Interpretation of the Two-Stage Analysis of Paired-Associate Memorizing[1]

MICHAEL HUMPHREYS

University of British Columbia, Vancouver 8, Canada

AND

JAMES G. GREENO

University of Michigan, Ann Arbor, Michigan 48104

Four groups were run with response difficulty and stimulus difficulty varied factorially. A two-stage Markov model fit the data adequately. The parameter associated with the first stage depended on stimulus difficulty as well as response difficulty, refuting an interpretation of the first stage as response learning. The learning parameters associated with the second stage seemed to depend only on stimulus difficulty. The results suggest that the first stage of learning involves storage of the stimulus–response pair in memory, and the second stage involves learning to retrieve the item reliably.

Since Underwood, Runquist and Schulz (1959) introduced the idea that paired-associate memorizing involves two stages, response learning and associative learning, many investigators have used the idea in interpreting experimental results. The immediate stimulus for the present study was Kintsch's (1963) hypothesis that these two stages could be represented by a Markov chain with three states. In the first state, neither the response for an item nor the stimulus–response association has been learned. On some trial, the response becomes learned and the item goes into the intermediate state where there is some probability that the subject will give the correct response. Then on that trial or some later trial, the subject learns the association and this permits the subject to give the correct response on all later trials.

The experiment was designed so that the difficulty of response learning and the difficulty of association learning would vary independently. Two levels of stimulus difficulty and response difficulty were used with the intention of observing the changes in parameter values between groups. If Kintsch's version of the response-learning, association-learning hypothesis is correct, we should expect the probability of accomplishing the first stage of learning to differ between groups with different responses, but not to depend on stimulus difficulty.

EXPERIMENTAL METHOD

Materials. Response difficulty was manipulated by varying pronounceability, and stimulus difficulty was manipulated by varying the number of elements in each stimulus and thus the number of common elements between stimuli. There was no guarantee that these manipulations would have completely separate effects on the processes of response learning and association learning, although they had considerable face validity.

Two lists of stimuli and two lists of responses were combined to make four paired-associate lists. These were learned by separate groups. The stimuli were numerals and the responses were trigrams. The lists are shown in Table 1. The abbreviations EE, HE, EH, and HH will be used throughout this paper; the first letters refer to the stimulus difficulty and the second letters refer to the response difficulty.

TABLE 1

Experimental Lists

EE	EH	HE	HH
1-HAZ	1-HPF	11-RAS	11-GPS
2-MAK	2-IPW	12-MAK	12-HPF
3-GAW	3-NPE	13-JAV	13-BPC
4-RAS	4-GPS	21-BAQ	21-IPW
5-BAQ	5-JPV	22-HAZ	22-NPE
6-LAN	6-MPA	23-FAC	23-XPO
7-DAP	7-BPC	31-DAP	31-RPK
8-JAV	8-XPO	32-GAW	32-MPA

Procedure. Nine random orders of each list were constructed with the restrictions that each item started one order and ended one order, but no item could end one order and begin the next. The items were presented on a memory drum using the anticipation procedure at a 3.5:3.5-sec rate and a 21-sec intertrial interval. The subject was required to spell the responses. Each subject was run until he reached a criterion of three consecutive errorless trials on the list. Five subjects failed to reach the list criterion, one in group EH and four in group HH. These subjects were stopped after 34 to 37 trials.

Before learning the assigned experimental list, each subject was given six trials on a practice list with color names as stimuli and shape names as responses. These were presented at the same rate as the experimental list and the subjects were asked to spell their responses.

Subjects. Ten males and eight females were used in each group. The subjects were volunteers from the student body at Reed College. The subjects of each sex were randomly assigned within blocks of four to the experimental groups in the order of their participation in the experiment.

RESULTS

Sequences were analyzed through a criterion of five consecutive correct responses on each item. If there were not five consecutive correct responses on a sequence, the final run of successes was counted as the criterion run.

TABLE 2
Summary Data

	Errors		
Group	Before first correct	After first correct	Trials of last error
EE	3.10	1.17	5.71
HE	5.28	1.85	8.28
EH	5.06	1.57	8.01
HH	6.64	2.82	11.79

The data were analyzed using a relatively general Markov model where it is assumed that learning requires two stages, and there are no correct responses in the initial stage. The model is a generalization of the system analyzed by Bower and Theios (1964). In this application, we use statistical properties of the general model that were analyzed by Greeno (1968). There are four states: O, the initial state; E and C, intermediate states where errors and correct responses occur, respectively; and L, the learned state where only correct responses occur. The initial and transition probabilities are denoted

$$P[L(1), E(1), C(1), O(1)] = [t, (1 - s - t) r, (1 - s - t)(1 - r), s].$$

$$P = \begin{array}{c} \\ L(n) \\ E(n) \\ C(n) \\ O(n) \end{array} \begin{array}{|cccc} L(n+1) & E(n+1) & C(n+1) & O(n+1) \\ \hline 1 & 0 & 0 & 0 \\ d & (1-d)q & (1-d)p & 0 \\ c & (1-c)q & (1-c)p & 0 \\ ab & a(1-b)e & a(1-b)(1-e) & 1-a \end{array} \qquad (1)$$

