DECISION STRUCTURE AND TIME RELATIONS IN SIMPLE CHOICE BEHAVIOR*

LEE S. CHRISTIE[†] AND R. DUNCAN LUCE[‡]
GROUP NETWORKS LABORATORY,
RESEARCH LABORATORY OF ELECTRONICS,
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

The structure of simple decisions is considered in terms of a model which composes such decisions from hypothetical elementary decisions. It is argued that reaction-time data can be treated by the use of the Laplace transform so as to overcome difficulties which negated earlier attempts to analyze choice reactions. The general model leads to complex problems which are formulated but not solved. Two special cases of the model are worked out, and the statistical problem of evaluating the fit of the model is discussed. It is shown that treating decision processing as time-discrete leaves the essential features of the analysis unchanged. Two experimental proposals, to provide data which should be considered in further work on the model, are made.

I. *Introduction*. In this paper we propose a model for the way human beings organize the decisions required by simple choice situations into a collection of component decisions. It is our thesis that such an organization of decisions must be reflected in the distribution of reaction times and, therefore, that 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

* This work was carried out when the authors were consultants to the Control Systems Laboratory, University of Illinois. It is a revision of report R-53 of that laboratory issued April 1954, supported under contract DA-36-039-SC-56695.

[†] Present address: Operations Research Office, The Johns Hopkins University, Chevy Chase, Maryland.

[‡] Present address: Center for Advanced Study in the Behavioral Sciences, Stanford, California.

This article appeared in *Bull. math. Biophysics*, 1956, **18**, 89–112. Reprinted with permission.

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.

II. *Reaction Times*. Suppose that a subject receives a stimulus of a fixed type at time 0 and responds at time t with a fixed type of response. The time interval, t, between the stimulus and the response is called the *simple reaction time*. If the subject is presented with one of a set of stimuli and a choice of response contingent on the stimulus is required the corresponding time interval is known as the *disjunctive reaction time*. In either case, it is clear that to obtain stable and readily analyzable time distributions it is necessary that the stimulus be simple enough so that the mean reaction time is no more than a second or two. Otherwise unwanted stimuli may intervene between the test stimulus and the response, and the interaction among 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, 1938, chap. 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 decision (decision latency*) from the other time lags involved in the total process. One attempt to make this separation involved measuring the subject's response to a stimulus when no decision was to be made and subtracting this time from the time required to respond to the same stimulus with the same motor action when a decision was involved. This technique has been considered unsatisfactory for the following reason. If the subject has no decision to make he is able to bring his motor readiness for the specified response to a much higher

* We use *reaction time* when referring to the time of a process timed from stimulus presentation to motor response; *latency* when referring to times of distinguished parts of such a process.

pitch than he can when he is required to make a disjunctive reaction; thus, the *base time*—the time to react in a choice situation excluding the time for the decision itself—cannot be equated to any simple 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 make a decision.

Second, suppose that in one way or another the pure decision latency distribution 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 non-choice latencies (base times) from decision latencies might be an end in itself. If, however, the resulting decision latency 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 non-choice latencies (base times) can also be used, or adapted, to decompose the decision latencies into more primitive terms. Such a decomposition of the observed reaction-time 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.

III. *The Laplace Transform.* Let F be a real-valued function of a real variable t such that $F(t) = 0$ for $t < 0$. The real-valued function $L(F)$ of the real variable s defined by the equation

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

is called the *Laplace transform* of F. There is essentially no loss of information about F in making this transformation [see equation

(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 for they are well known (cf. Churchill, 1944).

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

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

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

iv. If a and b are constants,

$$L(aF + bG) = aL(F) + bL(G). \qquad (5)$$

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

$$L(F) = \frac{1}{\frac{s}{\lambda} + 1}. \qquad (6)$$

IV. *The Model.* Our proposal is based on assumptions which are intuitively acceptable, but which at the moment 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.

Assumption I. It is possible, for a given experimental situation, to divide the observed reaction time t into two latency 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. 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. Specifically, it is not directly dependent on the mode of stimulus presentation or 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 latencies are independent for a fixed experimental situation, it follows from condition 1 that

$$f(t) = \int_0^t f_b(\tau) f_c(t-\tau)\, d\tau \,. \qquad (7)$$

Our second major assumption concerns only the choice latencies 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 appropriate sense, elementary decisions built into him. If such a structure exists in human decision making, it is analogous to the structure of a decision process in a computing machine, which may be thought of as composed from a set of decisions which are elementary relative to that machine, i.e., the elementary decision capabilities 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 at least of the machine, and we shall suppose it is true of human beings. 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 in a human being.

