a position must handle, the latter to the number of messages.<sup>1</sup> Shaw (1964) discriminates further between input saturation (task information to be transmitted), noting at the same time that saturation refers to the total requirements placed upon an individual in a given position, including non-communication-related information demands such as data manipulation. Shaw (1954 a,b) showed that the central person in centralized nets tends to become overloaded by the communication requirements of his post.

Environmental complexity seems to have received less attention. Tuckman (1964) concluded that groups tend to become hierarchical under stress induced by environmental complexity. Shaw (1964) however concluded from his review of the evidence that centralized groups were more effective in solving simple problems (essentially the exchange or collation of pre-assigned data), while decentralized networks did better on more complex problems (solving arithmetic problems).

The concept of centralisation embodied in the experiments reported by Shaw has been criticised by Mackenzie ( 1968 ), who notes a confusion between an assumed network structure and the real structure derived from actual observed group interaction patterns.

---

1 Mackworth and Mackworth (1956) projected up to 12 sources of information simultaneously to subjects through separate windows. Decrements in performance were associated with (1) the number of windows employed and (2) the amount of overlap of messages.

Relationship and information transmission.

Our previous argument led us to believe that in order to communicate individuals have to learn two basic codes: a system of representations which determine how symbols are mapped to external referents, and a system of constitutive, or institutional, rules which determine how messages are to be taken, insofar as they specify responses on the part of the recipient, and imply attitudes on the part of the sender. Role relationships between individuals arise whenever standardized patterns of message-exchange emerge, either by agreement or from practice.

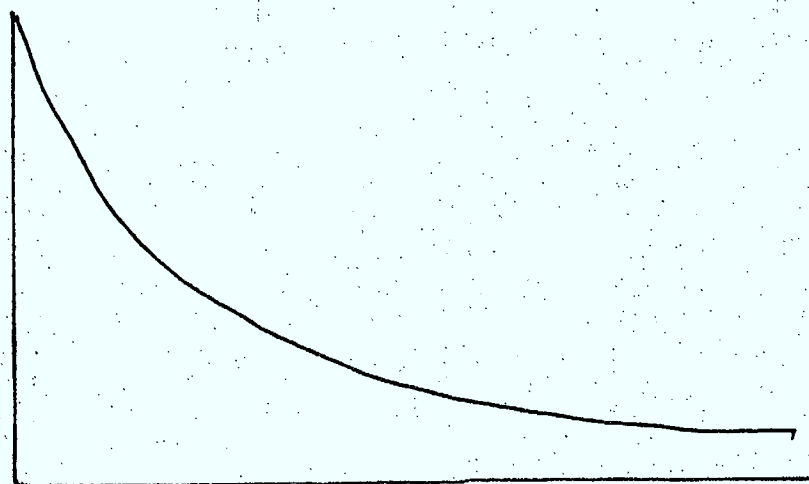
We expect stability of role relationship to be associated with frequency and length of interaction. In a frequently-cited experiment, Schacter (1951) introduced a confederate into a group programmed to produce deliberately discordant communications. Other members initially intensified the number of communications addressed to the "odd man out", and when he remained recalcitrant they then sharply reduced communications to him, effectively relegating him to Coventry. In general, continued interaction appears to lead to a measure of stabilisation (normalisation) of relations, at least to a reduction of uncertainty concerning the probable behavior of the other.

As uncertainty is reduced concerning the pattern of interpersonal relations, the number of communications concerned with this domain tends to diminish (although not necessarily in a smooth curve). Let us represent this assumption as a function:



FIG. 111-6

Proportion of  
communications  
related to  
Relationship/  
Referential  
information

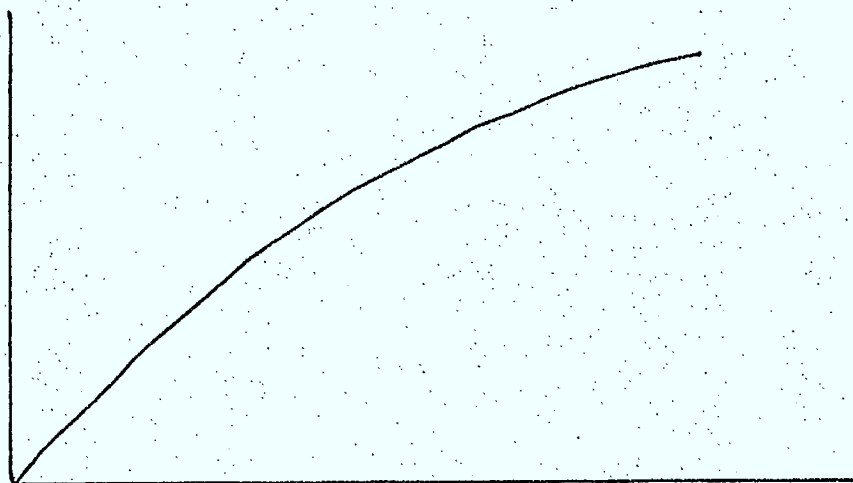


Duration of interaction

Then, as one's network increases, total uncertainty concerning relationships increases, as a positive function (the flatness of the curve reflecting on assumption that networks increment slowly):

