APPLICATION OF THE ALL-OR-NONE CONDITIONING MODEL

TO THE LEARNING OF COMPOUND RESPONSES

by

Gordon H. Bower

TECHNICAL REPORT NO. 37

July 7, 1961

PSYCHOLOGY SERIES

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

Applied Mathematics and Statistics Laboratories

STANFORD UNIVERSITY

Stanford, California
APPLICATION OF THE ALL-OR-NONE CONDITIONING MODEL
TO THE LEARNING OF COMPOUND RESPONSES

by

Gordon H. Bower

Over the past year, Estes (1961), Suppes and Ginsberg (1960) and the writer (1960) have been developing and experimentally testing an elementary theory of associative learning. To the present the theory has been applied to simplified experiments on paired-associates learning and it has provided an adequate, detailed account of the data obtained in several such experiments. In the customary paired-associate situation, the subject is required to learn the relevant responses (frequently nonsense syllables) and also to learn to associate each response with its appropriate stimulus member in the list of items. However, in experimental tests of the model, conditions have been arranged to circumvent the response-learning aspect of the typical experiments; instead, the major effort was devoted to studying the formation of associations between stimulus patterns and response alternatives that the subject already knew and had available for use throughout the experiment. Response learning was precluded in some of these experiments by using responses that were highly integrated units in the subject's repertoire (e.g., the integers 1, 2, 3, 4) and informing him in advance of the experiment of what were the relevant responses; in other experiments, the relevant response alternatives were immediately available to the subject by virtue of the construction of the testing

---

1 Some of the support for this work was provided to the writer by a research grant, M-3849, from the National Institutes of Mental Health, and other support from Group Psychology Branch of ONR under Contract NR 171-034.
situation (e.g., two push-buttons, one of which was pressed for each stimulus pattern).

In applying the theory to these simplified experiments, it has been assumed that each stimulus pattern is either fully conditioned to its correct (reinforced) response or else it is not conditioned at all. If the pattern is conditioned, then the subject gives the correct response to that pattern with probability 1; if the pattern is not conditioned, then the subject emits responses with some distribution having probability \( g_i \) of the \( i \)-th response alternative. The "guessing" distribution, \( g_i \), can be controlled in part by experimental factors (e.g., the proportion of items in the list having alternative 1 as their correct response). However, in most cases, the assignment of responses to stimuli in the list has been so balanced that it is plausible to assume that the "guessing" distribution is uniform over the \( N \) response alternatives. Under these conditions, the probability of a correct guess is \( 1/N \) for every item in the list.

The conditioning axiom of the theory states that each time we reinforce the correct response to a stimulus pattern, there is some fixed probability \( c \) that the correct association is learned (if not already learned); with probability \( 1-c \) the reinforcement is ineffective and the item remains unconditioned. It is assumed that the item begins in the unconditioned state and that the effect of successive trials is to provide repeated opportunities for the all-or-none association to be acquired.
A model of this theory is an absorbing two-state Markov chain. The two states correspond to the stimulus pattern being conditioned or not conditioned to its correct response, and $c$ represents the probability of a transition from the unconditioned to the conditioned state as the result of a single reinforcement. This model is mathematically tractable, so a large body of theorems have been derived referring to quantitative predictions about the sequences of responses to a given stimulus during the course of learning (Bower, 1960).

The success of this model in experimental applications indicates that the process of associating available responses to distinct stimulus patterns is effectively an all-or-nothing process. The next major objective is to formulate extensions and modifications of the theory that will permit an accurate description of how the relevant responses themselves are learned and become available for the associative hook-up with stimulus patterns. Available evidence (e.g., McGeogh and Irion, 1952; Underwood and Schultz, 1960) indicates that response learning is a complex process that is influenced greatly by prior learning and which, in turn, produces a number of interactive phenomena when combined with the associative process involved in ordinary paired-associate learning.

In the face of this complexity, the appropriate strategy would appear to be that of initially extending the theory to experimental situations which approach in some mild degree the complete response-learning situations. The purpose of this report is to describe such a situation and show how the theory applies to it and thereby, hopefully, bring us one step closer to an understanding of response learning as it operates in typical paired-associate learning experiments.
This situation to be considered was studied initially in experiments by E. J. Crothers at Indiana University and some subsequent experiments have been carried out by the writer at Stanford University. Crothers' idea was simply to require his subjects to learn concurrently two component responses to each stimulus pattern. For example, the first response might be **red** or **blue** and the second response **circle** or **square**; to the stimulus **DAX** the correct response might be "blue circle". We will refer to such an ensemble of components as a response compound. It may be noted that the subject's task in such experiments is analogous to that involved in certain concept identification problems studied by Bourne and Restle (1960); in their four-response problems, subjects were required to respond to stimulus patterns according to a two-way classification of relevant binary dimensions.

For general discussion, we may characterize a response compound as consisting of \( r \) components with the \( i \)-th component selected from among a set of \( N_i \) known alternatives which are homogeneous in some sense. We may distinguish procedures in which the subject is or is not, respectively, required to emit the \( r \) components in a fixed serial order. In the former case, the response compound may be conceived to be a miniature rote serial list of length \( r \) and we may expect the typical serial-position phenomena to appear. Evidence from experiments by Arthur Jensen indicates that the all-or-nothing law of conditioning holds for such serial components and that the serial position difficulty is reflected mainly in the conditioning parameter, \( c_i \), for the \( i \)-th component of the serial list. Later we shall consider in detail the case for two-components and the discussion is sufficiently general so that the cases
of ordered and unordered components are included. The differences in procedure presumably will affect the parameter values of the model.

In addition to their intrinsic empirical interest, experiments on response compounding are of interest because of the questions they raise concerning the identification of the "response alternatives" to which the theory applies. These identifications affect markedly the way in which the basic theory is applied to a given experiment and the resulting predictions vary considerably with the nature of the response identifications.

II. The Compound as the Response Unit.

According to one set of identifications, we may consider the entire compound as the response unit to which the theory applies. If compound \( Q \) has at least one component different from compound \( R \), then we consider them to be completely distinct alternatives. The number of distinct alternatives is \( N = N_1 \cdot N_2 \cdots N_r \); in our example with \( r = 2 \) and \( N_1 = N_2 = 2 \), the four alternatives would be red circle, red square, blue circle, and blue square. The subject is considered to give the "correct response" only when the entire compound is given correctly; all other compounds are lumped together indiscriminately and called "errors".

In applying the theory with such identifications, the conditional association is presumed to form between the stimulus pattern and the response compound as a unit. Prior to conditioning the subject is presumed to guess at random among the \( N \) alternative compounds. With these identifications, the model developed previously (Bower, 1960) will apply and the expected properties of the learning sequences are well known.
There are no sure rules for deciding when such identifications are appropriate (i.e., will make the model empirically correct). These identifications would seem to be appropriate when the r components are clustered in "meaningful" ways, or, alternatively, when the sequential dependencies between adjacent components of the correct compounds have been learned already by the subject. Three-letter words are an obvious example of meaningful clustering of the component-units (letters) for familiar speakers or readers of the language; in this case, the natural identification of the response alternatives would be the word-units. But such identifications would probably not be appropriate when the subject is familiar only with the characters of the language but not with words or spelling rules.