We shall describe the organization of decisions by a *directed graph*. (The terms *oriented graph* and *network* have also been employed in the mathematical literature 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 Figure 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 Figure 2a, and in the sense that there may be lines with opposite directions as in Figure 2b. 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.

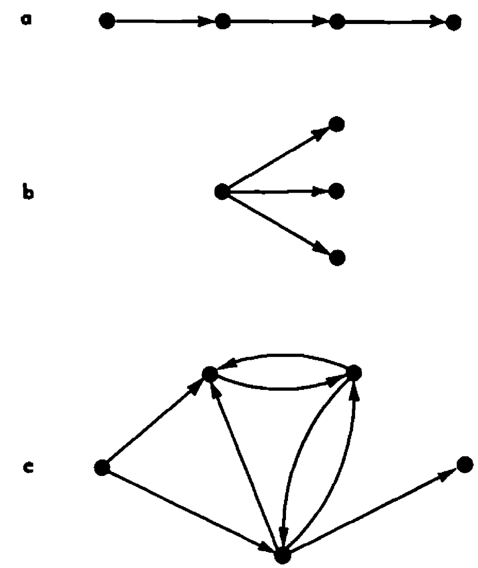

FIGURE 1

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 latency 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 activated by the experimental stimulus at time 0. There will also be at least one node, and again possibly more, which initiates no directed line,

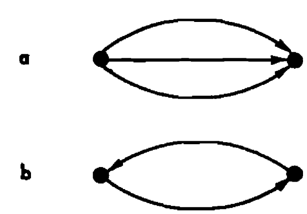

FIGURE 2

and it is only when the decisions at 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.

Assumption II. It is possible to find for each stimulus situation, σ, a set of stimulus situations, S, which all have the same base-time distribution, f_b, and an elementary decision latency, f_e, such that:

1. σ is an element of S.
2. For each choice situation ρ in S 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 S there is one whose directed graph satisfying conditions 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 distribution 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.

V. *Comments*. The above assumptions comprise the formal structure of our model; 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 to be forced to an analysis in terms of individual neurone firings. It is probable that 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 S 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 as "similar" the set of these situations generated as n ranges over the smaller integers. We should probably reject the model if they could not be put in the same set S, 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.

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

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

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 that 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 (Christie, 1952a, b). 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 (Christie, 1952b; Luce, 1953). 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.

VI. *The Problem.* Let S be a set of choice situations which are presumed to satisfy the assumptions of the model, i.e., S is a set of the type described in Assumption II. Let f_σ denote the reaction-time distribution associated with a typical member of S. The problem is then to find distributions f_b and f_e and a set of directed graphs N_σ, where σ ranges over S, such that each of the triples (f_b, f_e, N_σ) when composed according to the assumptions of Section IV yields the distribution f_σ. There may, of course, be no, one, or many solutions to the problem, but one hopes that by an appropriate choice of S 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 S, 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 preassigned transfer functions. If we identify the given reaction-time distribution with the transfer functions, 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 S 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

$$\begin{aligned} f_\sigma(t) &= \int_0^t f_b(r) f_c(t-r)\,dr , \\ f_1(t) &= \int_0^t f_b(r) f_e(t-r)\,dr . \end{aligned} \tag{8}$$

Taking the Laplace transform in each case and applying equation (2),

$$\begin{aligned} L(f_\sigma) &= L(f_b)L(f_c) , \\ L(f_1) &= L(f_b)L(f_e) . \end{aligned} \tag{9}$$

If we divide the first equation by the second in equation (9), we obtain

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

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

$$F(t) = \int_0^t f(r)\,dr .$$

(Throughout we shall use small Latin letters to denote distributions and the corresponding capitals to denote their 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 equation (10), into statements about the cumulative distributions. From equation (3) we have

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

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

$$\frac{L(F_\sigma)}{L(F_1)} = \frac{L(f_c)}{L(f_e)} . \tag{11}$$

Having eliminated f_b from our discussion, the problem of determining it remains. Since our division in equation (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

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

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

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

and so equation (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 equations (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, equation (4) implies

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

or integrating from 0 to t,

$$F_b = \frac{1}{\lambda} f_1 + F_1 . \tag{13}$$

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

VII. *Serial Decision Process.* An alternative program to solving the general problem discussed in Section VI 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. The first, the topic of this section, is the general serial case shown in Figure 3a, and the second, which will be discussed in Section VIII, is the parallel case shown in Figure 3b.