FIG. 111-7

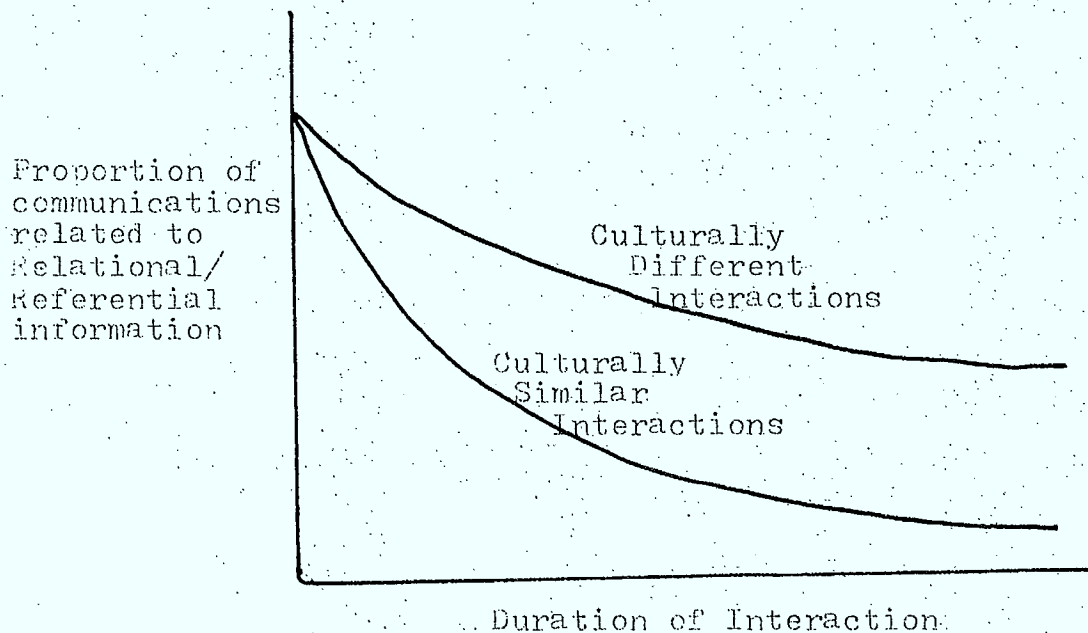
Relational  
Uncertainty



Size of Network

Furthermore, since, in a large, and culturally diverse, population, agreement concerning the code of constitutive institutional rules is likely to be highly variable, the functional relationship between duration of interaction and the proportion of communications devoted to Relational/Referential information is affected by the extent of cultural similarity.

FIG. III- 8



Hence the steepness of the curve in Figure III -8 is a function of cultural diversity. In general, also the larger the network, the more heterogeneous culturally it is likely to be, and hence the greater the probability of high relational uncertainty in large networks.

Assuming that each individual is only able to deal with a limited amount of total uncertainty, then an increase in relational uncertainty should produce a reduced efficiency with respect to the processing of referential information. This is the conclusion reached by Shaw as a result of his survey of relevant experiments. He defines a concept "independence" which is determined by the perception of individuals of their role, in the group, and then adds: "The effect of independence upon performance is due... to the individual's willingness and ability to perform under the more autonomous conditions. That is, lowered independence not only directly limits the possibilities for action (hence performance), but also reduces the person's willingness to perform at his optimal level".

#### Experimental support

-----

The literature appears to be relatively poor as it concerns the points raised above. Some support can be found in the work of Bales (Bales, 1953 ; Dunphy, 1966) to support the idea that the development of interpersonal systems proceeds in several phases, as more complex interaction patterns develop. To accommodate this idea the function shown in Figure III - 6 should incorporate a cyclical pattern.

Since some kind of balance between relationship-modifying and information-transmission processes is implied by our argument, we might ask, conversely, what happens to social organization under conditions of extreme pressure to transmit information. A number of experiments (e.g. Hovland and Weiss, 1952; Kelman, 1958) show that information transmission is affected by the perception of relation; we have not however been able

to discover any experimental evidence concerning the reverse situation, and particularly concerning the effects of information overload on the perception of relationship.