There are various border-line cases for which prior evidence and intuition are of little help in identifying the alternatives the subject uses (and which, presumably, would be the appropriate identifications for the theory to be correct). Such cases arise particularly when all components are selected from a common set of alternatives. Suppose, for example, \( r = 2 \) and the sets \( N_1 \) and \( N_2 \) both contain the integers 1, 2. Then the compound \((1,2)\) may be learned as the unit "twelve" rather than as a sequence of two-choice decisions \((a\,1,\,\text{then}\,a\,2)\). It is likely that in such cases the subject's response-set can be controlled to a large extent by the experimental instructions.

The procedure of identifying the compound as the response unit has disadvantages in a number of cases. For example, when the component-clustering (or within-component dependency) is weak or absent, these identifications will probably lead to empirically false predictions;
moreover, important data regarding serial learning within the compounds would be ignored by the data analysis dictated by such identifications. It is clear also that identifying the compounds as the response units will be of no help in accounting for phenomena of response generalization. To illustrate this last point, suppose a group of subjects have learned to call DAX a large red circle. Some of the subjects then learn to call it a large red square while others learn to call it a small blue square. There is little doubt that the former subjects would learn their task more quickly because of the common response components. But the "compound as the unit" identifications would not permit such analyses since, according to those identifications, all compounds are viewed as equally distinct responses.

II. Individual Components as the Response Units.

Alternatively, we may consider each component of a compound as a meaningful unit to which the theory may be applied. Over a series of learning trials, there will be some sequence of correct and incorrect occurrences of the i-th response component and the theory is applied directly to such sequences.

With two component responses (call them A and B) the observation each trial for a given stimulus pattern is a pair \((x_n, y_n)\) where \(x_n\) is 0 or 1 according as the A-component is correct or incorrect, and \(y_n\) is 0 or 1 according as the B-component is correct or incorrect. In the analysis of sequences of response pairs, we can distinguish four cases: \(x_n\) and \(y_n\) are mutually dependent, mutually independent, \(x_n\) is independent of \(y_n\) but not vice versa, \(y_n\) is independent of \(x_n\) but not vice versa. Within the model, the independence of the A-component
learning from the B-component learning means that the probability of conditioning the A-component is the same whether or not the B-component is conditioned. If the A-component is independent of the B-component, then the $x_n$ sequence may be analyzed according to the model developed previously (Bower, 1960). If conditioning of the A (or B) component depends upon the conditioning of the B (or A) component, then special derivations are required and these are illustrated in the section below. In this report all theorems referring to the $(x_n, y_n)$ sequences are derived for the general case where the learning processes are mutually dependent. The other three cases follow from this general treatment by special restrictions on the conditioning parameters. We will analyze in detail only the two component situation; the extension of the theory to the general case of r-components is both obvious and cumbersome.

III. Model for Learning Two Mutually Dependent Response Components.

Each stimulus pattern in the experiment is conceived to be either fully conditioned to its correct $i$-th response component or else it is not conditioned to its $i$-th component. For each stimulus requiring two response components we can identify four states of conditioning of this system: neither response is conditioned (state 0, or $s_0$), only the A-response is conditioned (state 1, or $s_1$), only the B-response is conditioned (state 2, or $s_2$), and both components are conditioned (state 3, or $s_3$). In our experiments, it has been reasonable to assume that each stimulus pattern begins in state $s_0$ for each subject. We will assume that a correction training procedure is used ensuring that both components are reinforced every trial to the appropriate stimulus pattern. It may be noted that the reinforcement can be provided either
serially (reinforce A-component after it's emitted, then the B-component after it's emitted) or at the end of the compound response (reinforce $A_i B_j$ after the subject emits a compound response). These two reinforcement procedures may differ in their effectiveness, but we will not inquire here into the possible reasons for this.

With the correction training procedure, the theory implies that the trial-sequence of conditioning states is a Markov chain where states 0, 1, 2 are transient and state 3 is absorbing; that is, after sufficient training, both components will be learned. To relate statements about conditioning states to the response sequences, we have the following response rules:

\[
P(x_n = 0 | s_0) = g_1 \quad P(y_n = 0 | s_0) = g_2 \\
P(x_n = 0 | s_1) = 1 \quad P(y_n = 0 | s_1) = g_2 \\
P(x_n = 0 | s_2) = g_1 \quad P(y_n = 0 | s_2) = 1 \\
P(x_n = 0 | s_3) = 1 \quad P(y_n = 0 | s_3) = 1
\]

These rules are a straightforward generalization of the previous response rules for single components. If component A isn't conditioned, then the probability that $x_n$ is a correct response is $g_1$; if component B isn't conditioned, then the probability that $y_n$ is a correct response is $g_2$. The error probabilities, of course, are $1 - g_1$ and $1 - g_2$, respectively. In our experiments, the component responses were randomized and balanced across stimulus items so that we could assume that $g_i = 1/N_i$ where $N_i$ is the number of available responses for the $i$-th component.
The conditioning rules of the theory are displayed in the following transition matrix where the $i,j$-th entry gives the conditional probability of a transition from $s_i$ (row state) to $s_j$ (column state) as the result of a single reinforcement.

\[
\begin{array}{cccc}
3 & 2 & 1 & 0 \\
3 & 1 & 0 & 0 \\
2 & c_1 & 1-c_1 & 0 & 0 \\
1 & c_2 & 0 & 1-c_2 & 0 \\
0 & a & b & d & 1-r \\
\end{array}
\]

where $r = a + b + d$.\(^2\) For example, $c_1$ is the probability that the first (A) component will become conditioned when the second component is already conditioned. Similarly, $a$ is the probability that both components are learned in a single trial.

Various cases discussed previously can be represented in terms of restrictions on the transition probabilities. Thus, if $b = d = 0$ we have the model treating the compound as a unit where both components are either jointly conditioned or unconditioned. If $a + d = c_1$, then the first component is learned independently of the second, and if $a + b = c_2$ the second component is learned independently of the first. If both these latter conditions hold, then the two learning processes are mutually independent and the constants in the lower row become $a = c_1c_2$, $b = c_2(1-c_1)$ and $d = c_1(1-c_2)$. In this case, we reduce to a two-parameter model, one learning parameter $c_1$ for each response component.

To begin our analysis, we require a few results from the Markov chain. First, we derive the quantities $n_0$ , $n_1$ and $n_2$ which are

\(^2\)In the interests of conserving space and reducing certain equations to manageable compactness, I have freely defined new quantities as certain functions of the basic parameters of the model. Each definition is noted as it is first used; for readers who begin to get lost in the notation, I have provided a reference table (at the end, pgs. 35 and 36) which summarizes the various quantities used throughout the text.
the mean number of trials the system will be in transient states 0, 1 and 2 (starting out in state 0) before absorption in state \( s_3 \). The standard result here (see Kemeny and Snell, 1960) is

\[
\begin{array}{c|ccc}
2 & 1/c_1 & 0 & 0 \\
1 & 0 & 1/c_2 & 0 \\
0 & b/rc_1 & d/rc_2 & 1/r \\
\end{array}
\]

2) \( N = (I - Q)^{-1} = \begin{array}{c|ccc}
2 & 1/c_1 & 0 & 0 \\
1 & 0 & 1/c_2 & 0 \\
0 & b/rc_1 & d/rc_2 & 1/r \\
\end{array} \)

