MULTI-PROCESS MODELS FOR MEMORY WITH APPLICATIONS TO A CONTINUOUS PRESENTATION TASK

R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin

Stanford University

TECHNICAL REPORT NO. 96

April 13, 1966

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Multi-Process Models for Memory with Applications to a Continuous Presentation Task[1]

R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin

Stanford University

Abstract

A multi-process model for memory and learning is applied to the results of two complementary experiments. In Experiment I the subject was required to keep track of the randomly changing responses associated with a fixed set of stimuli. The task involved a lengthy and continuous sequence of trials, each trial consisting of a test on one of the stimuli followed by study on that same stimulus paired with a new response. The size of the stimulus set, s, took on the values 4, 6, and 8. Experiment II differed from Experiment I in that a large number of stimuli were used even though in any experimental condition the subject was required to remember only 4, 6, or 8 stimuli at one time. In both experiments the basic dependent variable was the probability of a correct response as a function of the number of intervening trials between study and test on a given stimulus-response pair (called the "lag"). The lag curves were all near 1.0 at lag 0 and monotonically decreased as the lag increased; the lag curves for the three conditions (s = 4, 6, and 8) decreased at different rates in Experiment I, whereas in Experiment II these curves were identical. Using four estimated parameters the model generated accurate predictions for the various response measures collected.

---

[1]This research was supported by the National Aeronautics and Space Administration, Grant No. NGR-05-020-036.

A quantitative model for human memory and learning has been proposed by Atkinson and Shiffrin (1965). Specific versions of the general model have been used to predict serial position curves obtained from free-verbal recall and paired-associate experiments. The variables which have been successfully handled include list length, presentation rate, and in a study by Phillips, Shiffrin, and Atkinson (1966), confidence ratings. These previous studies were all conducted with a discrete-trial procedure, i.e., the presentation of an entire list of items was followed by a single test. In the present study it was desired to test the model in a situation involving a continuous succession of study and test items. Additionally, the present study involved the manipulation of certain experimental variables that have logical relationships to model parameters. The specific experimental variable manipulated was the size of the stimulus set being remembered by a subject.

The task employed in the experiments to be described here involves a modification of the typical paired-associate procedure which makes it possible to study the memory process under conditions that are quite uniform and stable throughout the course of an experiment. This is the case because the task is continuous and each subject is run for 10 to 12 daily sessions.[2] In essence the task involves having the subject keep track of the randomly changing response members of s different stimuli. Each trial of the experiment is divided into a test period and a study period. During the test

---

[2]The task is similar to those used by Yntema and Mueser (1962), Brelsford, Keller, Shiffrin, and Atkinson (1966), and Katz (1966).

phase a stimulus is randomly selected from among the set of s stimuli and the subject tries to recall the response last associated with that stimulus. Following the test, the study phase of the trial occurs. During this phase, the stimulus used in the test phase of the trial is re-paired with a new response for study. Thus every trial is composed of a test and study period on the same stimulus. Following each trial a new stimulus is chosen randomly from the set of s stimuli and the next trial begins. The instructions to the subject require that on a test he is to give the response that was paired with the stimulus the last time it was presented for study.

The number of trials intervening between study and test on a given stimulus-response pair will be referred to as the "lag" for that item. Thus, if the test occurs immediately following the study period the lag is zero. If one trial intervenes (involving test and study on another stimulus), then the lag is 1; and so on. It should be clear that in this task the number of stimulus-response pairs that the subject is trying to remember at any given time is fixed throughout an experimental session. Each time a stimulus is tested it is immediately re-paired with a new response, keeping the size of the to-be-remembered stimulus set always equal to s. Of course, in order to start an experimental session, an initial series of trials must be given with the test phase omitted. The stimuli presented during these study trials are the ones used throughout the rest of the experimental session. In the present experiments there were three experimental conditions in which the size of the stimulus set, s, was either 4, 6, or 8. For each daily session, a subject was randomly assigned to one of these three conditions. The principal dependent variable is the probability of a correct response as a function of lag.

Model

The model assumes three memory states: a very short-lived memory system called the sensory buffer; a temporary memory state called the memory (or rehearsal) buffer; and a long-term storage state called LTS. In the discussion of the model which follows, reference is frequently made to the term "stimulus-response item." Items are postulated to enter and leave the two buffers at various times. At the outset, the question arises, what is an item? In terms of the present model an item will be defined as that amount of information that allows one to make a correct recall when a stimulus is presented for a test. The specification of the exact form of this information (i.e., whether it be acoustic rehearsal, visual imagery, or some type of mnemonic) is not within the scope of the present paper. Nevertheless, in view of the work of Conrad (1964), Wickelgren (1965), and others on auditory confusions in short-term memory, we would be satisfied with the view that items in the memory buffer are acoustic mnemonics and are kept there via rehearsal, at least for experiments of a verbal character.