### Perspective

In this final chapter we have sketched in, all too lightly, the general outlines of an integrated theory of information and program overload. In doing so we have not been able to draw on the rich sources of experimental findings that were available at the level of the individual organism. There is, clearly, a great need for systematic theory and research at the group level. Individuals function within groups, and some of the primary side-effects of the communication explosion are not to be found by looking only at individuals. A government office, a garage, a university are systems, just as the individual is, and their processes can also be discussed with the same kind of precision as the cognitive processes of the person. We need however finer instruments than we now possess, to permit us to describe properly phenomena of communication in networks. The discussion in this chapter is intended to be a contribution towards this goal.

At the beginning of our investigation we warned that, although we had a practical goal in view, some of our discussion would seem abstract. This should not be read as an apology. There is a dynamic interplay between theory and practical which must never be neglected. Those responsible the **development of theory and research** must **learn from practical experience** by putting their ideas to the test. Those responsible for practical planning are equally likely to miss the boat by asking

questions which are set at too low a level: in no other century has it been so evident how practical is theory! We are constantly in need of conceptual tools; otherwise we risk the pursuit of empty information.

To quote Simon (1968):

"Science does not advance by piling up information -- it organizes information and compresses it... In the scientific endeavour, 'knowing' has always meant 'knowing parsimoniously'. The information that nature presents to us is unimaginably **redundant**. When we find the right way to summarize and characterize that information -- when we find the pattern hidden in it -- its vast bulk compresses into succinct equations, each one enormously informative.

Herein lies the real significance of today's information revolution. Information and the processing of information are themselves for the first time becoming the objects of systematic scientific investigation. We are laying the foundations for a science of information processing that we can expect will greatly increase our effectiveness in handling the information around us".

One important conclusion which can be reasonably drawn from our discussion is that the real problem of the wired nation is not that of information overabundance as such but of the undertaking of too many separately interesting tasks which together result in the condition we term "program overload". In chapter I of this report it was noted that the introduction of new technologies has two consequences: while it augments available information, it also leads individuals to interact within wider networks. The latter fact seems to us to be the more important, - and the more neglected. It implies that individuals tend to get involved, simultaneously, in a number of transactions.

The danger of overabundance of available information can be easily exaggerated. The mechanisms of perception are adapted to inhibit, or suppress, unneeded and unwanted information. Milgram (1970) has demonstrated that individuals living in information-rich New York City exhibit equal ability to select needed and screen out unwanted information in this environment (which they chose indeed because of its information characteristics, as Mumford and other urbanologists have long been at pains to point out).

By comparison, the individual seems disastrously non-adapted to deal with program overload. This we take to be the significance of a recent article by Lipowski (1971), who argues: "a major feature of the affluent, technological and open society is that it exposes its members to an overload of attractive stimuli". Lipowski defines "attractive stimuli" as those which "arouse appetitive and approach tendencies". They involve his capacity to "process, choose, approach and consummate". In our terms, this would be re-interpreted as the tendency to become involved in more activities than can be adequately handled by the individual.



The process is insidious: each new activity is initially attractive, and the individual may tend to underestimate its time requirements over a longer period. Eventually, he is committed to more activities than can be managed at once, each with accelerating information processing requirements. The final product may well be confusion and breakdown.

The evidence we do have indicates that program overload can prevent successful concept formation, and hence reduce the individual to stereotyped responses: As Meier (1962) writes: "The culture of cities cannot grow and develop unless man's interpretation of the universe and man's study of man supplies new concepts and images more rapidly than they are lost".

Another question which has received little attention in the available literature is the role of human communication. To a surprising extent, we have been led by the influence of a communication model based in engineering research to suppose that the analysis of information system could be conducted without regard to other communication variables. This situation needs to be rectified, particularly since it is at this level, above all others, that the social (as opposed to the psychological), effects of technological innovation will be most evident.

We have reasoned as follows: man requires information (verification and comparison of his hypotheses and perceptions), but he is vulnerable to control.

Communication has as one effect the linking of information transmission systems into chains, in which (because of the symbolic capability of man) one can serve another as his eyes and his hands. It is because of ~~this~~ "control" dimension of communication, even more than because of the informational, that individuals establish systems made up of statuses and roles, which (a) assure to individuals a certain stability in their communication relationships, and (b) create the basis of an organized society. What has not been asked in the literature on information overload is the effect that intensified information processing may be expected to have on the social structure which the information-control system supports. As the individual becomes overloaded, does the resulting stress make for increasing dependence on others? Or decline in trust? Under conditions of overload, does it become increasingly difficult to organize the necessary social arrangements? Does comprehension decline? Are there effects on how much information is accepted from others?

The answers to these and similar questions cannot be found in the existing literature, yet this information is of increased urgency: it will have gained us little if in extending our means to communicate with each others, we succeed, in the process, in destroying the very basis on which our society can be organized. We need to learn not only how to handle more information, but at the same time how to re-organize the accompanying social arrangements: "The breaking up of...organization, or its reduction to impotency due to communications overload, is...frequently due to inadequacies in the basic formula for doing business or in the informal 'rules of the game' set by occupations and professions" (Meier, 1962).

