

UNCLASSIFIED

A # 2154 Q/24

CONTROL SYSTEMS LABORATORY

SUGGESTIONS FOR THE ANALYSIS
OF REACTION TIMES AND
SIMPLE CHOICE BEHAVIOR

Report Number R-53

April 1954

Contract DA-36-039-SC-56695

Project 8-103A, D/A Project 3-99-10-101

UNIVERSITY OF ILLINOIS · URBANA · ILLINOIS

UNCLASSIFIED

U N C L A S S I F I E D

1/39

Report Number R-53

SUGGESTIONS FOR THE ANALYSIS OF REACTION
TIMES AND SIMPLE CHOICE BEHAVIOR

April 1954

Prepared by:

Lee S. Christie

R. Duncan Luce

CONTROL SYSTEMS LABORATORY
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS
Contract DA-36-039-SC-56695

Numbered Pages: 39

U N C L A S S I F I E D

Foreword

This paper deals with some aspects of the dynamics of making decisions. The main results seem to be two: (1) The dynamics for some idealized cases have been worked out, and (2) even the simplest assumptions lead to models of so considerable mathematical complexity, that the problem of testing hypotheses becomes extremely delicate. These are important results for anybody concerned with models of human information-processing.

The MS was prepared during the summer study session of 1953. Some experimental results obtained at this laboratory since that time might tend to modify some of the statements in this paper, but will not affect its main theses.

Henry Quastler

Suggestions for the Analysis of Reaction
Times and Simple Choice Behavior

Lee S. Christie¹⁾ and R. Duncan Luce²⁾

1. Introduction

In this paper we propose a model for the way human beings organize the decisions required by simple choice situations into a series of component decisions. It is our thesis that such an organization of decisions must be reflected in the distribution of reaction times and that, therefore, it may be possible to infer the organization from the reaction-time distribution. Although our thinking derives from empirical studies, we must describe this proposal as speculative, for the model is not firmly based on such studies. However, the development of the model has led us to suggest two experiments which we believe may help to determine what merit it has. These experiments will also help to decide whether it is desirable to pursue further work in an attempt to modify the model to accord better with reality, for we have little hope that the particular details of the present model have any lasting value.

2. Reaction Times

Suppose that a subject is stimulated at time 0 in a situation demanding that he make some decision and he responds at time t . The time interval, t , between the stimulus and the response is called the disjunctive reaction time. It is clear

1) Department of Systems Analysis, Tufts College, Medford, Mass.

2) Bureau of Applied Social Research, Columbia University, New York, N.Y.

6/39

that to obtain stable and readily analyzable time distributions it is necessary that the stimulus be simple enough so that the mean latency is no more than a second or two. Otherwise an unwanted stimulus may intervene between the test stimulus and the response and the interaction between the stimuli will cause a distortion of the time distribution which will be very difficult to analyze.

The study of reaction times, including disjunctive reaction times, has a long history in the literature of psychology (cf. Woodworth [15] ch. XIV). In recent years, however, relatively little interest has been evident in reaction-time studies. We may attribute this loss of interest to two related causes. First, there has been a failure to separate the time to make a choice from the nonchoice time lags involved in the total process. One attempt to make this separation involved measuring the subject's response to a stimulus when no choice was to be made and subtracting this time from the time required to respond to the same stimulus when a choice was involved. This technique has been considered unsatisfactory for the following reason: If the subject has no choice to make he is able to bring his motor readiness for the specified response to a much higher pitch than he can when he is required to make a disjunctive reaction and thus the base time (the time to react in a choice situation excluding the time for the choice itself) cannot be equated to the nonchoice reaction time. We may conclude that the base time will be determined, if at all, only from measurements taken when the subject is required to choose.

Second, suppose that in one way or another the pure choice distribution of reaction times has been obtained, then what? It is true that if these distributions were found to be extremely simple (in that they could be well approximated by some class of elementary mathematical functions) the separation of nonchoice from choice latencies could be an end in itself. If, however, the resulting choice distribution were of a complex character, the challenge to account for it in more primitive terms would remain.

We describe these as related difficulties, for it is not unreasonable to suppose that the method used to tease out the nonchoice latencies (base times) can also be used, or adapted, to decompose the choice latencies into more primitive terms. Such a decomposition of the observed latency distribution may be an entirely formal mathematical process with no empirical correlate or it may be based on a model which purports to describe the way a human being composes the finally observed decision from certain more elementary ones. It is with such a model that we are concerned.

At the heart of our proposal is the idea that the mathematical technique of the Laplace transform may be employed usefully in the study of reaction times. Since it is unlikely that every one of our readers will be familiar with the Laplace transform, we have devoted the next section to its definition and to a list of those of its elementary properties which we shall need.

3. The Laplace Transform

Let F be a real-valued function of a real variable t

8/39

such that $F(t) = 0$ for $t < 0$. The real-valued function $L(F)$ of the real variable s defined by the equation

$$L(F) = \int_0^{\infty} e^{-st} F(t) dt \quad (1)$$

is called the Laplace transform of F . There is essentially no loss of information about F in making this transformation (see eq. 4), but because of some of the special properties of the transform there is sometimes a distinct advantage to working with transformed functions. We shall list a few of the elementary properties of the transform which we shall need later; no proofs will be given since they may be found in Churchill [3].

$$i. \quad L \left\{ \int_0^t F_1(\tau) F_2(t-\tau) d\tau \right\} = L(F_1) L(F_2) \quad (2)$$

$$ii. \quad L \left(\frac{dF}{dt} \right) = sL(F) - F(0) \quad (3)$$

iii. If $L(F) = L(G)$, then $F = G + N$ where N is some function with the property that $\int_0^t N(t) dt = 0$ for all $T > 0$. If it is known that F and G are continuous, then N is continuous and so $N(t) \equiv 0$, i.e., $F = G$. (4)

iv. If a and b are constants, (5)

$$L(aF+bG) = aL(F) + bL(G).$$

v. If $F(t) = \lambda e^{-\lambda t}$, where λ is a constant, then (6)

$$L(F) = \frac{1}{\frac{s}{\lambda} + 1}$$

4. The Model

Our proposal is based on assumptions which are intuitively

acceptable, but which do not appear to be susceptible of direct verification. It is our impression that any empirical verification of the model must deal with the full set of assumptions rather than with each in isolation.

I. Our first major assumption is that it is possible, for a given experimental situation, to divide the observed reaction time t into two components t_b and t_c , called base time and choice time respectively, such that

1. $t = t_b + t_c$;

2. the value of t_b depends only on the mode of stimulus presentation and on the motor actions required of the subject and, specifically, it is not directly dependent on the character of the choice demanded;

3. the value of t_c depends only on the choice demanded and, specifically, it is not directly dependent on the stimulus mode or on the motor actions required.

Let the distributions of t , t_b , and t_c be denoted by f , f_b , and f_c respectively. Since conditions 2 and 3 imply that the two component reaction times are independent for a fixed experimental situation, it follows from 1 that

$$f(t) = \int_0^t f_b(\tau) f_c(t-\tau) d\tau \quad (7)$$

Our second major assumption concerns only the choice reaction times and requires the distribution f_c to be composed from more elementary distributions. The basic idea is that the final decision made by a person is organized into a set of simpler decisions which are, in some sense, elementary decisions built into him. If such a structure exists in human

10/39

decision making, it is analogous to the structure of a decision process in a computing machine, as composed from a set of decisions which are elementary relative to that machine, i.e., the elementary decisions built into the machine by the engineer. The actual organization of these elementary decisions to form a more complex one is a function both of the individual man or machine and of the nature of the decision being made. This is true for the machine at least, and we shall suppose it is true of people. In addition, the breakdown of a complex decision is not in general restricted to a serial process where one elementary decision is followed by another, for in a machine different portions may be simultaneously employed on different parts of the problem. There seems every reason to suppose this is also true of men.