There are nine theoretical parameters, but there are only seven identifiable parameters. However, certain hypotheses about the parameters are testable. We ignored responses made on the initial trial of the experiment because they all had to be errors. Therefore, the first counted responses were those on the trial following a single study presentation of all items. A natural assumption is that all items were in State O before the first study trial, and that the first study trial had the same effect as all later study trials. According to this assumption, the probabilities of the four states on the first counted trial should equal the probabilities in the bottom row of the transition matrix. In other words,

$$r = e,$$
$$s = 1 - a, \qquad (2)$$
$$t = ab.$$

These three restrictions on the theoretical parameters imply two restrictions on the seven identifiable parameters (Greeno, 1968, Eq. 62). Therefore, the hypothesis can be tested using a likelihood ratio test with $-2 \log \lambda$ asymptotically distributed as $\chi^2(2)$. The test was carried out using Subroutine Stepit (Chandler, 1965) to find parameters that minimize -2 times the log of the likelihood function (see Greeno, 1968, Eq. 41). The minimization was carried out separately for each group with all seven identifiable parameters free to vary, and then under the restrictions that follow from Eq. 2. The results are given in Table 3. The quantity assumed to be distributed as $\chi^2(2)$ is the difference between the two numbers in each row. Clearly, there is no evidence that we should reject the hypothesis. If we combined all the data, we would obtain a value of $-2 \log \lambda$ of 8.21, and the asymptotic distribution would be $\chi^2(8)$.

TABLE 3

Tests of the Hypothesis that the First Study Trial Was
Like Later Study Trials

| Stimuli | Responses | $-2 \log L$, | | p |
		Unrestricted	Restricted	
Easy	Easy	1685.52	1686.82	.25
Hard	Easy	2091.19	2094.30	.07
Easy	Hard	1932.34	1934.33	.16
Hard	Hard	2612.36	2614.17	.18

A further testable hypothesis is that $e = q$. If that is accepted, the hypothesis $b = d$ can be tested. The results of these tests are given in Table 4. The procedure used was the same as that one leading to Table 3. The results in Table 4 are differences between $-2 \log L$ minimized under the restrictions and $-2 \log L$ minimized under the more

TABLE 4

Tests of Hypotheses About Transition Parameters in Single Groups

Stimuli	Responses	Hypothesis: $e = q$		Hypothesis: $b = d$	
		$-2 \log \lambda$	p	$-2 \log \lambda$	p
Easy	Easy	0.81	.37	0.17	.65
Hard	Easy	1.21	.27	1.79	.18
Easy	Hard	1.07	.29	2.44	.12
Hard	Hard	10.13	.002	—	

general hypotheses. Each test has one degree of freedom. The column titled "Restricted" in Table 3 gives the general version used for testing $e = q$, and the value obtained with $e = q$ was the base line for testing $b = d$. The results indicate that both of the hypotheses $e = q$ and $b = d$ can be accepted for three of the groups, but $e = q$ has to be rejected for Group HH. In that group, then, $b = d$ is not a testable hypothesis. Since an important purpose of the investigation was to investigate parameter invariance among the different groups. it was decided that the analysis should be carried out with the largest number of free parameters needed for any group. Therefore, we let all the transition parameters be free, which means that there are five free identifiable parameters, and six theoretical parameters.

An important preliminary question is whether the two-stage theory is necessary for these data. We have found that we can specialize the theory in accord with a reasonable hypothesis about the first trial. Can we specialize it further to make it an all-or-none theory? The all-or-none theory that would result has two free parameters (Greeno and Steiner, 1964). The present theory has five. Thus, if we minimize $-2 \log L$ for the all-or-none theory and subtract the minimum value obtained with the two-stage theory, we obtain a statistic that may be distributed as $\chi^2(3)$, if learning really was all-or-none. The obtained values of this statistic were 185.3, 252.8, 222.7, 352.1, respectively, for the four groups. There seems little doubt that the learning process involved at least two stages.

Whether or not we can account for the data with just a two-stage theory depends on goodness-of-fit. First, we will present data obtained after the first correct response. Figure 1 shows the predicted and obtained distributions of errors after the first correct response in the four groups. The data are represented by the histograms, and the theoretical predictions are shown by the connected dots. The predictions were obtained using estimates of five parameters which maximized the likelihood for all the data, rather than for the specific distributions tested here. The goodness-of-fit to these distributions was tested using the Kolmogorov-Smirnov test and a chi-square test. In the K-S test, there is a probability of .20 of getting a maximum absolute discrepancy
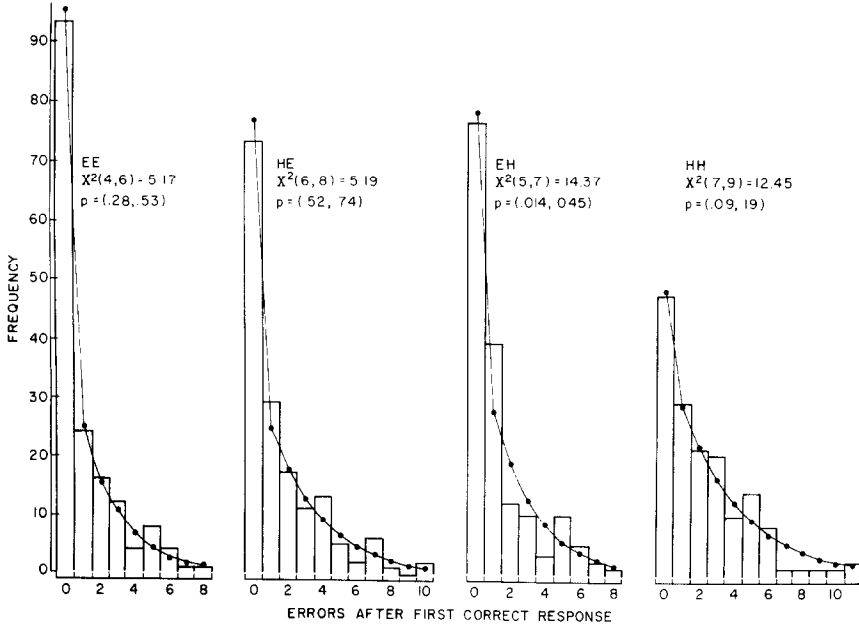
FIG. 1. Theoretical and empirical distributions of the number of errors after the first correct response.

between cumulatives as high as .084. The maximum discrepancies observed in these four cases were .018, .056, .021, and .046.