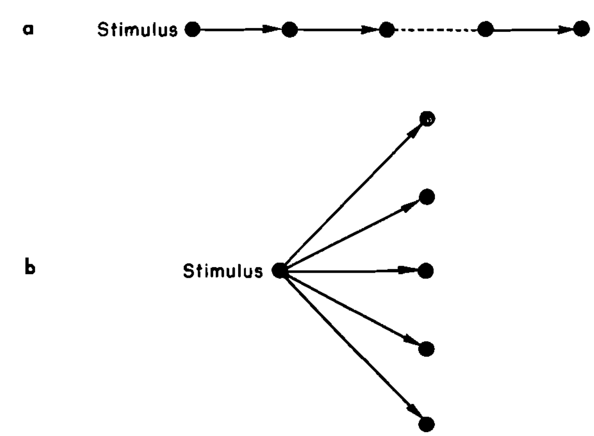

FIGURE 3

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

$$f_n(t) = \int_0^t \cdots \int_0^{t_3}\int_0^{t_2} f_b(t_1)f_e(t_2 - t_1)\ldots f_e(t - t_n)\,dt_1\,dt_2\ldots dt_n\,. \quad (14)$$

Applying the Laplace transform to equation (14) and using equation (2) we have

$$L(f_n) = L(f_b)L(f_e)^n\,, \quad (15)$$

or dividing by the case $n = 1$,

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

Equation (16) is the explicit form of equation (11) for the serial case. Clearly, if we have given numerical data we may determine (possibly numerically) 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, equation (16) becomes

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

In Figure 4 we have presented plots of $\dfrac{1}{\left(\frac{s}{\lambda} + 1\right)^n}$ vs. $\dfrac{s}{\lambda}$ 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 equations (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

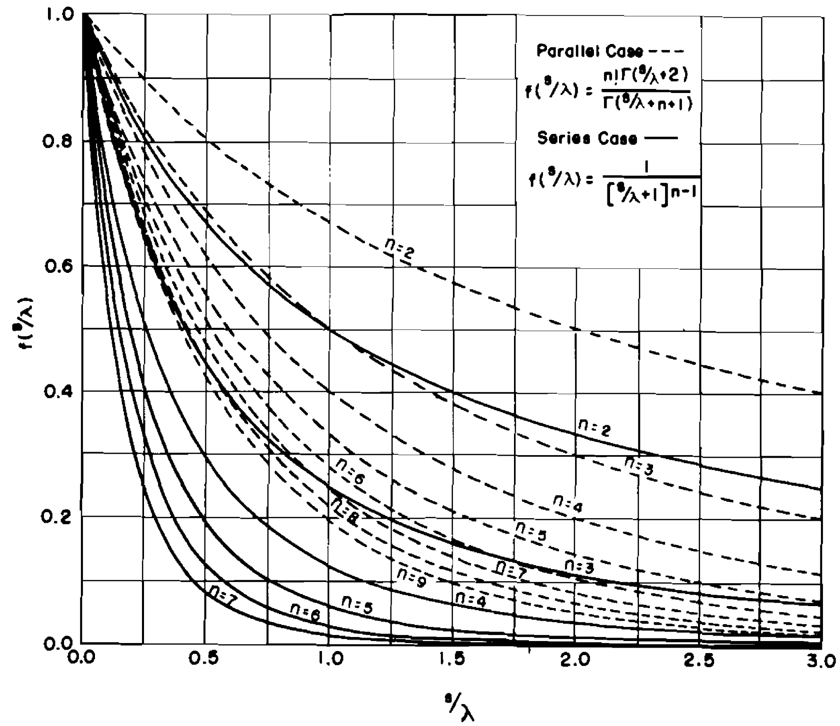

FIGURE 4

ready method may suffice until the statistical problems are formulated and solved. From the data we compute $\dfrac{L(F_n)}{L(F_1)}$ 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 $\dfrac{s}{\lambda}$, say $\dfrac{s}{\lambda} = \dfrac{1}{2}$, find in Figure 4 the corresponding value of $\dfrac{1}{\left(\frac{s}{\lambda} + 1\right)^{n-1}}$.

We know from equation (17) that this must be equal to $\dfrac{L(F_n)}{L(F_1)}$ 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 equation (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 equation (19)] and $(n - 1)/\lambda$ is a minimum; this yields the best possible fit at the point $s/\lambda = 1/2$ 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.