We shall describe the organization of decisions by what is called in mathematics a directed graph (the terms oriented graph and network have also been employed in mathematics and the term flow diagram is used in connection with computer coding). A directed graph consists of a finite set of points, which are called nodes, with directed lines between some pairs of them. Several examples are shown in Fig. 1. It is possible, in general, for more than one directed line to connect two points, both in the sense that we may have two or more in the same direction as in Fig. 2a, and in the sense that there may be lines with the opposite direction, as in Fig. 2b.

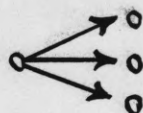
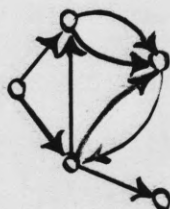


Figure 1.



a



b

Figure 2.

In this paper when we use the term, directed graph, we shall suppose that neither of these possibilities is allowed, that is, we shall suppose that between any pair of nodes there is at most one directed line.

We shall employ a directed graph to represent the organization of decisions in the following way: At each node we shall assume that an "elementary decision" will take place, the time distribution governing the decision at node i being denoted by f_i . The decision process is initiated at node i when, and only when, decisions have been made at each of those nodes j such that there is a directed line from j to i . We may think of the "demon" at node i waiting to begin making his decision until he has received the decisions of all the demons who precede him in the directed graph.

For the directed graphs we shall consider, there will be at least one node, possibly more, which is the terminal point of no line; these will be the decision points which are initiated by the experimental stimulus at time 0. There will also be at least one node, and again possibly more, which initiates no directed line and it is only when the decisions at

12/39

all these nodes have been taken that the motor actions (which signal the subject's response to the experimenter) are begun.

It is clear that for any individual and for any stimulus situation it is possible to find at least one directed graph N and elementary latencies f_i which compose as described above to give f_c . For example, let N have but one node and let $f_i = f_c$. We shall, however, make stringent assumptions about N and f_c which in general exclude this trivial solution. It is some of these assumptions which most likely will be abandoned or modified if the present model cannot cope with experimental data.

Let $S(b)$ denote the set of choice situations which all have the same base time distribution f_b .

II. Our second major assumption is that it is possible to find for each stimulus situation σ of $S(b)$ a subset $R(b)$ of $S(b)$ and a latency f_e such that

1. σ is an element of $R(b)$;
2. for each choice situation ρ in $R(b)$ there exists a directed graph N_ρ with the properties
 - a. each of the latency distributions at the nodes is the same, namely, f_e .
 - b. the decision time at node i is independent of that at node j , $j \neq i$,
 - c. f_c is a composition of N_ρ and f_e (as described above);
3. among the stimulus situations in $R(b)$ there is one whose directed graph satisfying condition II.2 is a single point.

In less formal terms, we require that there be groups of stimulus situations all of which have the same base time dis-

tribution and which can be built up according to a directed graph from elementary and independent decisions which all have the same latency distribution f_e . In addition, among the stimulus situations in this class we assume that there is one which employs but a single elementary decision. The latter assumption can be weakened, if we choose, to the assumption that there is one stimulus situation whose directed graph we know a priori, but in what follows we shall take the stronger form that the graph is a single point.

5. Comments

The above assumptions comprise the formal structure of our model; however, there are a series of auxiliary comments which are necessary.

Even if we were able to show that these assumptions can be met for certain wide classes of experimental data, but that in so doing we obtain elementary decision distributions f_e which are extremely complicated, it is doubtful that we should accept the model as an adequate description of the decision process. Equally well, if the directed graphs required are excessively complex we should reject the model. The hope is that it is possible to subdivide the total process into a relatively small set of subprocesses which are practically identical. But we do not want the analysis to be pushed down to the level of neurone firings. Assumption II.3 effectively prevents this extremity by requiring the existence of a stimulus situation which demands but one elementary decision for its response.

It is also implicit in our thinking (although not a part of the formal model) that the sets $R(b)$ of "similar" stimulus situations will include as subsets those experimental situations we naturally think of as being similar. For example, suppose the subject is presented with n points, one of which is colored differently from the others and he is required to signal the location of that one. We should want to consider the set of these situations generated as n ranges over the smaller integers as "similar", and we should probably reject the model if they could not be put in the same set $R(b)$, even if by great ingenuity we were able to find other less intuitively simple sets of situations for which the model held.

When the model is applied to experimental data we anticipate that the case of the directed graph being a single point will be identified with the intuitively "simplest" choice situation within the set of "similar" ones. This may in general prove to be the situation which involves one binary decision in the sense of information theory, i.e., it will be that situation requiring a choice between two equally likely alternatives [4, 5, 12].

In some of the following sections we shall make the following explicit assumption as to the form of f_e

$$f_e(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

where λ is a positive constant. There are two grounds for supposing this might be an appropriate assumption. First, let us suppose that when no decision has been reached by time

t following stimulation at time 0 then the probability the decision will be reached between t and $t + \Delta t$, where Δt is small, is approximately proportional to Δt , with a constant of proportionality λ . In this case, it is not difficult to show that the distribution of decisions is exponential [1, 2]. Whether this assumption is correct is an empirical problem, but it must be admitted that it has the virtue of simplicity. Second, and probably more relevant, it is a relatively common observation that as certain decision situations are made more and more simple, the observed latency is better and better approximated by an exponential distribution slightly displaced from the origin [2, 7]. The main error is generally on the rising limb. If this change toward simplicity is actually toward a directed graph consisting of one point, and if our other assumptions hold, then it seems plausible that the elementary decision latency is actually exponential but that the observed distribution is smeared by the convolution of the base time distribution and the decision time distribution.

6. The Problem

Let R be a set of choice situations which are presumed to satisfy the assumptions of the model, i.e., R is a set of the type $R(b)$ described in assumption II. Let f_{σ} denote the reaction time distribution associated with a typical member of R . The problem is then to find distributions f_b and f_e and a set of directed graphs N_{σ} , where σ ranges over R , such that each of the triples (f_b, f_e, N_{σ}) when composed according to the assumptions of section 4 yield the distribution f_{σ} . There may, of course, be no, one, or many solutions to this

problem, but one hopes that by an appropriate choice of R there will be exactly one solution.

It would appear that if the problem is to be solved in any degree of generality, it must be attacked somewhat indirectly. It may prove appropriate to solve first the following problem: Given a continuous distribution f , find the set of all triples (f_b, f_e, N) , where f_b and f_e are continuous, which satisfy the assumptions and which compose to form f . It seems very plausible to suppose that in general there are many solutions to this problem. However, if f and f' are two distributions associated with choice situations from the same set R , then it will be necessary to accept only those triples with the same f_b and f_e present in both cases. Further stimulus situations should serve further to restrict the possibilities.

These problems will not be attacked, let alone solved, in this paper; they appear to be of considerable difficulty. We know of only one important lead in this direction, but we have not investigated it. In recent years electrical engineers have been concerned with the problem of synthesizing in a systematic manner electrical networks to have a given transfer function. If we identify the given reaction time distribution with the transfer function, the graph N with the electrical network, and f_e with component characteristic, there is an analogy between the two problems. This is probably worth investigation, but it is almost certain that solving our problem will prove to be a major research undertaking.