The Sensory Buffer

It is assumed that all external stimulation coming into the system enters the sensory buffer, resides there for a short time (perhaps on the order of a few seconds), decays and is lost.[3] In the context of the present experiment it will be assumed that every item enters the sensory buffer. Furthermore, it will be assumed that a test follows the preceding study period closely enough in time so that an item will always be recalled

---

[3] We imagine that the form of the decay is roughly representable by the results from the Peterson and Peterson (1959) experiment on the decay of a consonant trigram in the absence of rehearsal.

correctly if it is tested immediately following its entry into the buffer. Therefore, since every item enters the sensory buffer, the probability of a correct recall at lag 0 will be unity. For lags greater than zero, items will have decayed, and the sensory buffer will have no further significance. For this reason, in the remainder of this paper, the term buffer when used by itself will refer to the memory buffer.

The Memory Buffer

The memory buffer is postulated to have a limited and constant capacity for homogeneous items. It may be viewed as a state containing those items which have been selected from the sensory buffer for repeated rehearsal. Once the memory buffer is filled, each new item which enters causes one of the items currently in the buffer to be lost. It is assumed that the series of study items at the start of each experimental session fills the buffer and that the buffer stays filled thereafter. The size of the buffer, r (defined as the number of items which can be held simultaneously), depends upon the nature of the items and thus must be estimated for each experiment. It is assumed that a correct response is given with probability one if an item is in the buffer at the time it is tested.

We have already said that every item enters the sensory buffer and that items are selected from there to be entered into the memory buffer. Assume that at the time items enter the sensory buffer they are examined. These items fall into one of two categories. They may be items which are already in the buffer, i.e., their stimulus member may already be in the buffer. Alternatively, their stimulus member may not currently be in the buffer. The former kind of item shall be referred to as an O-item ("old"

item), and the latter kind as an N-item ("new" item).[4] When an O-item is presented for study, it enters the memory buffer with probability one; the corresponding item, which was previously in the buffer, is discarded. Thus an O-item may be said to replace itself in the buffer. When an N-item is presented for study it enters the buffer with probability α. The value of the parameter α may be related in some manner to the particular scheme that a subject is using to rehearse the items currently in the buffer. When an N-item enters (with probability α) some item currently in the buffer is lost. This loss is called the "knockout process" and will be described below. With probability $(1-\alpha)$ an N-item fails to enter the buffer. In this case the buffer remains unchanged, the item in question decays from the sensory buffer, and is permanently lost from memory. For reference, the memory system is diagrammed in Fig. 1.

The memory buffer is arranged as a push-down list. The newest item that enters the buffer is placed in slot r, and the item that has remained in the buffer the longest is in slot 1. If an O-item is presented it enters slot r and the corresponding item is lost (in effect, the stimulus moves from its current slot to slot r and the response is changed). Then the other items move down one slot if necessary, retaining their former order. When an N-item is presented for study and enters the buffer (with probability α) it is placed in the r^{th} slot. The item to be knocked out is chosen according to the following scheme: with probability κ_j the item currently in slot j is the particular item that is discarded, where

[4] The reader should keep in mind that O-items and N-items are theoretical constructs and do not refer to observable experimental events.