The degrees of freedom for the chi-square test are not well defined since the estimation did not minimize this chi-square statistic (Chernoff and Lehman, 1954). However, bounds on the distribution are obtained by considering two chi-square distributions with $N - 1$ and $N - m - 1$ degrees of freedom, where m is the number of estimated parameters. Two parameters need to be estimated to determine the theoretical distributions in Fig. 1 (Greeno, 1968, Eq. 31). The chi-square statistics are shown in Fig. 1 in the following form: $\chi^2(i, j) = x, p = (y, z)$. i and j are the lower and upper bounds on the degrees of freedom, x is the obtained chi-square statistic, and y and z are the significance levels at which we can reject the hypothesis of an exact fit if we use the lower and upper bounds on the degrees of freedom.

Figure 2 shows the predicted and obtained number of trials after the first correct response on which the last error occurred. If there were no errors after the first correct response, this variable was set at zero. For the Kolmogorov-Smirnov test, the maximum absolute deviations were .031, .029, .025, and .062. The chi-square values are given in the figures again, in the same form as was used for the error distributions. Again, the theoretical distribution is determined by two parameters (Greeno, 1968, Eq. 32).
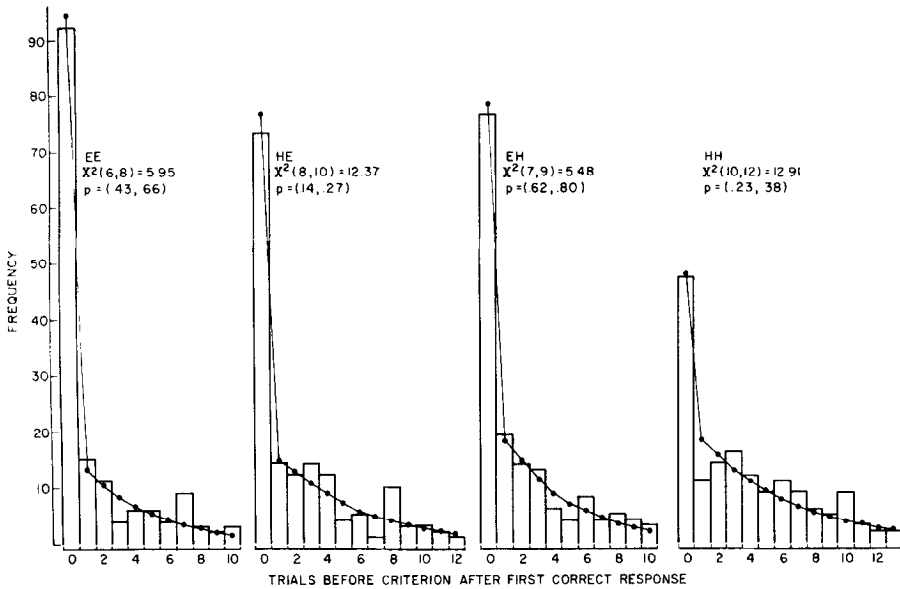
FIG. 2. Theoretical and empirical distributions between the first correct response and criterion.

A final prediction about responses after the first correct response is that errors and correct responses before criterion should be independent. A chi-square test with one degree of freedom is appropriate. The obtained values were 0.16, 0.27, 0.004, and 3.26. The last value has $p = .07$. When the item criterion is set at four consecutive correct, the obtained values of the chi-square statistic were 1.10, 1.93, .79, and .012; when set at six consecutive correct, the obtained values were .02, 1.00, 6.59, and 6.06. The criterion of five was chosen as a compromise, including as much of the error data as possible while still excluding the most divergent points obtained at the end of the experimental session.

It seems reasonable to conclude that the fit of the model to data after the first correct response was quite good. This is important, regardless of the outcome of the remaining analyses. It says that the learning which occurred on an item after the first correct response was all-or-none to a close approximation.

The next data are the distributions of numbers of errors before the first correct response. The distribution of the total number of sequences with j errors before the first correct response was evaluated by the Kolmogorov-Smirnov test. Again, the test has a critical value of .084 if we adopt $\alpha = .20$. The obtained maximum absolute deviations were .037, .051, .040, and .055. The data shown in Fig. 3 have the distributions broken down into components. The distributions in the lower panel have sequences with no errors after the first correct response. The distributions in the upper panel