To some extent the problem we pose may be simplified by using some of our assumptions and the Laplace transform. Let f_{σ} be the observed distribution of reaction times for a given stimulus situation σ , then by assumption II we know there exists a set $R(b)$ which includes σ and another stimulus situation whose directed graph consists of one point. Let f_1 denote the distribution of reaction times in the latter case. From assumption I we may write

$$f_{\sigma}(t) = \int_0^t f_b(\tau) f_c(t-\tau) d\tau$$

$$f_1(t) = \int_0^t f_b(\tau) f_e(t-\tau) d\tau$$
(8)

Taking the Laplace transform in each case and applying eq. 2,

$$L(f_{\sigma}) = L(f_b) L(f_c)$$

$$L(f_1) = L(f_b) L(f_e)$$
(9)

If we divide the first equation by the second in eq. 9 we obtain

$$\frac{L(f_{\sigma})}{L(f_1)} = \frac{L(f_c)}{L(f_e)}$$
(10)

This is a fairly crucial consequence of our assumptions, for it is seen that all mention of the base time has been eliminated. It is an equation relating the empirical data to f_e and N_{σ} .

At this point we should raise an important practical problem. Empirically, one does not obtain estimates of the distribution f , but rather approximations to the cumulative distribution

18/39

$$F(t) = \int_0^t f(\tau) d\tau$$

(Throughout we shall use small Latin letters to denote distributions and the corresponding capitals to denote the cumulatives.) Now, while approximations to F may be reasonably accurate, it is well known that numerical differentiation of data tends to magnify errors and is therefore to be avoided; so the question arises whether we can translate our results, in particular eq. 10, into statements about the cumulative distributions. From eq. 3 we have

$$L(f) = sL(F) + F(0)$$

Since we are speaking of empirical data we may assume $F(0) = 0$, and so eq. 10 becomes

$$\frac{L(F_1)}{L(F_2)} = \frac{L(f_1)}{L(f_2)} \quad (11)$$

Having eliminated f_b from our discussion, the problem of determining it remains. Since our division in eq. 11 assumes f_b is the same in the several cases, it will suffice to determine it from any one. The simplest, of course, is the case where the graph consists of one point, in which case

$$L(f_b) = \frac{L(f_1)}{L(f_e)} = \frac{L(F_1)}{L(F_e)} \quad (12)$$

As an example of how eq. 12 may be used, suppose f_e is exponential with time constant λ . Then by eq. 6,

$$L(f_e) = \frac{1}{\frac{s}{\lambda} + 1}$$

and so eq. 12 becomes

$$L(f_b) = \frac{s}{\lambda} L(f_1) + L(f_1)$$

If we make the reasonable assumption that $f_1(0) = 0$, then from eqs. 3 and 5 we find

$$L(f_b) = \frac{1}{\lambda} L\left(\frac{df_1}{dt}\right) + L(f_1) = L\left(\frac{1}{\lambda} \frac{df_1}{dt} + f_1\right)$$

Assuming that f_b is continuous and that f_1 has a continuous derivative, eq. 4 implies

$$f_b = \frac{1}{\lambda} \frac{df_1}{dt} + f_1$$

or integrating from 0 to t ,

$$F_b = \frac{1}{\lambda} f_1 + F_1$$

Since f_1 must be determined from empirical data, it is clear from eq. 13 that considerable data will be necessary to obtain accurate estimates of F_b .

7. Serial Decision Process

An alternative program to solving the general problem discussed in section 6 is to discover the consequences of certain explicit assumptions about the directed graph N and the elementary latency f_e . The results of this alternative program will, unfortunately, be much weaker than a solution of the general problem, but they may have considerable heuristic value. We may choose such extra assumptions on intuitive grounds, with the hope that they may be relevant for some experimental data. We shall examine two cases which are, in a sense, the two most extreme forms of the directed graph N .

have checked up to here

20/39

The first, the topic of this section, is the general serial case shown in Fig. 3a, and the second, which will be discussed in section 8, is the parallel case shown in Fig. 3b.

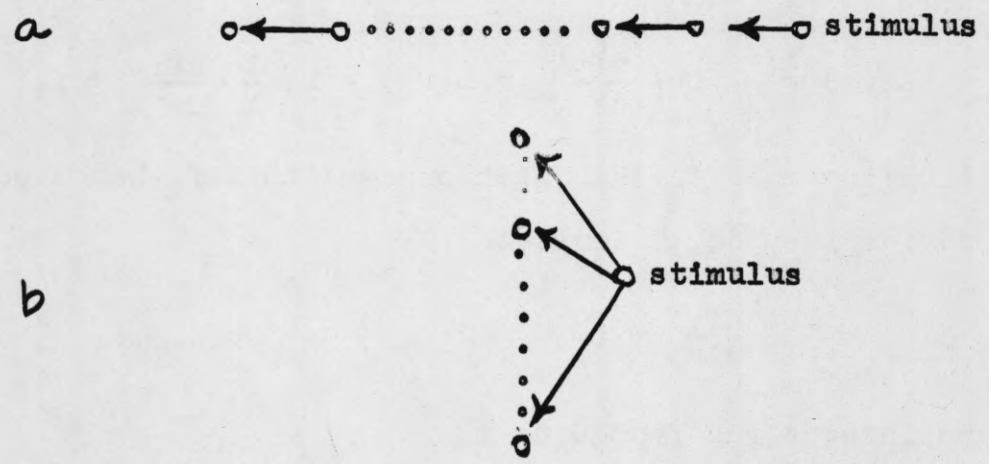


Figure 3

It follows immediately from assumptions I and II.2.b that the observed distribution f_n of a serial process having n nodes

is given by (14)

$$f_n(t) = \int_0^t \dots \int_0^{t_3} \int_0^{t_2} f_b(t_1) f_e(t_2-t_1) \dots f_e(t-t_n) dt_1 dt_2 \dots dt_n$$

Applying the Laplace transform to eq. 14 and using eq. 2 we have

$$L(f_n) = L(f_b) L(f_e)^n \tag{15}$$

or dividing by the case $n = 1$,

$$\frac{L(f_n)}{L(f_1)} = L(f_e)^{n-1} = \frac{L(F_n)}{L(F_1)} \tag{16}$$

Eq. 16 is the explicit form of eq. 11 for the serial case.

Clearly, if we have given numerical data we may solve (possibly numerically) for f_e for each value of n .

As an example of how this might be done when we know the general form of f_e , suppose f_e is exponential with the time constant λ . In that case, eq. 16 becomes

$$\frac{L(F_n)}{L(F_1)} = \frac{1}{\left(\frac{s}{\lambda} + 1\right)^{n-1}} \quad (17)$$

In Fig. 4 we have presented plots of $\frac{1}{\left(\frac{s}{\lambda} + 1\right)^n}$ vs. $\frac{s}{\lambda}$ for small values of n .