It is in this perspective that the present report has been undertaken. Our goal must be the intelligent planning of our own communication systems.

## BIBLIOGRAPHY

- Alluisi, E. A., P. F. Muller, & P. M. Fitts (1957) An information analysis of verbal and motor responses in a forced-pace serial task. J. exp. Psychol., 53, 153-8.
- Ashby, W. R. (1952) Design for a brain. London: Butler & Tanner, Ltd.
- Ashby, W. R. (1956) An introduction to cybernetics. London: Chapman & Hall.
- Averbach, E. & A. S. Coriell (1961) Short-term memory in vision. Bell Syst. tech. J., 40, 309-328.
- Bair, J. (1971) Human information processing in man-computer systems. Paper presented to the International Communications Association. Phoenix, Arizona.
- Baker, E. J. & E. A. Alluisi (1962) Effects of complexity, noise, and sampling rules on visual and auditory form perception. Marietta, Georgia: Human Factors Research Department, Lockheed-Georgia Co.
- Beebe-Center, J. G., M. S. Rogers, & D. N. O'Connell (1955) Transmission of information about sucrose and saline solutions through the sense of taste. J. Psychol., 39, 157-60.
- Biederman, I. (1966) Human performance in contingent information processing tasks. University of Michigan, Ann Arbor. Human Performance Center, Tech. rep. no. 3 (DDC-AD666-666). Dissertation (Ph. D.).
- Birdwhistell, R. L. (1970) Kinesics and context: Essays on body motion communication. Philadelphia: Univer. of Penn. Press.
- Blank, G. (1934) Brauchbarkeit optischer Reaktionsmessung. Indust. Psytechnik, 11, 140-150. Cited by J. G. Miller, Living Systems. (forthcoming).

- Botwinick, J., J. F. Brinley, & J. S. Robbin (1958) The interaction effects of perceptual difficulty and stimulus exposure time on age differences in speed and accuracy of response. Gerontologia, 2, 1-10.
- Brady, J. V. (1958) Ulcers in executive monkeys. Scientific Amer October, 95-100.
- Bricker, P. D. (1955) Information measurement and reaction time: A review. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Illinois: The Free Press.
- Broadbent, D. E. (1955) Noise, paced Performance, and vigilance tasks. Brit. J. Psychol., 44, 295-303.
- Broadbent, D. E. (1954) The role of auditory localization in attention and memory span. J. exper. Psychol., 47, 191-6.
- Broadbent, D. E. (1958) Perception and communication. Oxford & Toronto: Pergamon Press.
- Brock, L. J., J. S. Coombs, & J. C. Eccles (1953) Intracellular recording from antidromically activated motor neurons. J. Physiol., 122, 429-461.
- Bruner, J. S. (1958) Neural mechanisms in perception. In Solomon, Cobb, & Penfield (eds.), The brain and human behavior. Baltimore, Maryland: The Williams & Wilkins Company.
- Bruner, J. S., J. J. Goodnow & G. A. Austin (1956) A study of thinking. New York: Science Editions, Inc.
- Canhon, W. B. (1932) The wisdom of the body. London
- Chomsky, N. (1957) Syntactic structures. The Hague: Mouton & Co.
- Chomsky, N. (1959) A review of R. F. Skinner's Verbal behavior. Language

Committee on Telecommunications (1969) Telecommunications for enhanced metropolitan function and form. Contract no. OEP-SE-69-101.

Washington, D. C.: Academy of Engineers.

Creutzfeldt, O. & O.-J. Grüsser (1959) Beeinflussung der Flimmerreaktion einzelner corticaler Neurone durch elektrische Reize unspezifischer Thalamuskern. Proceedings 1st Internl. Congr. Neurol. Sci.

Brussels, Vol. III. EEG Clinical Neurophysiology and Epilepsy.

London: Pergamon Press. See also K. H. Pribram (Ed.), Perception and action, Brain and behaviour 2. Middlesex, England & Baltimore,

Maryland: Penguin Books Ltd.

Cronbach, L. J. (1955) On the non-rational application of information measures in psychology. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Illinois: The Free Press.

Crossman, E. R. F. W. (1953) Entropy and choice time: The effect of frequency unbalance on choice response. Quart. J. exp. Psychol., 5, 41-51.

Crossman, E. R. F. W. (1955) The information capacity of the human operator in symbolic and non-symbolic control processes. In J. Drake (Ed.), The application of information theory to human operator problems. London: Ministry of Supply, Great Britain.

As cited by J. G. Miller, Living Systems (forthcoming).