VIII. *Parallel Decision Process*. If we suppose that the n elementary decision processes are carried out in parallel (see Figure 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 when all the elementary distributions are the same, namely F_e, reduces to

$$n f_e(t)[F_e(t)]^{n-1} .$$

If we denote the observed reaction-time distribution for the parallel case by g_n, then it follows from equation (7) that

$$g_n(t) = \int_0^t f_b(\tau) n f_e(t-\tau)[F_e(t-\tau)]^{n-1} d\tau . \tag{20}$$

Applying the Laplace transform and equation (2),

$$L(g_n) = L(f_b) L(n f_e F_e^{\,n-1}) . \tag{21}$$

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

To proceed further, we assume f_e is exponential, then

$$L(n f_e F_e^{\,n-1}) = n\lambda \int_0^\infty e^{-st} e^{-\lambda t}[1 - e^{-\lambda t}]^{n-1} dt ,$$

$$= n\lambda \int_0^\infty e^{-(s+\lambda)t} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k e^{-k\lambda t} dt ,$$

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

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

$$n \int_0^1 \Phi(x)\, dx = n \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \int_0^1 x^{\left(\frac{s}{\lambda}+k\right)} dx ,$$

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

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

and that

$$n \int_0^1 \Phi(x)\, dx = n \int_0^1 x^{s/\lambda} (1-x)^{n-1} dx ,$$

$$= nB\left(\frac{s}{\lambda} + 1, n\right) ,$$

$$= \frac{n\Gamma\left(\frac{s}{\lambda}+1\right)\Gamma(n)}{\Gamma\left(\frac{s}{\lambda}+n+1\right)} ,$$

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

From these results we easily obtain

$$\frac{L(g_n)}{L(g_1)} = \frac{nB\left(\frac{s}{\lambda}+1, n\right)}{B\left(\frac{s}{\lambda}+1, 1\right)} ,$$

$$= \frac{n!\Gamma\left(\frac{s}{\lambda}+2\right)}{\Gamma\left(\frac{s}{\lambda}+n+1\right)} . \tag{22}$$

In Figure 4 we have also 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

$$\mu_1(n) - \mu_1(1) = \frac{1}{\lambda}\sum_{i=2}^{n}\frac{1}{i}. \tag{23}$$

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

IX. *Model Selection.* Without a solution to the general problem described in Section VI, 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 is 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 VII) to determine the constants, but it is almost certain that such an *ad hoc* procedure is not optimal.

The difficulty of making a selection among different sets of assumptions is evidently quite serious for it can be seen from Figure 4 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 are 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 judgments as to goodness-of-fit before trying to formulate and to solve the statistical problems.

X. *The Perceptual Moment.* In Section II we remarked 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. 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 (Stroud, 1949a, b). 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, 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 we may assume that 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, for it may happen that a person is only able to assimilate information during 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 ith moment following the presentation, i.e., at time $i\delta$, then the probability of a decision in the ith moment is $\lambda\delta$. If we call the probability of a response by the ith moment P_i, then

$$\begin{aligned} P_i &= P_{i-1} + [1 - P_{i-1}]\lambda\delta , \\ &= (1 - \lambda\delta)P_{i-1} + \lambda\delta . \end{aligned} \tag{24}$$

With the initial condition $P_0 = 0$, the difference equation (24) is solved by

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

The probability of a decision in the ith moment is obviously

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

hence, we have

$$\lambda\delta(1 - \lambda\delta)^{i-1} , \tag{25}$$

as our distribution f_e.

If we replace this discrete distribution, equation (25), by a continuous one Φ_ϵ 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 $\epsilon \longrightarrow o$ 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

$$f_n(t) = \lim_{\epsilon \to o} \int_0^t \cdots \int_0^t \int_0^t f_b(t_1)\, h(t_2 - t_1)\, \Phi_\epsilon(t_3 - t_2) \ldots \Phi_\epsilon(t - t_{n+1})\, dt_1 \ldots dt_{n+1} . \tag{26}$$

Applying the Laplace transform and using equation (2),

$$L(f_n) = \lim_{\epsilon \to o} L(f_b)\, L(h)\, L(\Phi_\epsilon)^n = L(f_b)\, L(h) \left[\lim_{\epsilon \to o} L(\Phi_\epsilon)\right]^n . \tag{27}$$