have sequences with one or more errors after the first correct response. All five of the identifiable parameters are involved in the distribution (Greeno, 1968, Eq. 28), which means that the distribution of the chi-square statistic is bounded between $\chi^2(N-1)$ and $\chi^2(N-6)$. The values of the chi-square statistic and bounds on the significance levels at which we could reject the hypothesis of an adequate fit are given in Fig. 3.
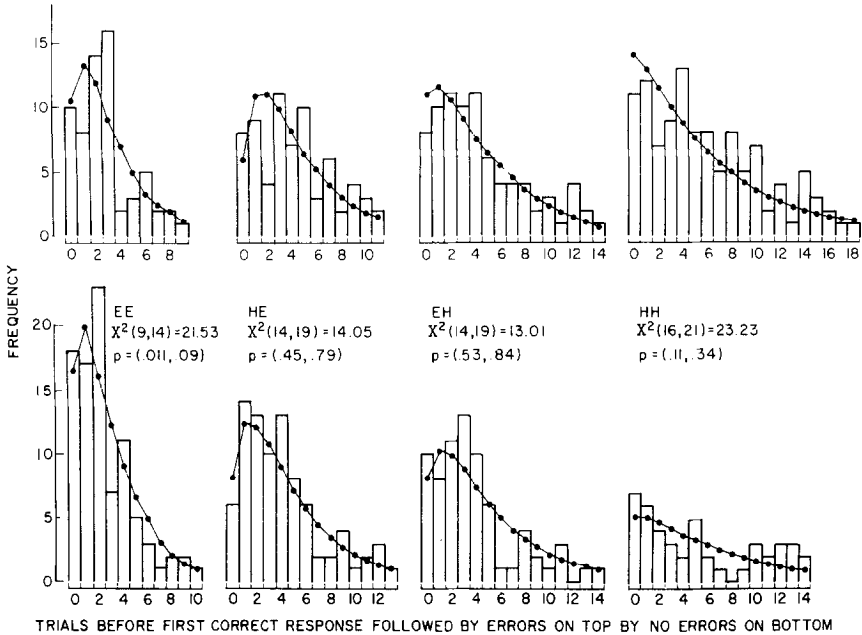

FIG. 3. Theoretical and empirical distributions of the number of errors before the first correct response, partitioned according to the presence or absence of errors after the first correct response.

Finally, we consider statistics based on complete sequences. The predicted and obtained distributions of total errors are in Fig. 4. The values of the Kolmogorov-Smirnov test statistic are .032, .068, .044, and .036. The theoretical distribution can be obtained by knowing three identifiable parameters (Greeno, 1968, Eq. 36). The chi-square statistics are given in Fig. 4 as before.

The distributions of trial of last error are given in Fig. 5. The Kolmogorov-Smirnov test statistics were .056, .063, .049, and .049. All five identifiable parameters are needed (Greeno, 1968, Eq. 37) so the bounds on the distribution of chi-square involve distributions differing in five degrees of freedom.

It seems to us that the statistical tests warrant the conclusion that the model and the data agree to a satisfactory approximation. Of 20 distributions that were examined,
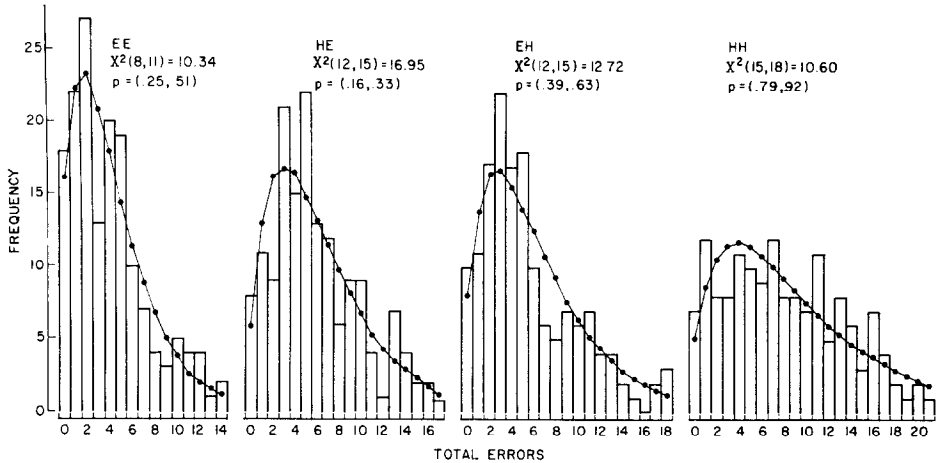
FIG. 4. Theoretical and empirical distributions of the total number of errors.
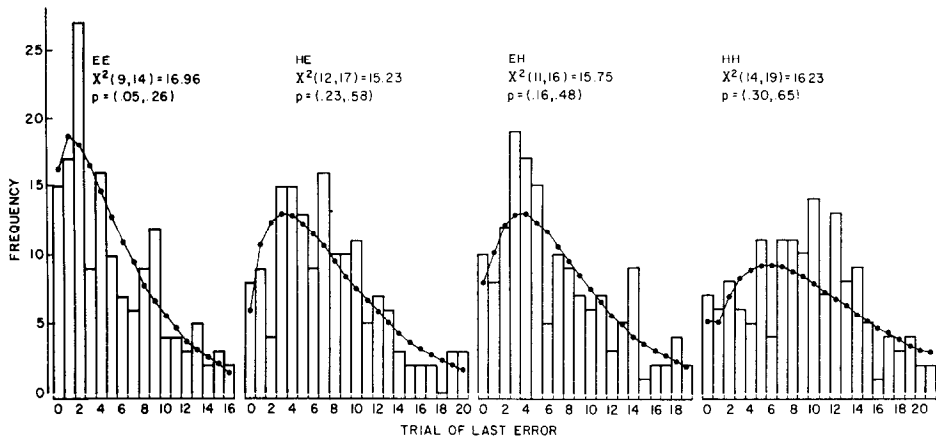
FIG. 5. Theoretical and empirical distributions of the trials of the last error before criterion.

there were no cases in which the model could be rejected by the Kolmogorov-Smirnov test. Using the chi-square test, one distribution permits rejection at $\alpha = .05$ if we adopt a criterion that probably is too weak, and three rejections at $\alpha = .05$ result if we use a criterion that probably is too strong. The hypothesis of independence between the first correct response and the criterion could not be rejected. Therefore, it seems reasonable to proceed with the investigation of parameter invariance.