For our purposes, the entries of interest are those in the bottom row which are, respectively, \( n_2 \), \( n_1 \) and \( n_0 \). From these three quantities, we can predict mean total errors and the number of occurrences over the experiment of various response-pairs. For the following, we let \( E(X) \) be the expectation of the random variable \( X \). The results on mean total errors are

\[
E(T_x) = E \left[ \sum_{n=1}^{\infty} x_n \right] = (1 - g_1) (n_0 + n_2)
\]

3) \( E(T_y) = E \left[ \sum_{n=1}^{\infty} y_n \right] = (1 - g_2) (n_0 + n_1) \)

These results are obtained easily. Errors for the A-component can occur only in states \( s_0 \) and \( s_2 \), when that component is not yet conditioned. The mean trials in these states is \( n_0 + n_2 \) and \( (1 - g_1) \) is the probability of an error on the A-component for every trial that that component is not conditioned. From the basic random variables \( x_n \) and
we may form a new random variable \( z_n = x_n + (1 - x_n)y_n \) which is 1
if an error occurs on either the A or B component and it is 0 if both
components are correct. From these considerations we have the result

\[
E(T_z) = E \left[ \sum_{n=1}^{\infty} z_n \right] = (1 - g_1)n_2 + (1 - g_2)n_1 + (1 - g_1 g_2)n_o
\]

If \( g_1 \) and \( g_2 \) are known, then relations (3) and (4) provide estimates
of \( n_o, n_1 \) and \( n_2 \) from empirical data. The natural estimates of the
parameters are then

5) \[
r = \frac{1}{n_o}, \quad d = \frac{n_o n_1}{n_o} , \quad b = \frac{n_o n_2}{n_o}
\]

An alternative (though not independent) way to estimate the \( n_1 \) is to
consider the average number of occurrences of the response pairs (error, error), (success, error) and (error, success). The results for these
are

\[
E \left[ \sum_{n=1}^{\infty} x_n y_n \right] = (1 - g_1) (1 - g_2)n_o
\]

6) \[
E \left[ \sum (1 - x_n)y_n \right] = g_1 (1 - g_2)n_o + (1 - g_2)n_1
\]

\[
E \left[ \sum x_n (1 - y_n) \right] = g_2 (1 - g_1)n_o + (1 - g_1)n_2
\]

If the response compounds are learned as units, then we expect that
\( n_1 = n_2 = 0 \) since \( b = d = 0 \). Hence, deviations from such a model
would appear in the average occurrences of the response pairs given in
Equation 6.
For more detailed analyses of the response sequences we must know the probability that the system is in some state 1 at the beginning of trial \( n \), given that we start off the process in \( s_0 \). We let \( w_{i,n} \) be these probabilities. The \( w_{i,n} \) would be the entries in the bottom row of the \((n-1)^{th}\) power of the transition matrix, \( P \). The results are

\[
\begin{align*}
    w_{0,n} &= (1 - r)^{n-1} \\
    w_{1,n} &= \frac{d}{r-c_2} \left[ (1 - c_2)^{n-1} - (1 - r)^{n-1} \right] \\
    w_{2,n} &= \frac{b}{r-c_1} \left[ (1 - c_1)^{n-1} - (1 - r)^{n-1} \right]
\end{align*}
\]

The result for \( w_{0,n} \) follows from the fact that there is a constant probability, \( 1 - r \), that we stay in \( s_0 \) every trial. The result for \( w_{2,n} \) is the sum over all \( k \) of the probability of staying in \( s_0 \) \( k-1 \) trials, moving to \( s_2 \) with probability \( b \) and then staying in \( s_2 \) for \( n-1-k \) trials, each with probability \( 1 - c_1 \). The sum is

\[
w_{2,n} = \sum_{k=1}^{n-1} (1 - r)^{k-1} b (1 - c_1)^{n-1-k}
\]

which results in the expression in (7) above. Similar reasoning is involved in obtaining \( w_{1,n} \). It will be noted that the sum over all \( n \) of the probabilities of being in \( s_1 \) is just \( n_1 \). In case \( c_1 = r \), the expression for \( w_{2,n} \) modifies to

\[
w_{2,n} = \frac{b n (n-1)}{2} (1 - r)^{n-2}
\]
Similarly, if $c_2 = r$, then $w_{1,n}$ is as above except $b$ is replaced by $d$.

Given the $w_{i,n}$, the average error probabilities on trial $n$ are

$$q_{1,n} = P(x_n = 1) = (1 - g_1) (w_{o,n} + w_{2,n})$$

8) $$q_{2,n} = P(y_n = 1) = (1 - g_2) (w_{o,n} + w_{1,n})$$

$$q_n = P(z_n = 1) = (1 - g_1 g_2) w_{o,n} + (1 - g_1) w_{2,n} + (1 - g_2) w_{1,n}$$

Summing these expressions for mean error probability over all trials results in the expressions for expected total errors given in Equations 3 and 4.

From the $w_{i,n}$ it is easy to compute various conditional probabilities obtaining between the values of $x_n$ and $y_n$. For example, suppose on trial $n$ we observe an error on the A-component; what is the conditional probability that the B-response will also be wrong? To compute this, we use the standard formula

$$P(y_n = 1| x_n = 1) = \frac{P(x_n = 1, y_n = 1)}{P(x_n = 1)}$$

and substitute

$$P(x_n = 1, y_n = 1) = (1 - g_1) (1 - g_2) w_{o,n}$$

and

$$P(x_n = 1) = (1 - g_1) (w_{o,n} + w_{2,n})$$
Substitution of these quantities yields

\[ P(y_n = 1|x_n = 1) = \frac{1 - g_2}{1 - \frac{b}{r-c_1} + \frac{b}{r-c_1} \left[ \frac{1-c_1}{1-r} \right]^{n-1}} \]

Now if \( r > c_1 \), as it is expected to be, then this expression approaches zero as \( n \) becomes large; that is, after a large amount of training, it is almost certain that the B-component is learned and hence no errors occur on it. We may reverse the order of conditionalization and ask, given an error on the second component on trial \( n \), how likely is it that the first component was also wrong? The result obtained here is

\[ P(x_n = 1|y_n = 1) = \frac{1 - g_1}{1 - \frac{d}{r-c_2} + \frac{d}{r-c_2} \left[ \frac{1-c_2}{1-r} \right]^{n-1}} \]

In similar manner, if we conditionalize upon a success, the two results are

\[ P(y_n = 1|x_n = 0) = \frac{(1-g_2) \left[ g_1 w_{o,n} + w_{l,n} \right]}{1 - (1-g_1) \left[ w_{o,n} + w_{2,n} \right]} \]

\[ P(x_n = 1|y_n = 0) = \frac{(1-g_1) \left[ g_2 w_{o,n} + w_{2,n} \right]}{1 - (1-g_2) \left[ w_{o,n} + w_{l,n} \right]} \]
For large \( n \), the limit of both of these expressions is zero since both numerators go to zero as the denominators approach one.

IV. Analyses for a single component.

In the following, we seek further theorems about the response sequences. Here we concentrate only upon one of the component responses (say, the \( x_n \) sequence) and ignore for the moment the other components. We derive a number of properties of the \( x_n \) sequence; by symmetry, these same results hold for the \( y_n \) sequence when we replace \( g_1 \) by \( g_2 \), \( b \) by \( d \), \( c_1 \) by \( c_2 \).