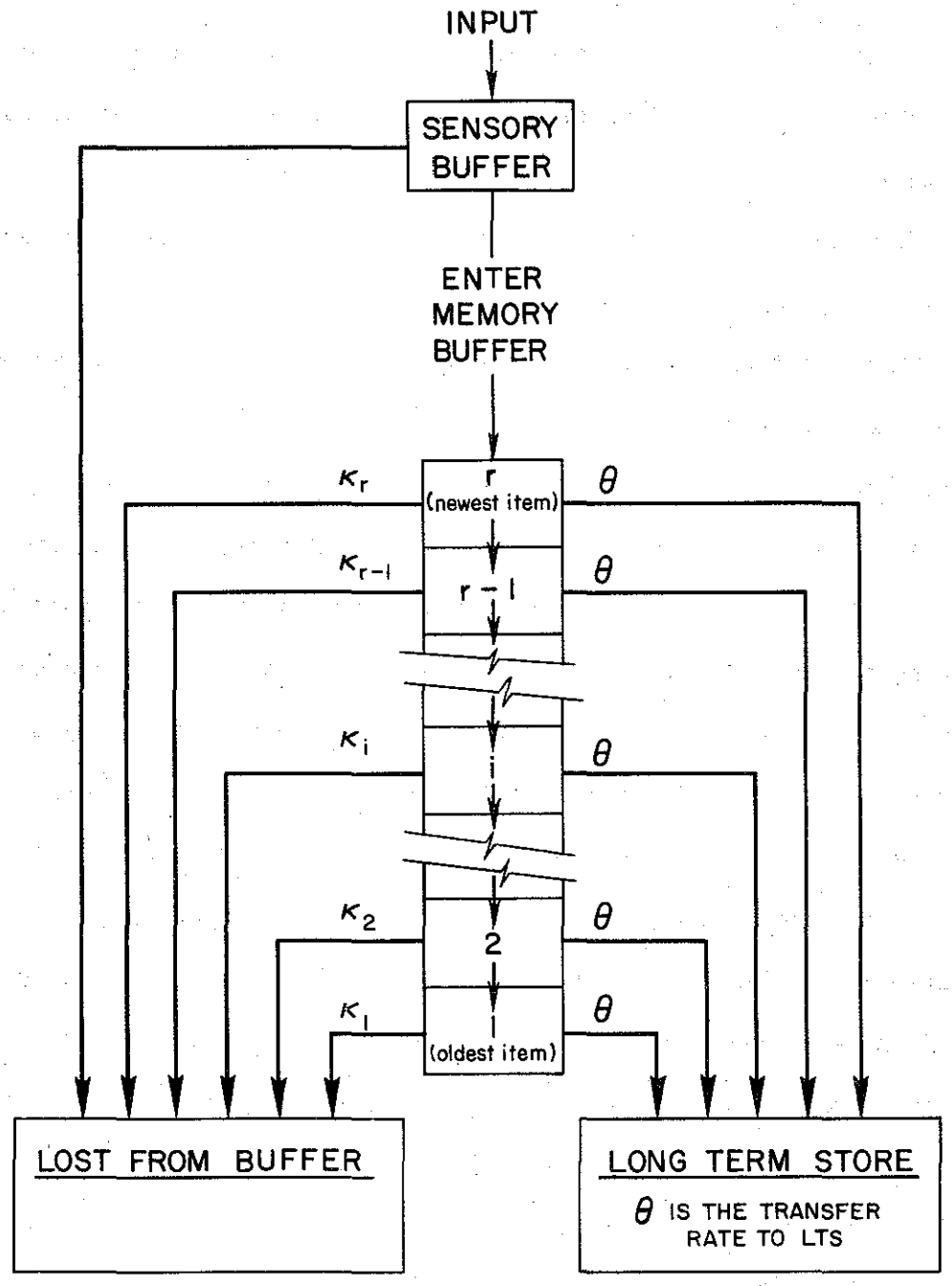

Figure 1. A flow chart characterizing inputs to the memory system.

$\kappa_1 + \kappa_2 + \cdots + \kappa_r = 1$. When the j^{th} item is discarded each item above the j^{th} slot moves down one, and the new item enters the r^{th} slot. Various schemes can be used to develop the κ_j's. The simplest is to let $\kappa_j = \frac{1}{r}$, in which case the item to be knocked out is chosen independently of the buffer position. However, in some experiments it has been necessary to postulate more general schemes which require that the longer the item has been in the buffer the greater its probability of being knocked out (Atkinson and Shiffrin, 1965).

Long-Term Storage

LTS is viewed as a memory state in which information accumulates for each item.[5] It is assumed that information about an item may enter LTS only during the period that an item resides in the buffer. We postulate that the status of an item in the buffer is in no way affected by transfer of information to LTS. Whereas recall from the buffer was assumed to be perfect, recall from LTS is not necessarily perfect and usually will not be. At the time of a test on an item, a subject gives the correct response if the item is in the sensory or memory buffer, but if the item is not in either of these buffers the subject searches LTS. This LTS search is called the retrieval process. Two features of the LTS retrieval process must be specified. First it is assumed that the likelihood of retrieving the correct response for a given item improves as the amount of information stored concerning that item increases. Second, the retrieval of an item gets worse the longer the item has been stored in LTS. This may simply mean that there is

---

[5]The term "information" is not used here in a technical sense. We use the term to refer to codes, mnemonics, images or anything else the subject might store that would be retrievable at the time of test.

some sort of decay in information as a function of the length of time information has been stored in LTS.