Deese, J. (1967) General psychology. Boston: Allyn & Bacon.

Deininger, R. L. & P. M. Fitts (1955) Stimulus-response compatibility, information theory and perceptual-motor performance. In H. Quastler (Ed.), Information theory in psychology. Glencoe, Illinois:

The Free Press.



- DeValois, R. (1958) Discussion in S. H. Bartley, Some facts and concepts regarding the neurophysiology of the optic pathway. A. M. A. Arch. Ophthalmol., 60, 784-5.
- Diebold (1952) Automation. New York.
- Eriksen, C. W. & J. S. Lappin (1964) Luminance summation--contrast reduction as a basis for certain forward and backward masking effects. Psychon. Sci., 1, 313-4.
- Eriksen, C. W. & J. F. Collins (1964) Backward masking in vision. Psychon. Sci., 1, 101-2.
- Birren, J. E. & J. Botwinick (1955) Speed of response as a function of perceptual difficulty and age. J. Geront., 10, 433-6.
- Fehrer, E. & I. Biederman (1962) A comparison of reaction time and verbal report in the detection of masked stimuli. J. exp. Psychol., 64, 126-30.
- Fehrer, E. & D. Raab (1962) Reaction time to stimuli masked by metacontrast. J. exp. Psychol., 63, 143-7.
- Final Report President's Task Force on Communications Policy (1968)  
Washington: U. S. Printer's Office.
- Fitts, P. M. & C. M. Seeger (1953) S-R compatibility: Spatial characteristics of stimulus and response codes. J. exp. Psychol., 46, 199-210.
- French, J. D. (1957) The reticular formation. Scientific American, 66.
- Garner, W. R. (1953) An informational analysis of absolute judgements of loudness. J. exp. Psychol., 46, 373-80.
- Garner, W. R. (1962) Uncertainty and structure as psychological concepts.  
New York: John Wiley.

Glucksberg, S. (1965) Decay and interference in short-term memory.

Paper presented to Psychonomic Society. Chicago, Illinois.

Granit, R. (1955) Centrifugal and antidromic effects on ganglion cells of retina. J. Neurophysiol., 18, 388-411.

Granit, R. & C. G. Phillips (1956) Excitatory and inhibitory processes acting upon individual purkinje cells of the cerebellum in cats. J. Physiol., 113, 520-47.

Gregory, R. L. (1966) Eye and brain, the psychology of seeing. New York & Toronto: McGraw-Hill Book Company.

Hake, H. W. & W. R. Garner (1951) The effect of presenting various numbers of discrete steps on scale reading accuracy. J. exp. Psychol., 42, 538-66.

Harris, J. D. (1950) Some relations between vision and audition. Springfield, Illinois: Charles C. Thomas.

von Helmholtz, H. (1963) Handbook of physiological optics. J. P. C. S. Southall, Dover Reprint.

Hernandez-Peon, Scherrer, & Velasco (1956) Central influences on afferent conduction in the somatic and visual pathways. Acta Neurol. Latinamer., 2, 8-22. See also H. Pribram (Ed.), Perception and action, Brain and behaviour 2. Middlesex, England & Baltimore, Maryland: Penguin Books Ltd.

Hess, W. R. (1957) The functional organization of the diencephalon. New York: Grune & Stratton, Inc.

Hick, E. E. (1952) On the rate of gain of information. Quart. J. exp. Psychol., 4, 11-26.

Hilgard, E. R. & G. H. Bower (1966) Theories of learning. New York: Appleton-Century-Crofts.

- Hilgendorf, L. (1966) Information input and response time. Ergonomics, 9, 31-7.
- Hirsh, I. J. (1950) The relation between localization and intelligibility. J. acoust. Soc. Amer., 22, 196-200.
- Hovland, C. I. & W. Weiss (1952) The influence of source credibility on communication effectiveness. Public Opinion Quart., 15, 635-50.
- Hubel, D. H. & T. N. Wiesel (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol., 160, 106.
- Hull, C. L. (1952) A behavior system. New Haven, Conn.: Yale Univer. Press.
- Hunt, E. B. (1962) Concept learning: An information processing approach. New York: John Wiley.
- Hyman, R. (1953) Stimulus information as a determinant of reaction time. J. exp. Psychol., 45, 188-96.
- Jacobson, H. (1950) The information capacity of the human ear. Science, 112, 143-4.
- Jacobson, H. (1951) The information capacity of the human eye. Science, 113, 292-3.
- Jakobson, R. (1944) Toward a linguistic typology of aphasic impairments. In A. V. S. deReuck & M. O'Connor (Eds.), Disorders of Language, CIBA Foundation Symposium. Boston: Little, Brown & Co.
- Jakobson, R. & M. Halle (1956) Fundamentals of Language. 's-Gravenhage: Mouton.
- Jung, R. (1958) Coordination of specific and nonspecific afferent impulses at single neurons of the visual cortex. In H. H. Jasper, et. al. (Eds.), Reticular formation of the brain. Boston: Little, Brown & Co.