In concentrating our analysis upon the \( x_n \) sequence, it is of conceptual benefit to lump together states 3 and 1 since these both represent "component A is conditioned". The resulting three-state chain has the transition matrix given below.

\[
P = \begin{pmatrix}
1 & 0 & 0 \\
1/3 & 1-c_1 & 0 \\
1/3 & 1/3 & 1/3 & 1-r
\end{pmatrix}
\]

States 0 and 2 are transient. A nice feature for analytical purposes is that the probability of an error, \( P(x_n = 1) \), is the same regardless of which transient state, 0 or 2, the item is in on trial \( n \). The states are distinguished, however, by the probability of conditioning the A-component.

We begin by deriving some sequential properties of the \( x_n \) sequence; specifically, we seek the expectation of \( r_j \), the number of error runs of length \( j \) in an infinitely long sequence of trials. For these purposes,
it is convenient to first work with the numbers \( u_{j,n} \) which count j-tuples of errors beginning with an error on trial \( n \). Thus,

\[
u_{j,n} = x_n x_{n+1} \cdots x_{n+j-1}
\]

In obtaining the expectation of \( u_{j,n} \), we note that an error can occur on trial \( n \) in either state 0 or state 2. Suppose first that the subject is in \( s_2 \) at the beginning of trial \( n \); then the probability of \( j \) consecutive errors is \((1-c_1)^j (1-c_2)^{j-1}\), where \((1-c_1)^{j-1}\) gives the probability that the subject does not move from \( s_2 \) into the conditioned state \( s_j \) in \( j-1 \) reinforcements. Suppose, on the other hand, that we begin in \( s_0 \) on trial \( n \). In order to have \( j \) consecutive errors, we must not have the response conditioned during the \( j-1 \) reinforcements. During these \( j-1 \) reinforcements, the system may stay in \( s_0 \), with probability \( w_{0,j} \), or move at some time to \( s_2 \), with probability \( w_{2,j} \). Hence, the final expression for the expectation of \( u_{j,n} \) is

\[
10) \quad E(u_{j,n}) = (1 - c_1)^j \left\{ w_{2,n} (1 - c_2)^{j-1} + w_{0,n} \left[ w_{0,j} + w_{2,j} \right] \right\}
\]

where

\[
w_{0,j} = (1 - r)^{j-1},
\]

and

\[
w_{2,j} = \frac{b}{r-c_1} \left[ (1 - c_2)^{j-1} - (1 - r)^{j-1} \right]
\]
Now we define $u_j$ to be the overall trial-sum of $u_{j,n}$. Summing over $w_{2,n}$ and $w_{o,n}$, the expectation of $u_j$ is found to be

$$E(u_j) = E\left[ \sum_{n=1}^{\infty} u_{j,n} \right] = (1 - g_1)^j \left[ n_2 (1 - c_1)^{j-1} + n_o [w_{o,j} + w_{2,j}] \right]$$

where $n_2$ and $n_o$ are the mean number of trials the system is in states $s_2$ and $s_o$, and were derived before in Equation 2.

The $u_j$ bear a simple relationship to the $r_j$. The relation is

$$r_j = u_j - 2u_{j+1} + u_{j+2}$$

$$R = \sum_{j=1}^{\infty} r_j = u_1 - u_2$$

We have derived $E(u_j)$; the relations above show that $E(r_j)$ is a simple linear combination of $E(u_j)$ values. Hence, we've also derived $E(r_j)$ from the model.

A closely related random variable is $c_{k,n} = x_n x_{n+k}$ which counts joint errors $k$ trials apart. It has the same features as an auto-correlation of lag $k$. By the same reasoning as above (except we omit the necessity for intervening errors), the expectation of $c_{k,n}$ is found to be

$$E(c_{k,n}) = E[x_n x_{n+k}] = (1-g_1)^2 \left[ w_{2,n} (1-c_1)^k + w_{o,n} [w_{o,k+1} + w_{2,k+1}] \right]$$

We define $c_k$ to be the sum of $c_{k,n}$ over all trials $n$. Its expectation, summing over $w_{2,n}$ and $w_{o,n}$, is
13) \[ E(c_n) = E \left[ \sum_{n=1}^{\infty} x_n x_{n+k} \right] = (1-g_1)^2 \left[ n_2 (1-c_1)^k + n_o \left( w_{o,k+1} + w_{2,k+1} \right) \right] \]

To complete this section on sequential statistics, we compute the mean number of alternations of successes and failures that will occur during the course of learning. We define a random variable \( A_n = x_n (1 - x_{n+1}) + (1 - x_n) x_{n+1} \) which will count an alternation (fail success or success fail) between trials \( n \) and \( n + 1 \). Consider first this measure conditional upon the process being in \( s_2 \) at the beginning of trial \( n \). The probability of a success-then-error is \( g_1 [(1-c_1)(1-g_1)] = g_1 \tau \) and the probability of error then success is \( (1-g_1) [c_1 + (1-c_1)g_1] = (1-g_1)(1 - \tau) \). Similarly consider the process to be in \( s_0 \) at trial \( n \). The probability of success-error is \( g_1 [(1-a-d)(1-g_1)] = g_1 \sigma \), and of error-success is \( (1-g_1)[a + d + g_1(1-a-d)] = (1-g_1)(1-\sigma) \). Combining these results, taking consideration of the likelihood of being in \( s_0 \) or \( s_2 \) on trial \( n \), the expectation of \( A_n \) is

14) \[ E(A_n) = w_{o,n} [g_1 \sigma + (1-g_1)(1-\sigma)] + w_{2,n} [g_1 \tau + (1-g_1)(1-\tau)] \]

We then define \( A \) as the sum of \( A_n \) over all trials \( n \), having expectation

15) \[ E(A) = E \left[ \sum_{n=1}^{\infty} A_n \right] = n_o [g_1 \sigma + (1-g_1)(1-\sigma)] + n_2 [g_1 \tau + (1-g_1)(1-\tau)] \]

where we have defined \( \sigma = (1-a-d)(1-g_1) \)

\( \tau = (1-c_1)(1-g_1) \)

19
We now consider various features of the total errors partitioned about particular points in the learning sequence. The first item to consider is the number of errors that follow an error occurring on trial \( n \). Call this random variable \( e_n \). In an all-or-none learning model, if the conditioning probability for an unlearned response is constant over trials, then the expectation of \( e_n \) is a constant independent of \( n \). This aspect of the model has been confirmed experimentally a number of times. In the present model if the conditioning parameter for a given component depends on the conditioning of the other component, then \( E(e_n) \) depends on \( n \) because the probability that the other component is conditioned varies with \( n \).

For the present case, we begin by supposing that the error on trial \( n \) came about with the process in \( s_2 \). On trial \( n \), the response is conditioned with probability \( c_1 \) (so no more errors would occur) but fails to be conditioned with probability \( 1 - c_1 \) (so \( (1 - g_1)/c_1 \) more errors are expected). Hence, we have

\[
E(e_n | s_2, n) = (1 - c_1) \frac{(1 - g_1)}{c_1} = \frac{\tau}{c_1}
\]

Now suppose the process were in \( s_0 \) on trial \( n \). The possible one-step transitions (from trial \( n \) to \( n + 1 \)) are indicated below along with the expected number of subsequent errors given that transition:

\[
\begin{align*}
& s_0 \xrightarrow{b} s_2 \\
& s_0 \xrightarrow{0} (s_1 \text{ or } s_3) \\
& (s_0) \xrightarrow{1 - \tau} (s_0)
\end{align*}
\]

\[\frac{e_n}{0} = \frac{(1 - g_1)}{c_1} = (1 - g_1)(n_0 + n_2) = E(T_x)\]
Weighting each of these right-hand values by the probability of the transition, the result is

\[ 17) \quad E(e_n | s_{o,n}) = b \frac{(1-g_1)}{c_1} + (1-r) E(T_x) \]

We know that if \( a + d = c_1 \), then learning of the A-component is independent of the B-component; therefore, with \( a + d = c_1 \) we expect and find that Equations 16 and 17 are equal and we have for the A-component the constancy result on \( E(e_n) \) referred to previously.