A second equation may be obtained by observing that the mean, $\mu_1(n)$, of a serial process with n exponential elementary decisions is given by

$$\mu_1(n) = \mu_1(b) + \frac{n}{\lambda} \quad (18)$$

where $\mu_1(b)$ is the mean base time. Thus,

$$\mu_1(n) - \mu_1(1) = \frac{n-1}{\lambda} \quad (19)$$

We may now use eqs. 17 and 19 to attempt to decide whether a given set of data is adequately fit by the assumptions of the model, plus the added assumptions of a serial directed graph and exponential elementary latencies. There are serious statistical questions as to how this may best be done, but the following ready method may suffice until the statistical problems are formulated and solved. From the data we compute $L(F_n)/L(F_1)$ as a function of s ; this we may assume is in the form of a plot, which we shall call plot A. For each (reasonable) value of n and for some value of $\frac{s}{\lambda}$, say $\frac{s}{\lambda} = \frac{1}{2}$, find

in Fig. 4 the corresponding value of $\frac{1}{\left(\frac{s}{\lambda} + 1\right)^{n-1}}$. We know from eq. 17 that this must equal $L(F_n)/L(F_1)$ if our assumptions are correct and if the correct value of n has been chosen. We thus enter plot A at this point and determine the value of s . Since we selected $\lambda = 2s$, this determines λ . But eq. 19 presents a relation between the observed means, λ , and n which will be satisfied if our assumptions are valid. We choose the value of n such that the error between the observed means (the left side of eq. 19) and $\frac{n-1}{\lambda}$ is a minimum; this yields the best possible fit at the point $\frac{s}{\lambda} = \frac{1}{2}$ for the model with the added assumptions of a serial graph and exponential f_e . Using these values of λ and n one may add the theoretical curve $\frac{1}{\left(\frac{s}{\lambda} + 1\right)^{n-1}}$ vs. s to plot A and a comparison between the two curves will give some indication of the adequacy of the assumptions. Clearly, a less subjective criterion of the quality of this fit is needed.

8. Parallel Decision Process

If we suppose that the n elementary decision processes are carried out in parallel (see Fig. 3b), the choice latency distribution is the distribution of the largest of n selections, one from each of the elementary distributions. This is known to be given by

$$\frac{d}{dt} \prod_{i=1}^n F_i(t)$$

which in the case all the elementary distributions are the same, namely F_e , reduces to

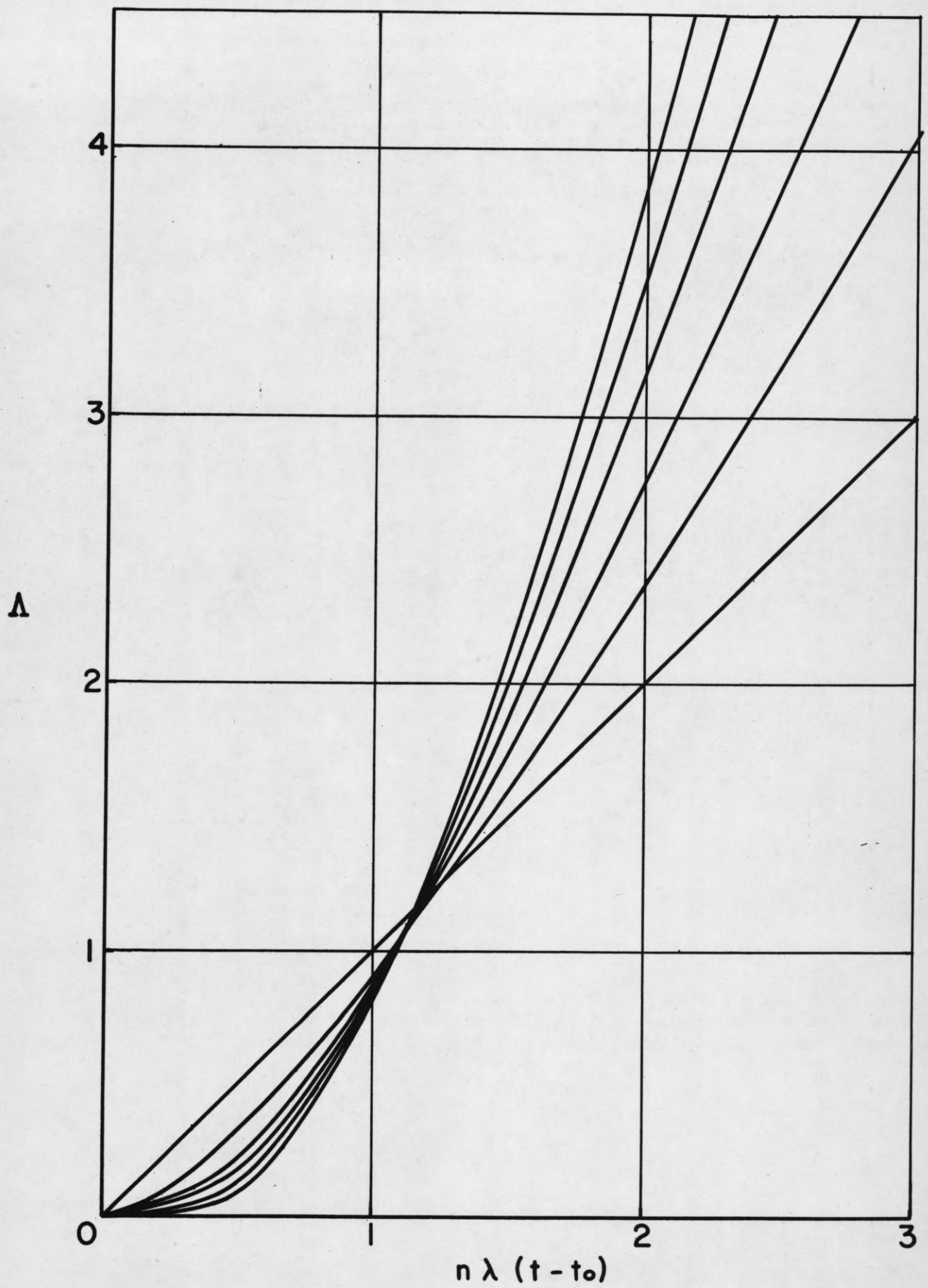


FIG. 4 Δ FOR SERIES CASE WITH MEAN 1.

$$nf_e(t) [F_e(t)]^{n-1} \quad (20)$$

If we denote the observed reaction time distribution for the parallel case by g_n , then it follows from eq. 7 that

$$g_n(t) = \int_0^t f_b(\tau) nf_e(t-\tau) [F_e(t-\tau)]^{n-1} d\tau \quad (21)$$

Applying the Laplace transform and eq. 2,

$$L(g_n) = L(f_b) L(nf_e F_e^{n-1}) \quad (22)$$

As before, we may divide by $L(g_1)$ to eliminate $L(f_b)$.

To proceed further, we assume f_e is exponential, then

$$\begin{aligned} L(nf_e F_e^{n-1}) &= n\lambda \int_0^{\infty} e^{-st} e^{-\lambda t} [1 - e^{-\lambda t}]^{n-1} dt \\ &= n\lambda \int_0^{\infty} e^{-(s+\lambda)t} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k e^{-k\lambda t} dt \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{1}{\frac{s}{\lambda} + k + 1} \end{aligned}$$

To evaluate the above sum, consider the function

$$\phi(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k x^{\frac{s}{\lambda}} x^k = x^{\frac{s}{\lambda}} (1-x)^{n-1}$$

Observe that

$$\begin{aligned} n \int_0^1 \phi(x) dx &= n \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \int_0^1 x^{\left(\frac{s}{\lambda} + k\right)} dx \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{1}{\frac{s}{\lambda} + k + 1} \end{aligned}$$

$$= L(n f_e F_e^{n-1})$$

And that

$$\begin{aligned} n \int_0^1 \phi x dx &= n \int_0^1 x^{\frac{s}{\lambda}} (1-x)^{n-1} dx \\ &= n B\left(\frac{s}{\lambda} + 1, n\right) \\ &= n \frac{\Gamma\left(\frac{s}{\lambda} + 1\right) \Gamma(n)}{\Gamma\left(\frac{s}{\lambda} + n + 1\right)} \end{aligned}$$

where $B(m, n)$ is the Beta function and $\Gamma(n)$ is the Gamma function.