We were not clear at the outset whether or not the data would provide strong leverage on questions of invariance for a model with five parameters. We began by

testing a few sample hypotheses, but this did not lead to any striking insights about the matter of invariance. What we finally did was to test all possible hypotheses about parameter invariance regarding all four pairs of groups where either the stimuli or the responses were the same. The pattern of results that emerged from this bulldozing operation turned out to make quite a bit of sense.

The results of the analysis are given in Table 5. The entries on the left under "Hypothesis" show which parameter(s) was held constant across groups for that test. A "1" indicates that a parameter was held constant, and a "0" indicates that a parameter was permitted to vary between groups. For example, the first row of Table 5 has tests of the hypothesis that a was a constant between pairs of groups; the second row has tests of the hypothesis that b and c were constant between pairs of groups.

The entries in Table 5 are test statistics for the likelihood ratio test and in parentheses the probability under the null hypothesis of a test statistic that large or larger. With no restrictions about parameter invariance, the minimum value of $-2 \log L$ for two groups is just the sum of the values of $-2 \log L$ for those groups from the second column in Table 3. For each test indicated in Table 5, we used Stepit to minimize $-2 \log L$ for the data of two groups, under the restriction given. The test statistic is the difference between the value obtained under the restriction and the value obtained without any restrictions. For example, for each test in the first column, the test statistic is the difference between the minimum $-2 \log L$ under the stated invariance restriction and the value $1686.82 + 2094.30 = 3781.12$.

TABLE 5

Tests of Parameter Invariance Across Pairs of Groups

Hypothesis: abcdeq	Groups			
	EE HE	EH HH	EE EH	HE HH
100000	5.97(.015)	6.69(.010)	14.23(.0002)	16.23(.00006)
011000	1.21(.28)	0.09(.77)	0.47(.50)	0.00(1.0)
010100	0.98(.33)	0.01(.93)	0.25(.62)	0.00(1.0)
010010	1.17(.28)	1.22(.27)	0.15(.70)	6.24(.013)
010001	5.69(.018)	0.04(.85)	4.66(.031)	0.08(.78)
001100	0.02(.89)	5.34(.021)	0.00(1.0)	0.30(.59)
001010	0.16(.69)	1.80(.18)	0.44(.51)	7.82(.006)
001001	3.44(.064)	0.05(.83)	8.80(.004)	1.61(.21)
000110	0.30(.59)	1.15(.29)	0.53(.47)	6.69(.010)
000101	6.16(.014)	2.32(.13)	5.60(.018)	2.77(.10)
000011	0.00(1.0)	1.69(.20)	1.84(.18)	8.61(.004)
111000	6.59(.038)	9.40(.010)	14.77(.0007)	17.17(.0002)
110100	6.41(.041)	6.69(.036)	16.03(.0004)	16.28(.0003)

TABLE 5 (*continued*)

Hypothesis: abcdeq	Groups			
	EE HE	EH HH	EE EH	HE HH
110010	13.14(.002)	6.76(.035)	18.09(.0002)	16.57(.0003)
110001	12.14(.003)	8.41(.015)	23.95(nil)	18.68(.00009)
101100	6.02(.050)	13.32(.002)	14.23(.0009)	17.45(.0002)
101010	9.36(.010)	6.69(.036)	16.20(.0004)	17.07(.0002)
101001	9.82(.008)	6.77(.034)	22.90(nil)	17.92(.0002)
100110	9.94(.007)	6.77(.034)	16.14(.0004)	16.73(.0003)
100101	13.79(.002)	9.50(.009)	21.72(nil)	20.43(.00004)
100011	7.82(.021)	6.69(.036)	14.86(.0006)	17.43(.0002)
011100	1.21(.55)	5.38(.08)	0.74(.70)	0.69(.71)
011010	2.88(.24)	7.99(.019)	1.16(.56)	10.57(.006)
011001	5.92(.052)	0.14(.94)	9.23(.010)	1.61(.45)
010110	0.88(.65)	0.22(.90)	1.36(.51)	6.71(.035)
010101	9.88(.008)	2.32(.32)	7.49(.024)	2.77(.26)
010011	6.94(.032)	3.24(.20)	10.12(.007)	11.50(.004)
001110	0.31(.86)	8.16(.017)	0.64(.73)	9.33(.01)
001101	6.17(.046)	5.38(.068)	9.05(.011)	2.77(.26)
001011	3.71(.16)	1.91(.39)	9.20(.011)	9.08(.011)
000111	6.18(.046)	3.99(.14)	7.31(.026)	11.34(.004)
111100	6.59(.087)	16.15(.002)	16.25(.002)	19.87(.0002)
111010	14.46(.003)	13.79(.004)	18.76(.0004)	11.08(.012)
111001	12.25(.007)	9.46(.024)	30.48(nil)	18.86(.0003)
110110	13.49(.004)	7.76(.052)	18.88(.0003)	16.98(.0008)
110101	15.84(.002)	10.73(.014)	26.88(nil)	11.25(.011)
110011	19.14(.0003)	8.96(.030)	28.61(nil)	22.52(.00006)
101110	9.95(.019)	13.37(.004)	16.40(.001)	18.65(.0004)
101101	13.79(.004)	13.40(.004)	24.03(.00003)	20.49(.0002)
101011	14.11(.003)	7.77(.052)	26.16(nil)	18.59(.0004)
100111	16.68(.0009)	9.52(.024)	22.91(.00005)	21.31(.0001)
011110	2.90(.41)	17.56(.0006)	2.48(.48)	12.32(.007)
011101	10.52(.015)	5.43(.15)	9.50(.024)	2.77(.43)
011011	7.34(.062)	8.28(.05)	11.55(.010)	12.04(.008)
010111	10.58(.015)	4.39(.23)	12.71(.006)	12.60(.006)
001111	6.20(.11)	8.38(.04)	9.63(.022)	11.88(.008)
111110	14.48(.006)	23.74(.0001)	19.96(.0006)	22.85(.0002)
111101	17.27(.002)	16.20(.003)	32.05(nil)	21.56(.0003)
111011	19.99(.0006)	14.01(.008)	30.48(nil)	23.02(.0002)
110111	24.46(.00007)	10.23(.037)	32.20(nil)	23.90(.00009)
101111	16.94(.002)	13.46(.010)	26.76(.00003)	21.53(.0003)
011111	12.96(.012)	17.92(.002)	12.72(.013)	16.71(.003)
111111	25.83(.0001)	24.06(.0003)	32.21(nil)	28.32(.00004)