Jung, R. (1958) Excitation, inhibition, and coordination of cortical neurons. Exp. cell. Res., 5 (suppl ), 262-71.

Jung, R. & G. Baumgartner (1955) Hemmung-smechasismen und bremsend Stabilisierung an einzelnen Neuronen des optischen Cortex. Pflüg. Arch. ges. Physiol., 261, 434-56. As cited by J. G. Miller, Living Systems. (forthcoming).

Jung, R., O. Cruetzfeldt, & O.-J. Grüsser (1957) Die Mikrophysiologie Kortikaler Neurone und ihre Bedeutung für die Sinnes--und Hirnfunktionen. Dtsch. med. Wschr., 82, 1050-59. See also R. Jung, Neuronal integration in the visual cortex and its significance for visual information. In K. H. Pribram (Ed.), Perception and action, Brain and behaviour 2. Middlesex, England & Baltimore, Maryland: Penguin Books, Ltd.

Kelman, H. C. (1958) Compliance, identification, and internalization; Three processes of attitude change. J. conflict Resol., 2, 51-60.

Kirchner, W. K. (1958) Age differences in short-term retention of rapidly changing information. J. exp. psychol., 55, 352-8.

Klemmer, E. T. & F. C. Frick (1953) Assimilation of information from dot and matrix patterns. J. exp. Psychol., 45, 15-9.

Klemmer, E. T. & P. F. Muller, Jr. (1953) The rate of handling information: Key pressing responses to light patterns. HFORL Memo Report No. 34. USAF, ARDC. As cited by D. R. L. Laming, Information theory of choice reaction times. London & New York: Academic Press.

Klemmer, E. T. & P. F. Muller, Jr. (1969) The rate of handling information: Key pressing responses to light patterns. J. mot. Behav., 1, 135-47.

- Kock, W. E. (1950) Binaural localization and masking. J. acoust. Soc. Amer., 22, 801-4.
- Kornblum, S. (1967) Choice reaction times for repetitions and non-repetitions. Acta psychol., 27, 178-87.
- Kornblum, S. (1968) Serial-choice reaction-time: Inadequacies of the information hypothesis. Science, 159, 432-434.
- Kornblum, S. (1969) Sequential determinates of information processing in serial and discrete choice reaction time. Psychol. Rev., 76, 113-31.
- Laming, D. R. J. (1962) A statistical test of prediction from information theory in a card-sorting situation. Quart. J. exp. Psychol., 14, 38-48.
- Laming, D. R. J. (1968) Information theory of choice-reaction times. London and New York: Academic Press.
- Lashley, K. S. (1942) The problem of cerebral organization in vision. In H. Kluever (ed.), Biological symposia, Vol. VII: Visual mechanisms. Lancaster: Jaques Cattell Press. See also H. Pribram Perception and action, Brain and behaviour 2. Middlesex, England & Baltimore, Maryland: Penguin Books, Ltd.
- Leonard, J. A. (1959) Tactual choice reactions: I. Quart. J. exp. Psychol., 11, 76-83.
- Leonard, J. A. (1959) Choice reaction time experiments and information theory. In E. C. Cherry (Ed.), Information theory. London: Butterworths.
- Lipowski, Z. J. (1971) Surfeit of attractive information inputs: A hallmark of our environment. Behav. Sci., 16, 467-71.

- Mackworth, J. (1963) The duration of the visual image. Canad. J. Psychol., 17, 62-81.
- Mackworth, J. & N. H. Mackworth (1956) The overlapping of signals for decisions. Amer. J. Psychol., 69, 24-47.
- Mandler, G. (1964) The interruption of behavior. In D. Levine (Ed.), Nebraska symposium on motivation: 1964. Lincoln, Nebraska: Univer. of Nebraska Press.
- Mandler, G. (1967) Verbal learning. In T. M. Newcomb (Ed.), New directions in psychology III. New York: Holt, Rinehart & Winston, Inc.
- Meier, R. L. (1962) A communications theory of urban growth. Cambridge: M. I. T. Press.
- Melzack, R. & P. D. Wall (1965) Pain mechanisms: A new theory. Science, 150, 971-9.
- Merkel, J. (1885) Die zeitlichen Verhältnisse der Willensthatigkeit. Philos. Studien (Wundt), 2, 73-127. As cited by J. G. Miller, Living Systems. (forthcoming).
- Milgram, S. (1970) The experience of living in cities. Science, 169, 1461-8.
- Miller, G. A. (1951) Language and communication. New York: McGraw-Hill Book Co.
- Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 63, 81-97. Reprinted in G. A. Miller, The psychology of communication, Baltimore, Maryland: Basic Books, Inc.
- Miller, G. A. (1967) The psychology of communication. Baltimore, Maryland: Basic Books, Inc.