From these results we easily obtain

$$\begin{aligned} \frac{L(g_n)}{L(g_1)} &= \frac{n B\left(\frac{s}{\lambda} + 1, n\right)}{B\left(\frac{s}{\lambda} + 1, 1\right)} \\ &= \frac{n! \Gamma\left(\frac{s}{\lambda} + 2\right)}{\Gamma\left(\frac{s}{\lambda} + n + 1\right)} \end{aligned}$$

In Fig. 5 we have presented plots of $\frac{n! \Gamma\left(\frac{s}{\lambda} + 2\right)}{\Gamma\left(\frac{s}{\lambda} + n + 1\right)}$ vs. $\frac{s}{\lambda}$ for small values of n .

The mean of the parallel process can be shown to be given by $\mu_1(n) = \mu_1(b) + \frac{1}{\lambda} \sum_{i=1}^n \frac{1}{i}$ and thus we have, as in the serial case, a second relation which must be met

$$\mu_1(n) - \mu_1(1) = \frac{1}{\lambda} \sum_{i=2}^n \frac{1}{i} \quad (24)$$

The procedure for curve fitting is the same as described

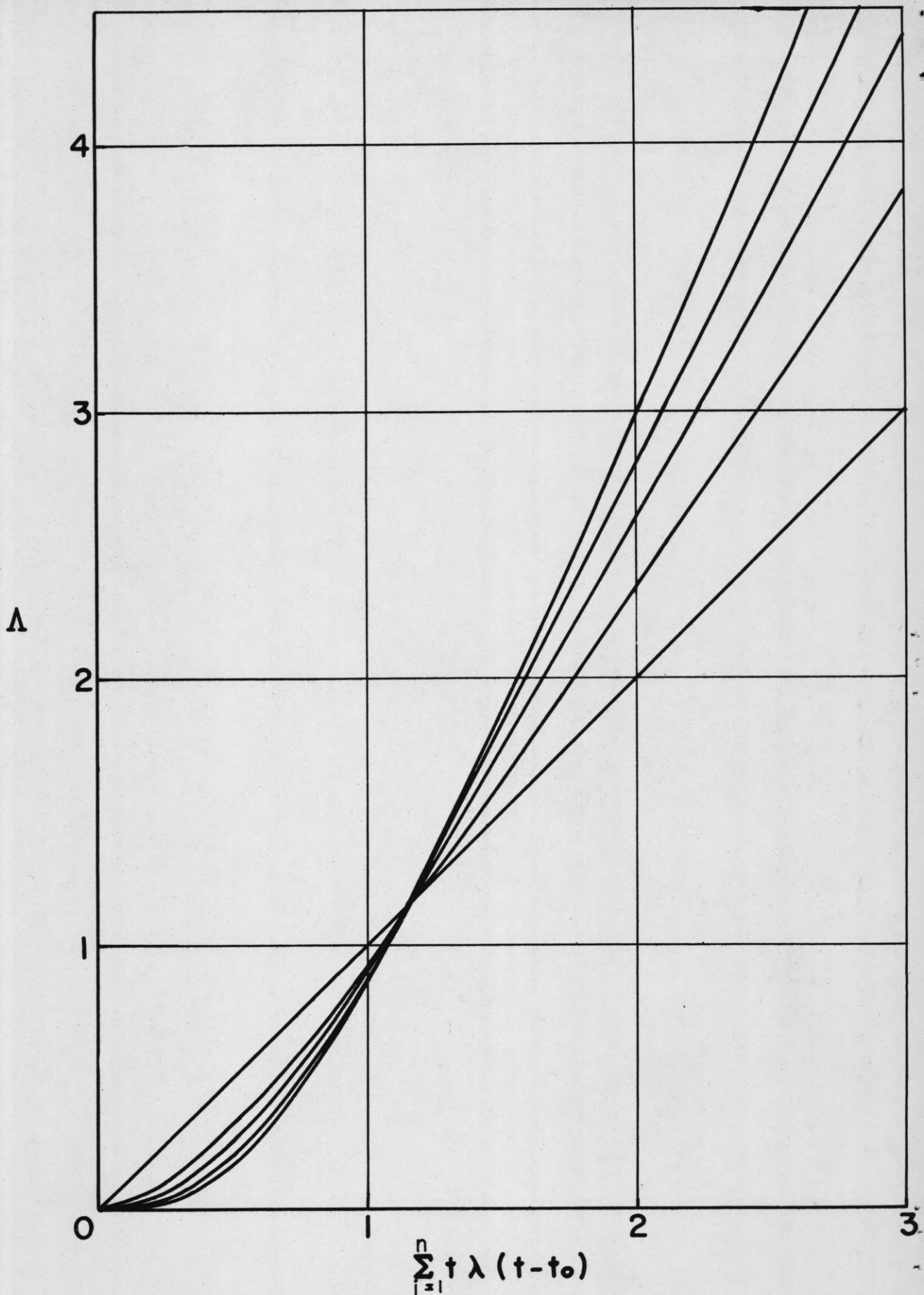


FIG. 5 Λ FOR PARALLEL CASE WITH MEAN 1

for the serial case except that $\frac{s}{\lambda} = 1$ seems to be a more favorable place to enter the graph than is $\frac{s}{\lambda} = \frac{1}{2}$.

9. Model Selection

Without a solution to the general problem described in section 6, there arise statistical problems as to how well a particular set of assumptions, such as serial directed graph and exponential f_e , fit the data and whether another set of similar assumptions are better, or not. In addition, within any one set of assumptions there are undetermined constants, such as λ and n , and there is a question as how best to choose them. We have indicated one procedure (end of section 7) to determine the constants, but it certainly is not clear that this is in any way optimal.

The difficulty of making a selection among different sets of assumptions is evidently quite serious for it can be seen from Figs. 4 and 5 that for almost any small value of n in one there is an n' in the other such that the two curves are fairly similar. Presumably, any other directed graph will produce curves which, in some sense, lie between these two extreme cases. Thus, the shape of the empirical data curves will not be extremely revealing of the proper directed graph to use -- an unfortunate situation.

It is clear that there a number of difficult statistical problems here, but in all likelihood it will prove to be more efficient first to do some experimental exploring using subjective judgements as to goodness-of-fit before trying to formulate and to solve the statistical problems.