The tests in Table 5 are grouped by degrees of freedom. Since a is an identifiable parameter, the hypothesis of its invariance is testable. However, the remaining parameters are not identifiable. This means that, for example, the value of b could be set arbitrarily (within limits) and the remaining parameters could be estimated and the goodness of fit would be just the same as if b had been set arbitrarily at a different value. Of course, in this situation a test of invariance of the parameter b, c, d, e, or q is not testable. However, if we test the simultaneous invariance of two of the above parameters, we have a hypothesis which does place a restriction on the data and is testable. We have set up Table 5 and calculated significance levels with the assumption that there is one degree of freedom if a is held constant and the number of additional degrees of freedom equals one less than the number of nonidentifiable parameters that are held constant. We do not have a formal proof that this is correct; however, we feel reasonably confident about the conjecture that this gives the right number of degrees of freedom for the chi-square distribution in the likelihood ratio test.

Now recall that the main theoretical interest in this investigation is the hypothesis that paired associates are learned in two stages, with the first stage consisting of response learning. If the first stage of learning in the Markov model consisted of response learning, then we should find that the parameter a is approximately the same in groups having the same responses. In the statistical tests shown in Table 5, we would expect to accept null hypotheses of parameter invariance for a in the first two columns. The first test shown (100000) is the one most directly relevant to the question, since it tests the invariance of a with all the other parameters free to vary. The results indicate rejection of the invariance of a. On the other hand, the tests involving the various parameters are not orthogonal, and many of the tests involving parameters other than a permit acceptance of null hypotheses. It is possible that a could be held constant across a pair of groups with the same responses if some acceptable restriction were imposed on the other parameters. However, this possibility did not materialize. There were 37 hypotheses about invariance of some subset of the parameters b, c, d, e, and q across Groups EE and HE or across Groups EH and HH that were acceptable ($\alpha = .05$). In each of these 37 cases, we can test the invariance of a and the other parameters as a null hypothesis, with the invariance of the other parameters as the alternative. In each of these tests, the test statistic $(-2 \log \lambda)$ is the difference between two values given in Table 5, and the test has one degree of freedom. (For example, 111000 is one null hypothesis, and 011000 is its alternative. For Groups EE and HE, the test statistic is $6.59 - 1.21 = 5.38$.) None of these 37 tests indicated that the invariance of a was an acceptable hypothesis at $\alpha = .05$. We conclude that the value of a must have depended on both the stimulus and response variables in this situation.

The situation regarding the remaining parameters is less clear, which is understandable since their estimates probably interact in complex ways. However, if we look at the tests with two degrees of freedom, we see that the hypothesis of invariance for b, c, and d could be accepted with $\alpha = .05$ for all pairs of groups; and in the tests with three

degrees of freedom, all of the acceptable invariance hypotheses involve at least two of the three parameters b, c, and d. It seemed reasonable, then, to test hypotheses about the invariance of b, c, and d across all groups. Recall that b, c, and d are not identifiable parameters for data from a single group. However, if two of these are assumed to be constant across four experimental conditions, we have a testable restriction. We are working with a model that has five identifiable parameters; thus, the general parameter space for four groups has 20 dimensions. Estimating values of a, e, and q separately for four groups uses 12 dimensions. Now suppose that we assume that b and c are constant across four groups, with d varying. This gives four values of d and a single value of b and one of c; the effect is to add 6 parameters to the 12 used for a, e, and q, giving 18 in all. Since this is two fewer than the number of identifiable parameters with no restrictions, we conclude that assuming that b and c are constant across the four groups is a testable restriction. Of course, the same argument holds regarding the assumption that b and d are constant with c varying, and the assumption that c and d are constant with b varying. We have assumed that under each of these null hypotheses, the test statistic for the likelihood ratio test is distributed approximately as chi-square with two degrees of freedom.

The tests were carried out using the same procedure as with pairs of groups. Stepit was used to minimize $-2 \log L$ across all four groups under each restriction. The minimum of $-2 \log L$ without a restriction is equal to the sum of the four values in the second column of Table 3. Tables 6, 7, and 8 show the obtained parameter values under the three restrictions involving two parameters at a time. All of these tests permit acceptance of the null hypothesis.