- Miller, G. A., E. Galanter, & K. H. Pribram (1960) Plans and the structure of behavior. New York: Holt, Rhinehart & Winston, Inc.
- Miller, J. G. (1960) Information input overload and psychopathology. Amer. J. Psychiat., 116, 695-704.
- Miller, J. G. (1961) Sensory overloading. In B. E. Flaherty (Ed.), Psychophysiological aspects of space flight. New York: Columbia Univer. Press.
- Miller, J. G. (1962) Information input overload. In M. C. Yovits, G. T. Jacobi, & G. D. Goldstein (Eds.), Self-organizing systems. Washington, D. C.: Spartan Books.
- Miller, J. G. (1963) The individual as an information processing system. In W. S. Fields and W. Abbott, Information storage and neural control. Springfield, Illinois: Charles C. Thomas.
- Miller, J. G. (1963) Living systems and mechanical systems. Preprint 112. Ann Arbor, Michigan: Mental Health Research Institute, Univer. of Michigan.
- Miller, J. G. (1964) Adjusting to overloads of information. In D. McK. Rioch & E. A. Weinstein (Eds.), Disorders of communication. Baltimore, Maryland: Williams & Wilkins.
- Miller, J. G. (1964) Coping with administrators' information overload. J. med. Edu., 29, (11, part 2), 47-54 and 182-189).
- Miller, J. G. (1964) Psychological aspects of communication overloads. In R. W. Waggoner & D. J. Carek (Eds.), International psychiatry clinics: Communication in clinical practice. Boston: Little, Brown & Co.
- Miller, J. G. (1965) The dynamics of information adjustment processes. In J. Masserman (Ed.), Science and psychoanalysis, VIII: Communication community. New York: Grune & Stratton.



- Miller, J. G. (forthcoming) Information input overload. In Living Systems.
- Miller, W. H., F. Ratliff & H. K. Hartline (1961) How cells receive stimuli. Scientific Amer., Sept.
- Montcastle, V. B. (1966) The neural replication of sensory events in the somatic afferent system. In J. C. Eccles (Ed.), Brain and conscious experience. New York: Springer-Verlag.
- Mountcastle, V. B., P. W. Davies, & A. L. Berman (1957) Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. J. Neurophysiol., 20, 374-407.
- Morrell, F. (1967) Electrical signs of sensory coding. In The neurosciences: A study program. New York: Rockefeller Univer. Press.
- Mowbray, G. H. (1954) The perception of short phrases presented simultaneously for visual and auditory reception. Quart. J. exp. Psychol., 4, 86-92.
- Mowbray, G. H. (1960) Choice reaction times for skilled responses. Quart. J. exp. Psychol., 16, 216-23.
- Mowbray, G. H. (1964) Subjective expectancy and choice reaction times. Quart. J. exp. Psychol., 16, 216-23.
- Mowbray, G. H. & M. V. Rhoades (1959) On the deduction of choice reaction times with practice. Quart. J. exp. Psychol., 11, 16-23.
- Neisser, U. (1967) Cognitive psychology. New York: Appleton-Century-Crofts.
- Newcomb, T. M. (1953) An approach to the study of communicative acts. Psychol. Rev., 60, 393-404.
- Olds, J. (1956) Pleasure centers in the brain. Scientific Amer. Oct., 105-16.
- Pollack, I. (1952) The information of elementary auditory displays. J. acoust. Soc. Amer., 24, 745-9.

- Pollack, I (1953) The information of elementary auditory displays, II. J. acoust. Soc. Amer., 25, 765-9.
- Posner, M. I. (1965) Effects of size and location of informational transforms upon short-term retention. J. exp. Psychol., 70:5, 496-505.
- Poulton, E. C. (1956) Listening to overlapping calls. J. exp. Psychol., 52 334-9.
- Quastler, H. & B. Brabb (1956) Human performance in information transmission. Tech. Rep. R-70. Urbana: Control Systems Laboratory, Univer. of Illinois.
- Quastler, H. & V. J. Wulff (1955) Human performance in information transmission. I. Tech. Rep. R-62. Urbana: Control Systems Laboratory, Univer. of Illinois.
- Quastler, H. & V. J. Wulff (1955) Human performance in information transmission. II. Tech. Rep. R-62. Urbana: Control Systems Laboratory, Univer. of Illinois.
- Sanders, A. F. (1967) Some aspects of reaction processes. Acta Psychol., 27, 115-30. Also known as A. F. Sanders (Ed.), Attention and performance. Amsterdam: North-Holland.
- Schacter, S. (1964) The interaction of cognitive and physiological determinates of emotional states. In L. Berkowitz (Ed.), Advances in american social psychology. New York: Academic Press.
- Schiller, P. H. & M. C. Smith (1966) Detection in metacontrast. J. exp. Psychol., 71, 32-39.
- Schipper, L. M. (1967) Extreme probabilities in learning and decision making. J. exp. Psychol., 73(1), 149-51.
- Schroder, H. M., M. J. Driver & S. Streufert (1967) Human information processing. New York: Holt, Rinehart & Winston.