10. The Perceptual Moment

In section 2 we observed that in reaction time studies the mean reaction time should be of the order of one second if unwanted interactions with other stimuli are to be avoided. This means that the data will be in a range where certain peculiar phenomena have been observed. One group of phenomena involved periodicities in data of the order of 1/10 of a second and unexpected results from flicker experiments when short groups of slowly (10 to 30 cycles per second) flickering light are administered. It has also been noted that the amount of information, in the sense of information theory, reported by a subject remains approximately constant if the period of stimulus presentation is varied from a few milliseconds to as much as half a second. To explain these observations, it has been proposed that a subject processes information very rapidly at certain discrete times and that he is in a refractory period between them. The period from the beginning of one such hypothetical event to the beginning of the next has been termed the perceptual moment by Stroud [13, 14]. Unfortunately, relatively little direct experimentation has been conducted on this problem, and so it is not possible at this time to give a formal statement of the properties of the moment. Indeed, there are investigators who doubt its existence. In the case that it does exist, our analysis will be applied to situations where it most probably will have an effect. It is therefore of interest whether the analysis can be adapted to cope with it. In this section we shall make a simple hypothesis as to the nature of the moment, not with

any belief that it is correct, but only to indicate that the general features of the analysis remain unchanged.

Let us assume the moment is of fixed duration of, say δ seconds, and that while a person may receive information at any time during that period it will only serve as a stimulus at the end of the period. Furthermore, we will assume that all intermediate (elementary) decisions occur at multiples of δ . Since there is no correlation between the stimulus presentation and the timing of the moment, we may assume the stimulus is presented according to a uniform distribution h in the interval 0 to δ . This assumption may be inappropriate in that a person may be able to assimilate information during only part of the moment; we shall return to this point later.

The question now arises as to the discrete form we should assume for the elementary decision process. In the continuous case we took it to be exponential, and so we shall use the discrete analogue. We assume that if no decision has been reached by the i th moment following the presentation, i.e. at time $i\delta$, then the probability of a decision in the i th moment is $\lambda\delta$. If we call the probability of a response by the i th moment P_i , then

$$\begin{aligned} P_i &= P_{i-1} + [1 - P_{i-1}] \lambda \delta \\ &= (1 - \lambda \delta) P_{i-1} + \lambda \delta \end{aligned} \tag{25}$$

With the initial condition $P_0 = 0$, the difference eq. 25 is solved by

$$P_i = 1 - (1 - \lambda \delta)^i.$$

The probability of a decision in the i^{th} moment is obviously

$$[1 - P_{i-1}] \lambda \delta ;$$

hence we have

$$\lambda \delta (1 - \lambda \delta)^{i-1} \quad (26)$$

as our distribution f_e .

If we replace this discrete distribution, eq. 26, by a continuous one ϕ_e which has rectangles of width ϵ and height $\frac{\lambda \delta (1 - \lambda \delta)^{i-1}}{\epsilon}$ centered about the point $i \delta$, then it is clear that in the limit as $\epsilon \rightarrow 0$ this becomes the discrete distribution.

Let the base time distribution be denoted by f_b as before, then the observed data in the discrete serial case is given by

$$f_n(t) = \lim_{\epsilon \rightarrow 0} \int_0^t \dots \int_0^t \int_0^t f_b(t_1) h(t_2 - t_1) \phi_e(t_3 - t_2) \dots \phi_e(t - t_{n+1}) dt_1 \dots dt_{n+1} \quad (27)$$

Applying the Laplace transform and using eq. 2,

$$L(f_n) = \lim_{\epsilon \rightarrow 0} L(f_b) L(h) L(\phi_e)^n = L(f_b) L(h) \left[\lim_{\epsilon \rightarrow 0} L(\phi_e) \right]^n \quad (28)$$

Observe,

$$\begin{aligned} L(\phi_e) &= \int_0^{\infty} e^{-st} \phi_e(t) dt \\ &= \sum_{i=1}^{\infty} \int_{i\delta - \frac{\epsilon}{2}}^{i\delta + \frac{\epsilon}{2}} e^{-st} \frac{\lambda \delta (1 - \lambda \delta)^{i-1}}{\epsilon} dt \end{aligned}$$

$$\begin{aligned}
&= \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon} \sum_{i=1}^{\infty} \lambda \delta (1-\lambda \delta)^{i-1} e^{-is\delta} \\
&= \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon} \lambda \delta (1-\lambda \delta)^{-1} \sum_{i=1}^{\infty} \left\{ (1-\lambda \delta) e^{-s\delta} \right\}^i
\end{aligned}$$

But,

$$\lim_{\epsilon \rightarrow 0} \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon} = 1,$$

so,

$$\lim_{\epsilon \rightarrow 0} L(\phi_\epsilon) = \frac{\lambda \delta e^{-s\delta}}{1-(1-\lambda \delta) e^{-s\delta}}$$

Substituting in eq. 28 and dividing by the case $n = 1$, we

have

$$\frac{L(f_n)}{L(f_1)} = \left[\frac{\lambda \delta e^{-s\delta}}{1-(1-\lambda \delta) e^{-s\delta}} \right]^{n-1} \quad (29)$$

which is the crucial equation for the discrete serial case.

The mean of the discrete distribution f_ϵ is given by

$$\sum_{i=1}^{\infty} i \delta \lambda \delta (1-\lambda \delta)^{i-1} = \frac{1}{\lambda}$$

Thus, the relation between observed means is

$$\mu_1(n) - \mu_1(1) = \frac{n-1}{\lambda} \quad (30)$$

Now, if we know the value of δ , i.e., the length of the moment, then these two sets of equations may be used in exactly the same fashion as were eqs. 17 and 19 of section 7. We have no theoretical value of δ , so it will be necessary to perform independent measurements of it. It is clear that if the per-

ceptual moment is a real phenomenon it will be important to ascertain its properties prior to analyzing experiments on reaction time.

One further comment of some interest: If we ignore f_b and let $n = 1$, the convolution of h and ϕ_ϵ , when $\epsilon \rightarrow 0$, is a step function such as that shown in Fig. 6. The convolution

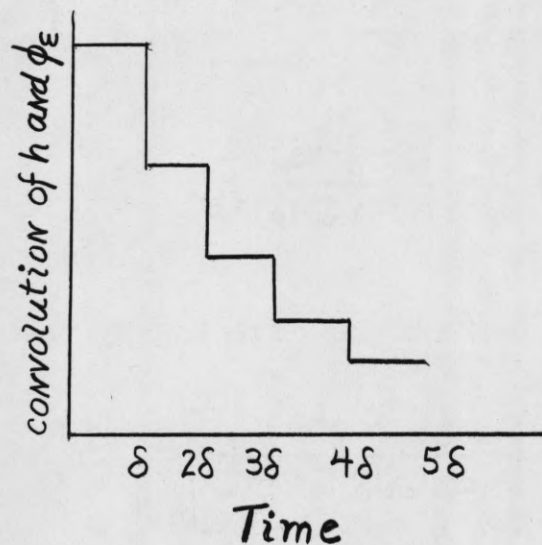


Figure 6

of this function with f_b , for reasonable f_b , will serve to smear the steps but it will not utterly destroy them. Smearing will also result if n is larger than 1, the amount depending on the value of n . Thus, if our assumption as to the moment is roughly correct, we should expect, at least for comparatively simple situations, to find the observed latency distribution somewhat lumpy. Indeed, in the literature [15] it has been remarked not only that the data are lumpy but that there is an oscillation superimposed on the distribution curve. This effect could easily be obtained analytically if we were to assume h uniform over only a small portion of the interval 0 to δ , in other words, if we assume the vast majority of the

moment is truly a refractory period during which there is no intake of information.

These considerations bring out even more strongly the need for comprehensive experiments to determine the properties of the moment.

We shall not attempt, as before, to study the parallel case. The reasons are that the mathematical problem is rather complex and with so little information on the nature of the moment it hardly seems worthwhile to carry out the analysis. Furthermore, we are of the opinion that it is unlikely that information accepted in different moments is dealt with other than serially. It may happen, however, that the information accepted in one moment is processed in parallel; we shall return to this point in the next section.