We also tested the hypothesis that all three of the parameters b, c, and d were constant across all four groups. However, this null hypothesis was unacceptable for our data. With b, c, and d held constant across all four groups there are 15 parameters estimated, so when we compare the minimum $-2 \log L$ with the value obtained without invariance restrictions the test statistic should have the chi-square distribution with five degrees of freedom. The obtained value of $-2 \log \lambda$ was 14.14, which has $p < .02$.

TABLE 6

Parameters and $-2 \log \lambda$ for Invariance of b and c

Stimuli	Responses	a	b	c	d	e	q
Easy	Easy	.30	.26	.00	.36	.74	.54
Hard	Easy	.21	.26	.00	.28	.63	.60
Easy	Hard	.19	.26	.00	.33	.52	.64
Hard	Hard	.13	.26	.00	.25	.11	.64

$-2 \log \lambda = 3.95; \ p = .14$

TABLE 7

Parameters and $-2 \log \lambda$ for Invariance of b and d

Stimuli	Responses	a	b	c	d	e	q
Easy	Easy	.29	.16	.19	.22	.55	.43
Hard	Easy	.22	.16	.10	.22	.68	.56
Easy	Hard	.18	.16	.18	.22	.39	.56
Hard	Hard	.14	.16	.06	.22	.17	.61

$-2 \log \lambda = 2.03; p = .37$

TABLE 8

Parameters and $-2 \log \lambda$ for Invariance of c and d

Stimuli	Response	a	b	c	d	e	q
Easy	Easy	.29	.24	.18	.17	.57	.42
Hard	Easy	.22	.06	.18	.17	.63	.51
Easy	Hard	.19	.14	.18	.17	.47	.54
Hard	Hard	.13	.00	.18	.17	.06	.56

$-2 \log \lambda = 5.62; p = .06$

Up to this point, the study of parameter invariance had been fairly mechanical, and apparently we are close to the limit of acceptable hypotheses of invariance. Two more hypotheses were considered that arose in an *ad hoc* consideration of the results obtained thus far. First, we went back to the results in Table 4 where in three of the four groups it was possible to test the hypothesis $b = d$ independently of other assumptions about parameters. In all of those cases the hypothesis was acceptable. Therefore, it seemed that it would be worthwhile to investigate the effect of introducing that restriction now, in the context of a restricted parameter space with some of the parameters held constant across groups. The second consideration came from examining Tables 6, 7, and 8. In all those cases restrictions on two learning parameters were acceptable, and the third learning rate then apparently covaried with stimulus difficulty, more or less independently of response difficulty. This fact made us suspicious of the invariance obtained earlier. We know that acceptance of a hypothesis about parameters may be misleading because the model is not identifiable. Therefore, the apparent equality of parameters across all four groups could arise because in each group one parameter can be adjusted to take up slack.

The fact that at least one learning parameter covaries with stimulus difficulty led us to the following hypotheses about the two stages of paired-associate learning. Suppose that the learning of a paired associate requires two achievements: (1) the

pair (i.e., the stimulus–response combination) has to be stored in memory, and (2) the subject must acquire a retrieval strategy which permits him to use the stimulus of the pair as a cue, find the pair in memory, and perform the response. Then the first stage of learning would involve storage of the pair in memory. Before that happened the probability of giving the correct response would be zero. The learning process would be in the intermediate stage on the trials when the item was stored in memory but a retrieval strategy had not yet been acquired. During the intermediate stage there would be some (presumably) nonzero probability of performing the correct response. The final stage of learning would involve the acquisition of the retrieval strategy which would permit the subject to use the stimulus as a cue for finding the pair in memory and then outputing the response. This hypothesis about paired-associate memorizing is similar in structure to Kintsch and Morris' (1965) hypothesis about free recall learning.

A specific version of the preceding story is partially illustrated in Fig. 6.
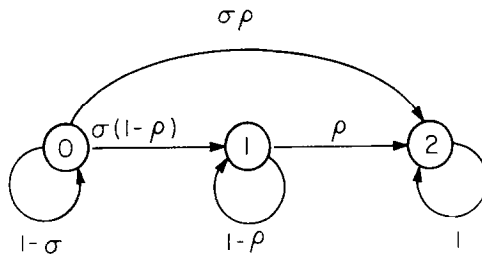

FIG. 6.   Graph representing a model with stimulus–response storage and acquisition of retrieval strategy.

The nodes represent stages of learning: In State 0, the pair has not yet been stored in the subject's memory. In State 1, the pair is in memory, but the subject does not have a retrieval strategy for it. And in State 2, the pair has become learned. σ and ρ represent transition probabilities as follows: σ is the probability of storing a pair in memory, ρ is the probability of acquiring a retrieval strategy. It is assumed that in State 0, the probability of a correct response is zero; in State 1 a correct response occurs with some nonzero probability.

The story relates to the parameters in the following way: First, since we are supposing that storage in memory involves the stimulus–response combination, the value of σ should depend on both stimulus and response variables. On the other hand, if the retrieval strategy depends mainly on learning how to use the stimulus as a search cue, the value of ρ should depend mainly on stimulus variables, and especially on the similarity between the stimuli.