Observe,

$$L(\Phi_\epsilon) = \int e^{-st}\, \Phi_\epsilon(t)\, dt ,$$

$$= \sum_{i=1}^{\infty} \int_{i\delta - \frac{\epsilon}{2}}^{i\delta + \frac{\epsilon}{2}} e^{-st}\, \frac{\lambda\delta(1-\lambda\delta)^{i-1}}{\epsilon}\, dt ,$$

$$= \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon} \sum_{i=1}^{\infty} \lambda\delta(1-\lambda\delta)^{i-1} e^{-is\delta} ,$$

$$= \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon}\, \lambda\delta(1-\lambda\delta)^{-1} \sum_{i=1}^{\infty} \{(1-\lambda\delta)\, e^{-s\delta}\}^i .$$

But,

$$\lim_{\epsilon \to o} \frac{e^{\frac{s}{2}\epsilon} - e^{-\frac{s}{2}\epsilon}}{s\epsilon} = 1 ,$$

so,

$$\lim_{\epsilon \to o} L(\Phi_\epsilon) = \frac{\lambda\delta e^{-s\delta}}{1 - (1-\lambda\delta)\, e^{-s\delta}} .$$

Substituting in equation (27) and dividing by the case $n = 1$, we have

$$\frac{L(f_n)}{L(f_1)} = \left[\frac{\lambda\delta e^{-s\delta}}{1 - (1-\lambda\delta)\, e^{-s\delta}}\right]^{n-1} , \tag{28}$$

which is the crucial equation for the discrete serial case. The mean of the discrete distribution f_e is given by

$$\sum_{i=1}^{\infty} i\delta\lambda\delta(1-\lambda\delta)^{i-1} = \frac{1}{\lambda} . \tag{29}$$

Thus, the relation between observed means is

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

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 equations (17) and (19) of Section VII. We have no theoretical value of δ, so it will be necessary to perform independent measurements of it. It is clear that if the perceptual 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 $\epsilon \longrightarrow 0$, is a step function such as that shown in Figure 5. The convolution 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 (cf. Woodworth, 1938) it has been remarked not only that the data are lumpy but that there is an oscillation superimposed on the distribu-

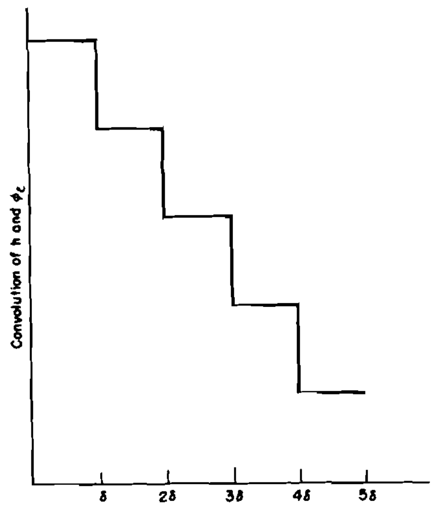

FIGURE 5

tion 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. The latter remark is a possible hint for developing an explanation of the effect of changing the number of "psychological dimensions" in an information display.

XI. *Experimental Proposals*. 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 run into the difficulty of possibly 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 following proposals suffers from the latter difficulty; the second from the former.

1st Experiment: Digit Difference Perception

Stimuli: White 3" × 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 appropriate 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 lines 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 nth 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 ½ 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; green

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 green card	Three red squares in vertical line on green 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.

LITERATURE

Christie, L. S. 1952a. "The Measurement of Discriminative Behavior." *Psychol. Rev.*, **59**, 443–52.

Christie, L. S., R. D. Luce, and J. Macy, Jr. 1952b. "Communication and Learning in Task-Oriented Groups." (Technical Report No. 231.) Cambridge: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Churchill, R. V. 1944. *Modern Operational Mathematics in Engineering.* New York: McGraw Hill.

Luce, R. D., J. Macy, Jr., L. S. Christie, and H. Hay. 1953. "Information Flow in Task-Oriented Groups." (Technical Report No. 264.) Cambridge: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Stroud, J. B. 1949a. "The Moment Function Hypothesis." M. S. Thesis. Stanford University.

Stroud, J. B. 1949b. "The Psychological Moment in Perception." H. von Forrester, Ed. *Sixth Conference on Cybernetics*, Josiah Macy, Jr. Foundation, 27–63.

Woodworth, R. S. 1938. *Experimental Psychology.* New York: Holt.

RECEIVED 5-5-55