Without more specific knowledge of the parameter values, we know only that \( E(e_n) \) will be bounded between the values given by Equations 16 and 17. Specifically, the relation is

\[ 18) \quad E(e_n) = \frac{w_{o,n} E(e_n | s_{o,n}) + w_{2,n} E(e_n | s_{2,n})}{w_{o,n} + w_{2,n}} \]

and this will be monotone increasing or decreasing depending on the parameter-values.

V. Distributions of Error-Quantities.

We now turn to deriving the distributions of the trial number of the last failure, total errors, and errors between the \( k \)-th and \((k + 1)\)st success, all referred to the \( x_n \) sequence. Before proceeding with these, it is helpful to first obtain two other quantities: \( b_2 \), which is the probability that there are no more errors following a guess (correct or incorrect) in \( s_2 \); and \( b_0 \), referring to the similar probability of no more errors following a guess in \( s_0 \). A recursion for \( b_2 \) is
19) \[ b_2 = c_1 + (1-c_1)g_1 b_2 \]
or\[ b_2 = \frac{c_1}{1-g_1(1-c_1)} \]

With the probability \( c_1 \) the response was conditioned by the current reinforcement; with probability \( 1 - c_1 \) that reinforcement failed to condition the response, the subject guessed correctly the next trial and then makes no more errors with probability \( b_2 \). In like manner, a recursion for \( b_o \) is found to be

20) \[ b_o = a + d + b g_1 b_2 + (1-r)g_1 b_o \]
or\[ b_o = \frac{a + b + b g_1 b_2}{1-g_1(1-r)} \]

Here with probability \( a + d \) the A-component was conditioned by the current reinforcement; with probability \( b \) the B-component was conditioned and the probability of a correct guess on the A-component on the next trial and no more errors is \( g_1 b_2 \); with probability \( 1-r \) the process stays in \( s_o \) and the probability of no more errors is \( g_1 b_o \).

Using these values of \( b_o \) and \( b_2 \), it is easy to write the distribution of \( n' \), the trial number on which the last error occurs during learning. If no errors occur, then we set \( n' = 0 \). The distribution is

21) \[ \Pr(n' = k) = \begin{cases} 
  g_1 b_o & \text{for } k = 0 \\
  (1-g_1)[b_o w_{o,k} + b_2 w_{2,k}] & \text{for } k = 1, 2, \ldots 
\end{cases} \]
We now consider the distribution of total errors, \( T \). The probability of zero errors is just \( g_1 b_0 \), as given in 21 above. For \( k > 0 \), we consider the probability that the subject makes his \( k \)-th (and last) error on trial \( j \) after having made \( k-1 \) errors someplace during the previous \( j-1 \) trials. Up until the trial of the last error, we are observing a fixed Bernoulli process with probability \( g_1 \) of a success and \( 1-g_1 \) of an error. We write the probability of \( k-1 \) errors in \( j-1 \) trials as

\[
B(k-1, j-1) = \binom{j-1}{k-1} g_1^{j-k} (1-g_1)^{k-1}
\]

The probability of obtaining exactly \( k \) errors (for \( k > 0 \)) is the sum over all \( j \) of the probabilities of making the last error on trial \( j \) and making \( k-1 \) errors during the preceding \( j-1 \) trials. That is,

\[
P(T = k) = \sum_{j=k}^{\infty} B(k-1, j-1) \cdot P(n' = j)
\]

By appropriate substitution from Equations 21 and 22 the result is

\[
P(T = k) = (1-g_1)^k b_0 \sum_{j=k}^{\infty} \binom{j-1}{k-1} g_1^{j-k} (1-r)^{j-1}
\]

\[
+ \frac{b(1-g_1)^k b_2}{r-c_1} \sum_{j=k}^{\infty} \binom{j-1}{k-1} g_1^{j-k} \left[(1-c_1)^{j-1}(1-r)^{j-1}\right]
\]

These sums on the right-hand side are all of the same form and reduce to simpler expressions. For example, considering only the first sum, it reduces to
\[(1-r)^{k-1} \sum_{i=0}^{\infty} \binom{k-1 + i}{k-1} \left[ g_1(1-r) \right]^i \]

which can be shown to equal

\[(1-r)^{k-1} \cdot \left[ \frac{1}{1-g_1(1-r)} \right]^k \]

Similar results obtain for the last two sums. If we employ the following abbreviations

\[\frac{(1-g_1)(1-r)}{1-g_1(1-r)} = \gamma \text{ and } \frac{(1-g_1)(1-c_1)}{1-g_1(1-c_1)} = \alpha \]

then the distribution of \( T \) can be written compactly as

\[P(T = k) = \frac{b_0 \gamma^k}{1-r} + \frac{b_0 b_2}{r-c_1} \left[ \frac{\alpha^k}{1-c_1} - \frac{\gamma^k}{1-r} \right] \text{ for } k \geq 1\]

We now derive the distribution of \( J_0 \), the number of errors before the first success. To have \( k \) errors before the first success, the process can either stay in \( s_0 \) \( k-1 \) reinforcements or can be in \( s_2 \) at the end of \( k-1 \) reinforcements. Given that the system was in \( s_0 \) on trial \( k \) (when the \( k \)-th consecutive error occurred), the probability of a success on the next trial is \( a + d + g_1(1 - a - d) = 1 - \sigma \); given that we were in \( s_2 \) on the \( k \)-th trial, the probability of a success on the next trial is \( c_1 + g_1(1-c_1) = 1 - \tau \). From these considerations, we may write the distribution of \( J_0 \) as
26) \[ P(J_0 = k) = \begin{cases} g_1, & \text{for } k = 0 \\ (1-g_1)^k \left[ w_{o,k}(1-\sigma) + w_{2,k}(1-\tau) \right], & \text{for } k \geq 1 \end{cases} \]

We now derive expressions for \( p_k \), the probability of no errors following the \( k \)-th success, and the distribution of \( J_k \), the number of errors between the \( k \)-th and \((k+1)\)st success. To obtain these quantities we first must derive the probability that the process is in \( s_0 \) on the trial of the \( k \)-th success (call this \( f_{o,k} \)) and the probability that the process is in \( s_2 \) on the trial of the \( k \)-th success (call this \( f_{2,k} \)). First, we derive \( f_{o,k} \), the probability that the \( k \)-th success occurs with the process in \( s_0 \). One way to realize this event is for the process to have \( k-1 \) successes in the first \( n-1 \) trials, be in state \( s_0 \) on trial \( n \) and have the response on trial \( n \) be a correct guess (with probability \( g_1 \)). If we then sum these probabilities over all values of \( n \geq k \) we will obtain \( f_{o,k} \):

27) \[
\begin{align*}
f_{o,k} = \sum_{n=k}^{\infty} w_{o,n} \binom{n-1}{k-1} g_1 g_1^{k-1} (1-g_1)^{n-k} \\
= \sum_{n=k}^{\infty} (1-r)^{n-1} \binom{n-1}{k-1} g_1^k (1-g_1)^{n-k} \\
= g_1^k (1-r)^{k-1} \sum_{i=0}^{\infty} \frac{[k-1]}{k-1} [(1-r)(1-g_1)]^i \\
= \frac{g_1^k (1-r)^{k-1}}{[1-(1-r)(1-g_1)]^k} = \frac{\mu^k}{(1-r)}
\end{align*}
\]
where we set \( \mu = \frac{g_1(1-r)}{1-(1-r)(1-g_1)} \).