11. A Possible Application

With the advent of information theory [12], there has been considerable effort expended by some psychologists to determine the information-processing capabilities of human beings [6,8,9,19,11]. We shall not attempt to summarize any of this work except to note that apparently some investigators had expected that a plot of maximum rate of information transmission as a function of the amount of information in the stimulus would be approximately constant. This is now known to be false,* rather, there is a maximum in the curve in the range of 5 to 10 bits. Let us examine this.

*) Recent experiments tend to confirm the approximate constancy of amounts of information transmitted over considerable ranges of stimulus information. Transmission rates of more than 20 bits were observed. H.Q.

To determine the maximum rate of information transmission is but to make a measurement; to hypothesize that this rate is constant as a function of information in the stimulus is to make a model about human behavior. This point does not seem to have been fully realized. There appear to be several different models one might propose each of which would predict a constant rate; probably the most intuitive can be obtained from the interpretation of information theory given by Fano [4,5]. This model supposes, in effect, that a person divides the possible stimulus situations into two approximately equally likely sets, decides which set the given stimulus event is in, divides that set in half, and so on until the event is isolated. If there are n possible occurrences and the probability of the i^{th} is p_i , then the expected number of halving operations is given by $H = - \sum_{i=1}^n p_i \log_2 p_i$.

Consider an experiment in which a subject is required to locate a point in a \sqrt{N} by \sqrt{N} array, assuming the point has an equal chance of occurring in any of the N cells. The above model says that a subject would continue halving the array until a single cell was isolated. This would require $\log_2 N$ halving operations and a decision associated with each as to which half the point were in. If this model were a correct description of a human being's organization of the matter, then, except for the relatively minor effect of the base time, the rate of information transmission should be independent of the value of N . This is not the case, and, indeed, anyone who has tried consciously to execute this sequence of halvings on a large array, say $N \gg 100$, is well aware of

its impracticability.

We propose that by studying the reaction-time distributions of such experiments with different values of N according to the methods suggested above, it may be possible to learn how human beings organize such a decision process. This knowledge, in turn, may allow the prediction of information transmission rates, a prediction which is not possible at present.

While it is irrelevant to our main topic, it may be of interest to pursue the above example a little further. From subjective observations it does not seem unreasonable to suppose that a person halves the grid until he has it in quarters, i.e. up to two bits of information. If this has not isolated the point, then it seems reasonable to suppose he knows which quarter it is in and therefore the corner of the grid nearest to the point. Let us suppose that he counts from that corner to the row it is in and thence to the column it is in, thus isolating the point. If the position of the point is uniformly distributed over the cells of the array, then on the average he will have to count $\frac{1}{4} \sqrt{N}$ in each direction. Thus the number, n , of decisions he has to make is given by

$$n = \begin{cases} H, & H \leq 2 \\ 2 + \frac{1}{2} \sqrt{N}, & H > 2 \end{cases}$$

Since $N = 2^H$, we have

$$n = \begin{cases} H, & H \leq 2 \\ 2 + 2^{\frac{H}{2}-1}, & H > 2 \end{cases}$$

Now, if we suppose the decision process is serial, which certainly seems a most reasonable assumption in this case, we may make an estimate of the rate of information transmission C as a function of H . Let us assume the mean base time is equal to the mean of an elementary decision, then up to a scale factor $C = H/(n+1)$. The computation is shown in Fig. 7 where the scale has been normalized so that $C = 1$ when $H = 1$. This curve seems to have approximately the characteristics which have been observed empirically. We have also plotted the parallel case in Fig. 7 and it is clearly inappropriate.

One should not, however, discard the parallel model simply because it does not seem applicable here, for it is known that, for a fixed value of H , C will vary depending on the mode of presentation of the stimuli. Two possibilities seem to exist if the concept of elementary decisions is valid: 1) The latency of the elementary decisions changes with the mode of presentation, or 2) The mental structuring, i.e. the directed graph, of the problem is different for different presentations. It is not known which of these possibilities is true, or whether both occur; a decision will have to rest on experimental evidence. We can, at least, adduce arguments to indicate that the second hypothesis cannot be discarded immediately. Suppose, for example, we were to compare the case of a fixed H in the array presentation with a stimulus having the same H distributed over H conceptually different dimensions. If H were 4, then an example of the latter would be the case of stimulus cards which the background is red or green, on which there are one or two distinguished areas (spots), and the areas are either triangles or circles and