Now we return to the parameters of the two-stage Markov model. Clearly, we have $a = \sigma$, and the story is consistent with our finding that a depended on both stimulus

and response difficulty. There are at least two ways to map ρ into the theory. One way is to suppose that on each trial in State 1, there is probability ρ that a retrieval strategy will be acquired. If we include the assumption that a retrieval strategy can be acquired on the trial when storage occurs, we have

$$b = c = d = \rho.$$

On the other hand, we may assume that the retrieval strategy is acquired though a process of selection, so that if a correct response occurs in State 1, the subject will not acquire a retrieval strategy; in other words, the transition from State 1 to State 2 occurs only on errors. This version of the story has

$$b = d = \rho, \qquad c = 0.$$

In either case, we expect ρ to depend mainly on stimulus variables.

Both versions of this hypothesis were tried, and both of them produced acceptable hypotheses about the parameters. Table 9 shows the results obtained using b, c, and d all equal, with one value for groups with easy stimuli and another value for groups with hard stimuli. Table 10 shows the results obtained using $c = 0$ in all groups, with b and d equal to one value for groups with easy stimuli and another value for groups with hard stimuli. In both cases there are 14 free parameters for the groups, which means that we have six degrees of freedom when we compare the minimum $-2 \log L$ with the value obtained without invariance restrictions.

Tables 9 and 10 raise some intriguing questions about the possible invariance of e and q that could lead to rather interesting hypotheses in connection with our proposed version of a two-stage learning process. For example, we investigated the hypothesis $e = q$ earlier, and we were able to accept that for all but the group with hard stimuli and hard responses. (This result was presented in Table 4.) Were we to investigate $e = q$ now with the invariance of parameters that has developed, it looks as if we could accept it for the groups with easy responses, but not for either group with hard

TABLE 9
Parameter Estimates and $-2 \log \lambda$ for $b = c = d = \rho$,
with ρ Depending Only on Stimulus Difficulty

Stimuli	Response	a	b	c	d	e	q
Easy	Easy	.29	.20	.20	.20	.55	.43
Hard	Easy	.22	.15	.15	.15	.68	.52
Easy	Hard	.18	.20	.20	.20	.34	.55
Hard	Hard	.13	.15	.15	.15	.00	.57

$-2 \log \lambda = 8.44; p = .21$

TABLE 10

Parameter Estimates and $-2 \log \lambda$ for $c = 0$, $b = d = \rho$,
with ρ Depending Only on Stimulus Difficulty

Stimuli	Response	a	b	c	d	e	q
Easy	Easy	.29	.34	.00	.34	.68	.54
Hard	Easy	.21	.26	.00	.26	.66	.60
Easy	Hard	.18	.34	.00	.34	.38	.64
Hard	Hard	.13	.26	.00	.26	.10	.64

$-2 \log \lambda = 4.37$; $p = .62$

responses. This could lead to some interesting speculations, but we feel that we should leave the situation as it now stands, pending further experimental work to check the reliability of the tentative conclusions presented already.

SUMMARY AND CONCLUSIONS

First, we found that the data from all groups could be described to a satisfactory approximation by the assumption of a two-stage learning process, with each stage consisting of an all-or-none event. Of course, this does not constitute a strong demonstration that the learning process involves exactly two discrete steps. However, we take the view that a two-stage process should be preferred over more complicated possibilities unless there is evidence requiring the move to a more complex structure.

Within the framework of a two-stage system, we have rather firm evidence against one hypothesis that we and others had considered as the most likely candidate. The refuted hypothesis postulates an initial stage of response learning, followed by a stage in which associative hookups are learned. The hypothesis led us to expect that the probability of accomplishing the first stage of learning should depend on response difficulty; actually, we found that the relevant parameter depended almost as strongly on stimulus difficulty as on response difficulty.

Our results led to an alternative hypothesis about the nature of the two learning stages. We propose that the first stage involves storage in memory of the stimulus–response pair, and that the second stage involves learning how to retrieve the correct response using the stimulus as a search cue. The evidence regarding our constructive suggestion seems to us to be encouraging, but far from definitive. Mainly, our reticence comes from the biographical fact that we thought of the hypothesis after, rather than before, we saw the results of the analysis. We have the classical problem of cross-validation; we observed a pattern of correlations between parameters and variables, and we think there might be some interesting processes that produced the observed

pattern. On the other hand, the observed pattern could have been produced by accident, or by something peculiar to the specific experiment that was carried out. We intend to carry out further studies on this problem to see whether our results can be replicated. In the meantime, the evidence against the two-stage theory with response acquisition in the initial stage seems firm and worth putting into the record.

REFERENCES

BOWER, G. H., AND THEIOS, J. A learning model for discrete performance levels. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford Univer. Press, 1964. Pp. 1–31.

CHANDLER, J. P. Subroutine Stepit. Program OCPE 66. Quantum Chemistry Program Exchange, Indiana University, Bloomington, Indiana. 1965.

CHERNOFF, H., AND LEHMANN, E. L. Use of maximum likelihood estimate in chi-square tests of goodness of fit. *Annals of Mathematical Statistics*, 1954, **25**, 579–586.

GREENO, J. G. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika*, 1968, **33**, 173–215.

GREENO, J. G. AND STEINER, T. E. Markovian Processes with identifiable states: general considerations and application to all-or-none learning. *Psychometrika*, 1964, **29**, 309–333.

Kintsch, W. All-or-none learning and the role of repetition in paired-associate learning. *Science*, 1963, **140**, 310–312.

Kintsch, W., and Morris, C. J. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology*, 1965, **69**, 200–206.

UNDERWOOD, B. J., RUNQUIST, W. N., AND SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.