By the same reasoning, \( f_{2,k} \) is given by

\[
f_{2,k} = \sum_{n=k}^{\infty} w_{2,n} \binom{n-1}{k-1} g^k (1-g)^{n-k}
\]

\[
= \frac{b}{r-c_1} \left[ \frac{\beta^k}{1-c_1} - \frac{\mu^k}{1-r} \right]
\]

where

\[
\beta = \frac{g_1(1-c_1)}{1-(1-g_1)(1-c_1)}
\]

Given these quantities, the probability of no errors following the k-th success is

\[
P_k = 1 - f_{o,k} - f_{2,k} + f_{o,k} b_o + f_{2,k} b_2
\]

\[
= 1 - (1-b_o)f_{o,k} - (1-b_2)f_{2,k}
\]

The quantity \( 1 - f_{o,k} - f_{2,k} \) is the likelihood that the response was in the conditioned state at the k-th success, so with probability 1 no more errors would occur in that case; \( f_{o,k} b_o \) is the probability that the k-th success occurred in \( s_o \) and with probability \( b_o \) no subsequent errors occurred; \( f_{2,k} b_2 \) is the analogous expression for the k-th success occurring with the process in \( s_2 \). The limit of \( P_k \) is 1 since both \( f_{o,k} \) and \( f_{2,k} \) go to zero as \( k \) becomes large. This is reasonable: after a large number of successes, it's likely that the response is conditioned so no more errors are going to occur.
We now derive the distribution of \( J_k \), the number of errors occurring between the \( k \)-th and \( (k+1) \)-th success. We can have \( J_k = 0 \) in case (a) the \( k \)-th success occurred by prior conditioning with probability \( 1 - f_{0,k} - f_{2,k} \), or (b) the process is in \( s_0 \) at \( k \)-th success and we have a success on the next trial with probability \( a + d + g_1(l-a-d) = (l-\sigma) \), or (c) the process is in \( s_2 \) at the \( k \)-th success and we have a success on the next trial with probability \( c_1 + g_1(l-c_1) = (l-\tau) \). Thus,

\[
P(J_k = 0) = 1 - f_{0,k} - f_{2,k} + f_{2,k}(1-\tau) + f_{0,k}(1-\sigma)
\]

\[
= 1 - \sigma f_{0,k} - \tau f_{2,k}
\]

For \( J_k = n > 0 \): with probability \( f_{2,k} \) we can start in \( s_2 \), stay there for \( n \) trials while making \( n \) errors, then get a success on the next trial with probability \( 1 - \tau \); with probability \( f_{0,k} \) we start in \( s_0 \) and either stay there or move to \( s_2 \) during the \( n \) trials.

From these considerations, we may write the distribution as

\[
P(J_k = n) = \begin{cases} 
1 - \sigma f_{0,k} - \tau f_{2,k} & \text{for } n = 0 \\
(l-g_1)^n \left[ f_{0,k} w_{0,n+1}(1-\sigma) + f_{0,k} w_{2,n+1}(1-\tau) + f_{2,k} (1-c_1)^n (1-\tau) \right] & \text{for } n \geq 1
\end{cases}
\]

where, as before,

\[ w_{0,n+1} = (l-r)^n \]
\[ w_{2,n+1} = \frac{b}{r-c_1} \left[ (1-c_1)^n - (1-r)^n \right] \]

As our final derivation here, we consider the distribution of the number of successes intervening between adjacent errors (provided there is a next error). Call this random variable \( H \). In the simple one-element model with a constant probability of conditioning, the next error is an uncertain recurrent event and \( H \) has a simple geometric distribution which is independent of the trial number \( n \) on which the leading error occurred. Such simple results do not obtain for the present model since the distribution of \( H_n \) depends upon the state (\( s_0 \) or \( s_2 \)) giving rise to the error on trial \( n \), and the probabilities of being in these states changes with \( n \). This means that we must first find the distribution of \( H_n \) conditional upon the process being in \( s_0 \) on trial \( n \), and also find the distribution of \( H_n \) conditional upon being in \( s_2 \) on trial \( n \). Before proceeding to this problem, however, we first consider the distribution of \( H_0 \), the number of successes before the first error. This problem is easy since by assumption the subject is in \( s_0 \) at the outset of training. The distribution of \( H_0 \) is

\[ P(H_0 = j) = \begin{cases} \frac{1-g_1}{1-g_1 b_0} & \text{for } j = 0 \\ \frac{(1-g_1)}{1-g_1 b_0} g_1^j \left[ w_{o,j+1} + w_{2,j+1} \right] & \text{for } j \geq 1 \end{cases} \]

The division by \( 1-g_1 b_0 \) establishes the condition that at least one error will occur. Given that at least one error is going to occur and given that \( j \) reinforcements fail to condition the response, then the probability of \( j \) successes then an error is \( g_1^j (1-g) \).
We now consider the distribution of $H_n$ conditional upon $s_2$ obtaining on trial $n$. Call this distribution $P(H_n|s_{2,n})$. The probability distribution here is

$$P(H_n = j|s_{2,n}) = \frac{(1-c_1)(1-g_1)}{1-b_2} \left[ g_1(1-c_1) \right]^j = \left[ 1-g_1(1-c_1) \right][g_1(1-c_1)]^j$$

for $j = 0, 1, 2, \ldots$

The division by $1-b_2$ ensures at least one more error; the factor $[g_1(1-c_1)]^j$ gives the probability of $j$ consecutive correct guesses without conditioning occurring, and then an error on the next trial with probability $(1-c_1)(1-g_1)$.

The similar density function conditional upon the process being in $s_o$ at the time the error occurred on trial $n$ is called $P(H_n|s_{o,n})$ and is given by

$$P(H_n = j|s_{o,n}) = \begin{cases} 
\frac{(1-a-\delta)(1-g_1)}{1-b_0} & \text{for } j = 0 \\
\frac{(1-g_1)}{1-b_0} g_1^j \left[ w_{o,j+2} + w_{2,j+2} \right] & \text{for } j \geq 1 
\end{cases}$$

The reason for the $(j+2)$ subscripts on $w_o$ and $w_2$ is to express the fact that there are $j+1$ reinforcements that occur between the error on trial $n$ and the next error on trial $n+j+1$. The probability that the process remains in $s_o$ during these $j+1$ reinforcements is $(1-r)^{j+1}$, which is $w_{o,j+2}$ according to the notation established in Equation 7.