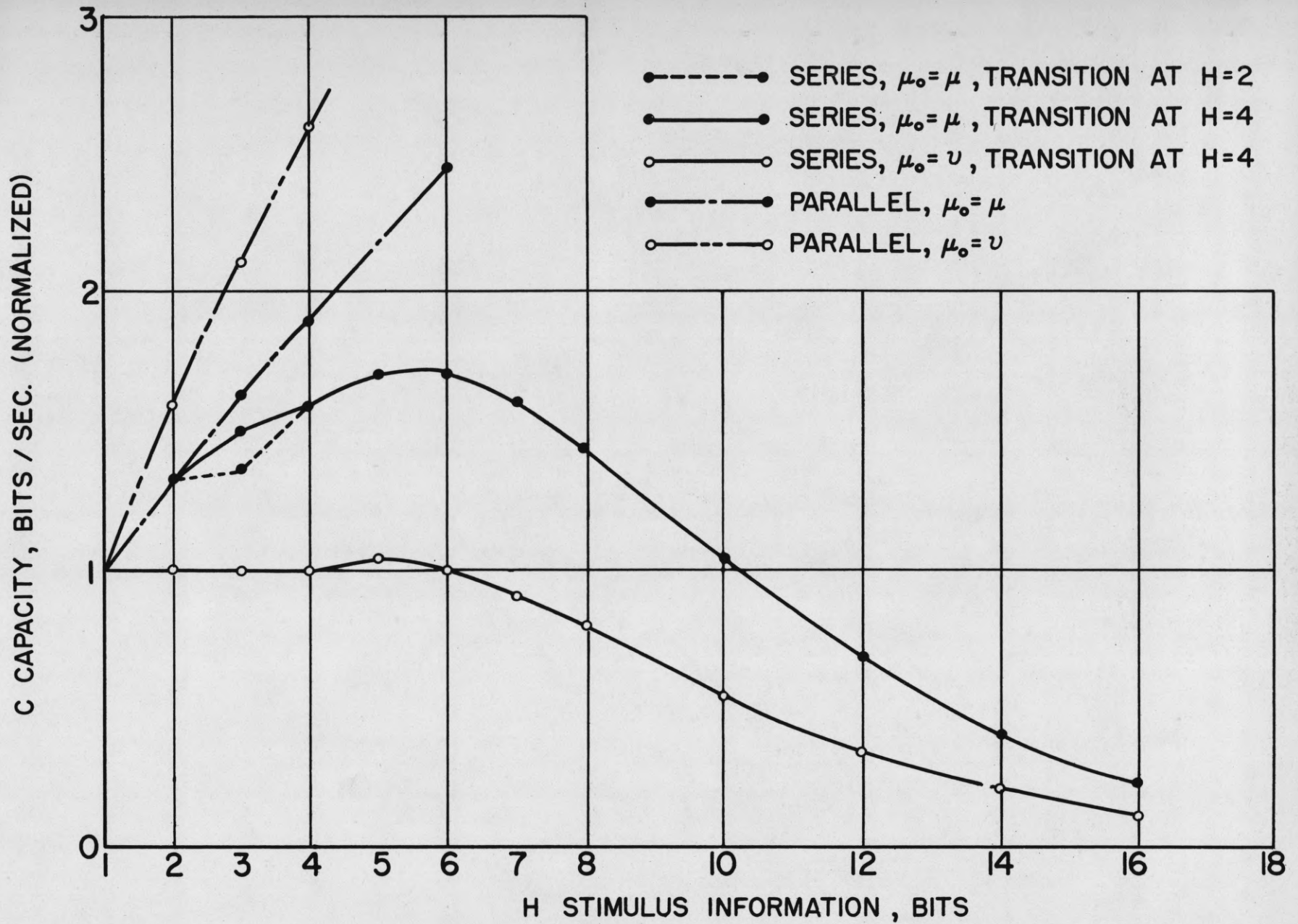


FIG. 7 THEORETICAL INFORMATION CAPACITY

they are colored either black or white. We suppose each possibility in each of the pairs has a probability of one-half. It would seem to be possible for a subject to deal with these four dimensions simultaneously, or nearly so, thus yielding as appropriate a model nearer the parallel case than the series case. However, it is certainly clear that one cannot increase the number of dimensions indefinitely and still retain a parallel model. We remarked at the end of the last section that parallel processing might occur within a moment; possibly the above conjectured parallel processing may be an example of it.

12. Experimental Proposals.

In the foregoing sections we have discussed some of our theoretical notions using a dot and matrix paradigm as the experimental model. In section 11 we adduced reasons for the belief that successive bifurcating of possibilities would break down and be replaced by a counting process when the number of matrix elements becomes large. For these reasons this experimental paradigm is not ideal. In this section we will propose designs for two other experiments which seem somewhat better suited for our purposes although each of them also suffers from limitations which make each less than ideal.

The key assumption in our analysis is that elementary decision processes can be found of such a sort that complex decisions can be built up from them in a way which leaves their characteristic λ value invariant. One should like to present experimental subjects with stimuli which vary in several dimensions but for which decisions on each of the dimensions have identical time characteristics. If one uses conceptually

different dimensions, we may be introducing several different λ values. If we use several objects with the same dimension relevant for each and with identical characteristics in every other respect, we have the difficulty that the reception of the stimulus may not be unitary, but broken down into several parts separated by receptor orienting acts such as eye movements. The first of the two proposals which follow suffers from this difficulty; the second from the former difficulty.

1st Experiment: Digit difference perception

Stimuli: White 3" x 5" cards with a triple-spaced typed, horizontal row of vertically aligned pairs of digits, 0 and 1, on each. The number of pairs per card to vary from one to sixteen. On each card either one pair or no pairs will be unlike digits, i.e., (0,1) or (1,0), the remainder like pairs, i.e., (1,1) or (0,0). The place of the unlike pair in the series of pairs to vary from the initial to the final position. Cards with the unlike pair in each of the positions from one to n will be included in the set with equal frequency and cards with no unlike pair will be included with the same frequency. The assignment of (1,1) or (0,0) to the remaining places will be made on an equiprobable random basis, and the choice of (0,1) or (1,0) for the unlike pair will be made on the same basis.

Responses: Experimenter will announce prior to each stimulus presentation how many pairs the card to be shown bears. Subject will respond yes or no, depending on whether the card does or does not bear an unlike pair, by pressing the appro-

priate one of two keys. The subject will be told that an unlike pair in each of the possible positions, including in no position, are equally likely events, and will be instructed to read the line of pairs from left to right. The data of primary interest will be the latencies of the no response to the cards which bear no unlike pair and the latencies of the yes response to the cards which bear an unlike pair in the n^{th} position.

- Apparatus:
1. Stimulus cards as described above,
 2. Light projector with fast shutter,
 3. Three telegraph keys; a) for the subjects to rest their fingers on prior to response so that the response will always start from the same situation.
 - b) for yes responses
 - c) for no responses
 4. A buzzer of 1/2 sec. duration as a warning signal to be sounded ending 1 sec. before shutter opens to illuminate stimulus.
 5. Recording chronoscope accurate to at least + 10 millisec.
 6. Timer for ready signal and shutter operation with silent starting key for the experimenter.

2nd Experiment: Multi-attribute perception

Stimuli: Ten decks of 32 cards each to be prepared using two values on each of five attributes according to the following scheme:

Attribute	Values
1. Number of spots	2;3
2. Color of spots	Red; black
3. Shape of spots	Round; square
4. Arrangement of spots	Horizontal line; vertical line
5. Background color	White; light blue

Responses: Experimenter will announce what pattern of attributes is to be responded to positively prior to each stimulus presentation. Subject to make a yes or no response by pressing the appropriate one of two keys as exemplified below:

Experimenter says	Stimulus presented	S to respond
1. Round red	Two black squares in horizontal line on white card	No
2. Vertical line of squares on blue card	Three red squares in vertical line on blue card	Yes

The instruction-stimulus pairs which call for a negative response should be half of the total number of stimuli presented in each attribute-pattern category so that the uncertainty of response prior to stimulus presentation will be equalized at the maximum. The data of primary interest will be the latencies of response to the set-stimulus pairs calling for a yes response.

Apparatus: Same as for the first experiment except for the stimulus cards.

Bibliography

1. Christie, L. S. The measurement of discriminative behavior Psychol. Rev., 59, 1952, 443-452.
2. Christie, L. S., Luce, R. D., and Macy, J., Jr. Communication and learning in task-oriented groups Technical Report No. 231, Research Laboratory of Electronics, Mass. Inst. of Technology, 1952, 251 pp.
3. Churchill, R. V. Modern Operational Mathematics in Engineering McGraw Hill, 1944.
4. Fano, R. M. The transmission of information Technical Report No. 65., Research Laboratory of Electronics, Mass. Inst. of Technology, 1949.
5. Fano, R. M. The transmission of information, Part II Technical Report No. 149, Research Laboratory of Electronics, Mass. Inst. of Technology, 1950.
6. Hay, H. The effects of certain communication changes on task-oriented groups, Ph.D. Thesis, Mass. Inst. of Technology, 1953.
7. Klemmer, E. T. and Frick, F. C. Assimilation of information from dot and matrix patterns J. Exp. Psych., 45, 1953, 15-19.
8. Krulee, G. K. and Sinclair, E. J. Some behavioral implications of information theory NRL Report 4119, Tufts College Project, Systems Coordination Division, Naval Research Laboratory, Washington, D. C., 1953.
9. Miller, G. A. Language and Communication McGraw Hill, New York, 1951.
10. Miller, G. A. and Frick, F. C. Statistical behavioristics and sequence of responses Psychol. Rev., 56, 1949, 311-324.
11. Pollack, I. The assimilation of sequentially encoded information Amer. Psychologist, 6, 1951, 266-267.
12. Shannon, C. and Weaver, W. The Mathematical Theory of Communication, Univ. of Ill. Press, Urbana, 1949.
13. Stroud, J. B. The moment function hypothesis M.S. Thesis, Stanford Univ., 1949.
14. Stroud, J. B. The psychological moment in perception in von Forrester, H. Ed: Sixth Conference on Cybernetics, Josiah Macy, Jr. Foundation, 1949, 27-63.
15. Woodworth, R. S. Experimental Psychology Holt, New York 1938.