- Shallice, T. & D. Vickers (1964) Theories and experiments on discrimination times. Ergonomics, 7, 37-49.
- Shannon, C. G. (1948) A mathematical theory of communication. Bell Syst. tech. J., 27, 374-423, 623-56.
- Shannon, C. G. & W. Weaver (1949) The mathematical theory of communication. Urbana, Illinois: The Univer. of Illinois Press.
- Sinaiko, H. W. & E.P. Buckley (1957) Human factors in the design of systems. NRL Report 4996, Naval Research Laboratory, Washington, D. C. Reprinted in H. W. Sinaiko (Ed.), Selected papers on human factors in the design and use of control systems. New York: Dover Publications, Inc.
- Skinner, B. F. (1957) Verbal behavior. New York: Appleton-Century-Crofts.
- Sokolov, E. N. (1960) Neuronal models and the orienting reflex. In A. B. Brazier (Ed.), The central nervous system and behavior. New York: Josiah Macey, Jr. Foundation. See also H. Pribram (Ed.), Perception and action, Brain and behaviour 2. Middlesex, England & Baltimore, Maryland: Penguin Books Ltd.
- Sperling, G. (1960) The information available in brief visual presentations. Psychol. Monog., 74, No. 11.
- Swets, J. A. (1962) Learning to identify nonverbal sounds: An application of a computer as a teaching machine. U.S. Naval Training Device Center, Tech. Rep. No. 789-1.
- Swets, J. A. , J. S. Harris, L. S. McElroy & H. Rudloe (1964) Further experiments on computer-aided learning of sound identification. U. S. Naval Training Device Center, Tech. Rep. No. 789-2.
- Swets, J. A., M. P. Tanner, & T. G. Birdsall (1964) Decision processes in perception. In J. A. Swets (Ed.), Signal detection and recognition by human observers: Contemporary readings. New York: John Wiley.

- Tasaki, I. & H. Davis (1955) Electric responses of individual nerve elements in the cochlear nucleus to sound stimulation (guinea pigs). J. Neurophysiol., 18, 151-8.
- Taylor, F. W. (1907) On the art of cutting metals. Transact. A. S. M. E., 28, 31-350.
- Taylor, F. W. (1911) The principles of scientific management. New York.
- Taylor, F. W. (1919) Shop Management. New York.
- Taylor, F. W. (1947) Scientific management. New York.
- Thomas, G. (1941) Experimental study of the influence of vision on sound localization. J. exp. Psychol., 28, 163-77.
- Thurmond, J. B. & E. A. Alluisi (1962) Choice time in reacting to stimulus pairs at various levels of dissimilarity. Marietta, Georgia: Human Factors Research Department, Lockheed-Georgia Co.
- Toffler, A. (1970) Future Shock. New York: Random House.
- Trowbridge, M. H. & H. Cason (1932) An experimental study of Thorndike's theory of learning. J. gen. Psychol., 7, 245-58.
- Vince, M. A. (1949) Rapid response sequences and the psychological refractory period. Brit. J. Psychol., 40, 23-40.
- Wall, P. D., J. Y. Lettvin, W. S. McCulloch & W. H. Pitts (1956) Factors limiting the maximum impulse transmitting ability of an afferent system of nerve fibers. In C. Cherry (Ed.), Information theory, third London symposium. New York: Academic Press.
- Webster, J. C. & P. O. Thompson (1953) Some audio considerations in air control towers. J. Audio Eng. Soc., 1, 171-5.
- Webster, J. C. & P. O. Thompson (1954) Responding to both of two overlapping messages. J. acoust. Soc. Amer., 26, 396-402.

Welford, A. T. (1960) The measurement of sensory-motor performance:  
survey and reappraisal of twelve years' progress. Ergonomics, 3,  
189-230.

Wiener, N. (1948) Cybernetics, or control and communication in the animal  
and the machine. Cambridge, Mass: The M. I. T. Press.

Woodbridge, D. E. (1963) The machinery of the brain. New York & Toronto:  
McGraw Hill Book Co.



87324

P  
91  
.C655  
.T39  
1972  
c.1

## Date Due

[illegible]

FORM 109