From these conditional distributions, we may obtain the unconditional distribution of $H_n$ by the relation
35) \[ P(H_n = j) = \frac{w_{o,n} P(H_n = j | s_{o,n}) + w_{s,n} P(H_n = j | s_{s,n})}{w_{o,n} + w_{s,n}} \]

These derivations above have been carried out with reference to \( H_n \) given an error occurring on trial \( n \). We could also count off errors and ask for the distribution of \( H'_{k} \), the number of successes between the \( k \)-th and \((k + 1)\)st error (provided there is a \((k + 1)\)st error). The conditional distributions - \( P(H'_k | s_2 \) at \( k \)-th error) and \( P(H'_k | s_0 \) at \( k \)-th error) - would be the same as the expressions in Equations 33 and 34, respectively. Combining these as in Equation 35 to obtain the unconditional distribution would require knowing the relative probabilities that the \( k \)-th error came about with the process in \( s_0 \) rather than \( s_2 \). Let \( t_{o,k} \) be the probability that a \( k \)-th error occurs and that the process is in \( s_0 \) at the time. The expression for \( t_{o,k} \) is

36) \[ t_{o,k} = \sum_{n=k}^{\infty} w_{o,n} \binom{n-1}{k-1} (1-g_1)^k g_1^{n-k} \]

\[ = \frac{(1-r)^{k-1}(1-g_1)^k}{[1-g_1(1-r)]^k} = \frac{k}{1-r} \]

Similarly, the probability that a \( k \)-th error occurs and the subject is in \( s_2 \) at the time is

37) \[ t_{s,k} = \sum_{n=k}^{\infty} w_{s,n} \binom{n-1}{k-1} (1-g_1)^k g_1^{n-k} \]

\[ = \frac{b}{r-c_1} \left[ \frac{c_1^k}{1-c_1} - \frac{y^k}{1-r} \right] \]
where \( \alpha \) and \( \gamma \) are as given before in Equation 25. Since we are assuming, in the conditionalization of \( H'_k \), that a \( k \)-th error occurs, the relative probability that it came about in \( s_o \) is given by \( t_{o,k} \) divided by \( t_{o,k} \) plus \( t_{2,k} \). Hence, the unconditional distribution of \( H'_k \), the number of successes between the \( k \)-th and \((k + 1)\)-th error (provided there is one), is

\[
P(H'_k) = \frac{t_{o,k} \frac{P(H'_k|s_o)}{t_{2,k} \frac{P(H'_k|s_o)}}}{t_{o,k} + t_{2,k}}
\]
V. Concluding Comments

What has been accomplished in the preceding pages is a detailed statistical analysis of response sequences for the general case of a subject learning two mutually dependent response components. The fundamental unit of analysis still remains the single response component (the $x_n$ sequence or $y_n$ sequence) rather than the composite score ($z_n$) representing joint performance on the two components. The main reason for this emphasis here was that of mathematical convenience; the $z_n$ sequence presents some special difficulties for extensive analyses, although it is clear that a few results could be easily derived (e.g., the trial of the last error).

A second reason for emphasizing the model's analysis of the single component is that the corresponding results have application to another class of problems having little to do with response compound learning, viz., those cases in which the learning parameter for a single response changes over trials. This latter situation is of more than academic interest in view of the prevalence of "warm-up" effects obtained with naive subjects in verbal learning experiments. In terms of the model, a "warm-up" effect in the learning of a single response would be reflected by a shift in the value of the conditioning parameter from the initial trials to the later trials of the experiment. Some evidence for such a shift with naive subjects has been obtained in experiments by the writer, but the effect is absent when well-practiced subjects are used. In applying the model to the "warm-up" phenomenon, we would identify two transient states which differ in the probability of conditioning the response. This scheme requires an additional parameter specifying the
probability of a transition each trial from the low-c state to the high-c state (the b constant in the matrix of Equation 9).

An issue not yet discussed here is that of the goodness of fit of the alternative models to the data of response compounding experiments. The alternatives mentioned before are (a) the AB compound is learned as a unit, (b) the A and B components are learned independently, and (c) some degree of dependence obtains between the two learning processes. These alternatives can be represented in terms of special values of the conditioning parameters of the general model. This being the case, the direct method for determining which case obtains for a set of data would involve estimating $c_1, c_2, a, b, \text{ and } d$ within the general model, and then testing special hypotheses about these parameters (e.g., that $b = d = 0$, or that $a + d = c_1$). The difficulty with this direct approach is that it requires some knowledge of the sampling distribution (at least the variance) of the several estimating statistics, and this knowledge is simply not available at the present time. It might be added that this situation is generally characteristic of the stochastic learning models studied to date, with a few notable exceptions (Restle, 1961). Failing in this direct approach, goodness of fit considerations must settle on second-best, approximate tests. One candidate for an approximate test is the Suppes-Atkinson (1960) chi-square test of how well a model reproduces the n-th order conditional probabilities of the response sequences. This is an approximate test of the present models because the chi-square procedure is really appropriate only when the model implies a finite Markov chain in the observable response sequence and none of the models considered here imply that
result. The second obvious candidate is the "visual inspection" test: a number of predictions of the alternative models are compared with data, and the model coming closer to the majority of the data statistics is accepted. The method is admittedly crude but for all that it has enabled research on stochastic models for learning to thrive over the past twelve years. Analysis is currently under way to determine which (if any) of the alternative models best describes results of an extensive experiment just completed on response compound learning.
Notation Used in Text

1. $g_1$ = probability of a correct guess on the $i$-th component.
2. $N$ = number of response alternatives.
3. $c_1, c_2, a, b, d$ = conditioning parameters.
4. $r = a + b + d$.
5. $n_i$ = the mean number of trials in transient state $i$ before absorption, starting out in $s_0$ on the first trial.
6. $T_x$ = total errors on response variable $x = \sum_{n=1}^{\alpha} x_n$.
7. $w_{i,n}$ = probability of being in state $i$ after $n$-1 reinforcements starting out in $s_0$.
8. $q_{i,n}$ = average error probability for $i$-th component on trial $n$.
9. $u_{j,n} = x_n x_{n+1} \ldots x_{n+j-1}$ = a $j$-tuple of errors.
10. $r_j$ = number of runs of exactly $j$ errors.
11. $c_{k,n} = x_n x_{n+k}$.
12. $\tau = (1-g_1)(1-c_1)$.
13. $\sigma = (1-g_1)(1-a-d)$.
14. $e_n$ = number of errors following an error on trial $n$.
15. $b_0$ = probability that there are no more errors following a guess with the subject in state 0.
16. $b_2$ = probability that there are no more errors following a guess with the subject in state 2.
17. $n'$ = trial of the last error.
18. $\gamma = \frac{(1-g_1)(1-r)}{1-g_1(1-r)}$
19. $\alpha = \frac{(1-g_1)(1-c_1)}{1-g_1(1-c_1)}$
20. $J_0$ = errors before the first success.

21. $J_k$ = errors between the $k$-th and $(k + 1)^{st}$ success.

22. $p_k$ = probability that there are no errors following the $k$-th success.

23. $f_{i,k}$ = probability that the subject is in state $i$ when the $k$-th success occurs.

24. $\mu = \frac{g_1(1-r)}{1-(1-g_1)(1-r)}$

25. $\beta = \frac{g_1(1-c_i)}{1-(1-g_1)(1-c_i)}$

26. $H_n$ = number of successes between an error on trial $n$ and the next error (provided there is a next error).

27. $H'_k$ = number of successes between the $k$-th and $(k + 1)^{st}$ error.

28. $t_{i,k}$ = probability that the subject is in state $i$ when the $k$-th error occurs.
REFERENCES


<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wright Air Development Center</td>
<td>Department of Philosophy</td>
<td>University of California, Los Angeles, CA, USA</td>
</tr>
<tr>
<td>Professor Ernest Adams</td>
<td>Michigan State University</td>
<td>East Lansing, MI, USA</td>
</tr>
<tr>
<td>Professor Alan Anderson</td>
<td>Yale University</td>
<td>New Haven, CT, USA</td>
</tr>
<tr>
<td>Professor S. K. Anderson</td>
<td>Columbia University</td>
<td>New York, NY, USA</td>
</tr>
<tr>
<td>Dr. R. F. Bales</td>
<td>Harvard University</td>
<td>Cambridge, MA, USA</td>
</tr>
<tr>
<td>Professor Edward H. Agate</td>
<td>Yale University</td>
<td>New Haven, CT, USA</td>
</tr>
<tr>
<td>Professor Gustav Bergman</td>
<td>State University of Iowa</td>
<td>Iowa City, IA, USA</td>
</tr>
<tr>
<td>Professor Max Black</td>
<td>Cornell University</td>
<td>Ithaca, NY, USA</td>
</tr>
<tr>
<td>Professor David Blackwell</td>
<td>University of California</td>
<td>Berkeley, CA, USA</td>
</tr>
<tr>
<td>Mr. Richard B. Rogers</td>
<td>U.C.L.A.</td>
<td>Los Angeles, CA, USA</td>
</tr>
<tr>
<td>Professor Lyle M. Burshe, Jr.</td>
<td>Department of Psychology</td>
<td>University of Utah, Salt Lake City, UT, USA</td>
</tr>
<tr>
<td>Professor C. J. Burke</td>
<td>Indiana University</td>
<td>Bloomington, IN, USA</td>
</tr>
<tr>
<td>Professor R. R. Bush</td>
<td>Pennsylvania State University</td>
<td>University Park, PA, USA</td>
</tr>
<tr>
<td>Dr. Donald Campbell</td>
<td>Department of Psychology</td>
<td>Northwestern University</td>
</tr>
<tr>
<td>Mr. J. Merrill Carlin</td>
<td>Department of Social Relations</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Professor Rudolf Carnap</td>
<td>Department of Philosophy</td>
<td>University of California, Los Angeles, CA, USA</td>
</tr>
<tr>
<td>Professor Edward C. Carterette</td>
<td>Department of Psychology</td>
<td>University of California, Los Angeles, CA, USA</td>
</tr>
<tr>
<td>Professor John Chomsky</td>
<td>Department of Philosophy</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>Professor C. V. Churchman</td>
<td>School of Business Administration</td>
<td>University of California, Berkeley, CA, USA</td>
</tr>
<tr>
<td>Professor James Coleman</td>
<td>Department of Social Relations</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>Dr. Clyde E. O'Connor</td>
<td>Department of Psychology</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Professor Gerard Debreu</td>
<td>Cowles Foundation for Research in Economics</td>
<td>Yale University, Box 2128</td>
</tr>
<tr>
<td>Dr. J. A. Dutsch</td>
<td>Center for Advanced Study in the Behavioral Sciences</td>
<td>Stanford, CA, USA</td>
</tr>
<tr>
<td>Professor Robert E. Kuczynski</td>
<td>Department of Economics</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Professor Burton Kuenen</td>
<td>Department of Philosophy</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Professor P. E. DeBose</td>
<td>Department of Psychology</td>
<td>Washington University</td>
</tr>
<tr>
<td>Dr. Ward Edwards</td>
<td>Department of Psychology</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Professor W. K. Estes</td>
<td>Department of Psychology</td>
<td>Indiana University</td>
</tr>
<tr>
<td>Professor Robert Fogel</td>
<td>Department of Psychology</td>
<td>University of Oregon</td>
</tr>
<tr>
<td>Professor Merrill M. Flood</td>
<td>271 West Engineering</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Professor Raymond Frankham</td>
<td>Department of Psychology</td>
<td>University of Illinois</td>
</tr>
<tr>
<td>Professor Milton Friedman</td>
<td>Department of Economics</td>
<td>University of Chicago</td>
</tr>
<tr>
<td>Dr. Eugene Galanter</td>
<td>Department of Psychology</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Professor John O'Malley</td>
<td>Department of Sociology</td>
<td>Columbia University</td>
</tr>
<tr>
<td>Name</td>
<td>Title</td>
<td>Department</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Dr. William Riesen</td>
<td>Professor</td>
<td>Center for Behavioral Sciences</td>
</tr>
<tr>
<td>Professor T. C. Keppel</td>
<td>Professor</td>
<td>Cowles Foundation for Research in Economics</td>
</tr>
<tr>
<td>Mr. Franklin Kreane</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor W. Krulik</td>
<td>Professor</td>
<td>Department of Statistics</td>
</tr>
<tr>
<td>Professor David LaFarge</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor Paul Lisenfeld</td>
<td>Professor</td>
<td>Department of Sociology</td>
</tr>
<tr>
<td>Dr. Bernhard Lieferman</td>
<td>Professor</td>
<td>Department of Social Relations</td>
</tr>
<tr>
<td>Professor Frank Logan</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor Donna Lube</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Dr. David McConnell</td>
<td>Professor</td>
<td>Economics</td>
</tr>
<tr>
<td>Professor William J. McMillan</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Dr. Kenneth Mardleke</td>
<td>Professor</td>
<td>Mechanical Engineering</td>
</tr>
<tr>
<td>Professor Jacob Marcus</td>
<td>Professor</td>
<td>Economics</td>
</tr>
<tr>
<td>Professor Samuel Melsch</td>
<td>Professor</td>
<td>Economics</td>
</tr>
<tr>
<td>Professor G. A. Miller</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor Richard A. Montague</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Dr. O. K. Moore</td>
<td>Professor</td>
<td>Department of Sociology</td>
</tr>
<tr>
<td>Professor Sirkyt Norman</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor Oscar Normand</td>
<td>Professor</td>
<td>Department of Economics and Social Institutions</td>
</tr>
<tr>
<td>Professor Frederick Nast</td>
<td>Professor</td>
<td>Department of Social Relations</td>
</tr>
<tr>
<td>Professor Ernest Nagel</td>
<td>Professor</td>
<td>Department of Behavioral Sciences</td>
</tr>
<tr>
<td>Professor Theodore N. Novitch</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor Jerry Newman</td>
<td>Professor</td>
<td>Department of Social Relations</td>
</tr>
<tr>
<td>Professor Vincent Novits</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor A. G. Papandreou</td>
<td>Professor</td>
<td>Department of Economics</td>
</tr>
<tr>
<td>Dr. Julian Popper</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Professor William F. Proctor, Jr.</td>
<td>Professor</td>
<td>Department of Psychology</td>
</tr>
<tr>
<td>Doug G. Pruitt</td>
<td>Professor</td>
<td>Political Science Department</td>
</tr>
<tr>
<td>Professor Hilary Pulman</td>
<td>Professor</td>
<td>Department of Philosophy</td>
</tr>
<tr>
<td>Professor Hillard V. Quine</td>
<td>Professor</td>
<td>Department of Philosophy</td>
</tr>
<tr>
<td>Professor Roy Radner</td>
<td>Professor</td>
<td>Department of Economics</td>
</tr>
<tr>
<td>Professor Edward Stowe</td>
<td>Professor</td>
<td>Department of Statistics</td>
</tr>
<tr>
<td>Dr. Anibal Rapport</td>
<td>Professor</td>
<td>Mental Health Research Institute</td>
</tr>
</tbody>
</table>