We shall specifically assume in this paper that information is transferred to LTS at a constant rate θ during the entire period in which an item resides in the buffer; θ is the transfer rate per trial. Thus, if an item remains in the buffer for exactly j trials (i.e., the j^{th} study item following the presentation of a given item causes it to be knocked out of the buffer), then that item accumulated an amount of information equal to $j\theta$. Next assume that each trial following the trial on which an item is knocked out of the buffer causes the information stored in LTS for that item to decrease by a constant proportion τ. Thus, if an item were knocked out of the buffer at trial j, and i trials intervened between the original study and the test on that item, the amount of information stored in LTS at the time of test would be $j\theta\tau^{i-j}$. We now want to specify the probability of a correct retrieval of an item from LTS. If the amount of information stored at the moment of test for an item is zero, then the probability of a correct retrieval should be at the guessing level. As the amount of information increases, the probability of a correct retrieval should increase toward unity. We define ρ_{ij} as the probability of a correct response from LTS of an item that had a lag of i trials between its study and test, and that resided in the buffer for exactly j trials. Considering the above specifications on the retrieval process,

$$\rho_{ij} = 1 - (1-g)\exp[-j\theta(\tau^{i-j})] \qquad (1)$$

where g is the guessing probability and in the present experiment is 1/26 since there were 26 response alternatives.

Lest the use of an exponential function seem entirely arbitrary, it should be noted that this function bears a close relation to the familiar linear model of learning theory. If we ignore for the moment the decay feature, then $\rho_{ij} = 1 - (1-g)\exp(-j\theta)$. It is easily seen that this is the linear model expression for the probability of a correct response after j reinforcements with parameter $e^{-\theta}$. Thus, the retrieval function ρ_{ij} can be viewed as a linear model with time in the buffer as the independent variable. To be sure, the decay process complicates matters, but the reason for choosing the exponential function becomes somewhat less arbitrary. A decay process is needed so that the probability of a correct retrieval from LTS will approach a chance level as the lag tends toward infinity.

Derivation of Lag Curves[6]

The basic dependent variable in the present experiment is the probability of a correct recall at the time of a test, given lag i. In order to derive this probability we need to know the length of time that an item resides in the memory buffer. Therefore, define

β_j = probability that an item (i.e., a specific stimulus-response pair) resides in the buffer for exactly j trials, given that it is tested at a lag greater than j.

In the general case we must define another quantity in order to find β_j; namely

---

[6]The derivations are for the case where $r < s$. If $r \geq s$ a given item will always remain in the buffer until it is tested and consequently performance will be perfect at all lags.

β'_{ij} = probability that an item (i.e., a specific stimulus-response pair) currently in slot i resides in the buffer for exactly j more trials, given that it is tested at some point following this period.

Remember that r represents the number of slots in the buffer, and κ_j is the probability that the item in the j^{th} slot will be knocked out when an N-item enters. The probability of an N-item (one not currently in the buffer) being presented on a trial is $(s - r)/s$, where s is the number of stimuli used in a given experimental condition; likewise, the probability of an O-item being presented is r/s. We shall define β'_{ij} recursively. Note that an item's buffer position on a trial is either the same, or one less on the succeeding trial (if it is not knocked out of the buffer). We therefore obtain the following difference equations:

$$\beta'_{i,j} = (1 - \frac{s - r}{s - 1}\alpha\kappa_1)\beta'_{1,j-1}$$

$$\beta'_{i,j} = \left\{\frac{r - i}{s - 1} + \frac{s - r}{s - 1}[(1-\alpha) + \alpha(\kappa_{i+1} + \kappa_{i+2} + \cdots + \kappa_r)]\right\}\beta'_{i,j-1}$$

$$+ \left\{\frac{i - 1}{s - 1} + \frac{s - r}{s - 1}\alpha(\kappa_1 + \kappa_2 + \cdots + \kappa_{i-1})\right\}\beta'_{i-1,j-1}$$

$$\beta'_{r,j} = \left\{\frac{s - r}{s - 1}(1-\alpha)\right\}\beta'_{r,j-1} + \left\{\frac{r - 1}{s - 1} + \frac{s - r}{s - 1}\alpha(1 - \kappa_r)\right\}\beta'_{r-1,j-1} \quad . \quad (2)$$

The initial conditions are $\beta'_{i,1} = \frac{s - r}{s - 1}\alpha\kappa_i$. Recall that when an N-item is presented it will enter the memory buffer with probability α. Also, note that the denominator in the terms denoting the probabilities of N-items and O-items is $(s - 1)$ rather than s. This is the case because β'_{ij} is a probability conditionalized upon the fact that we have yet to present the item in question for test. Now we can write:

$$\beta_j = \begin{cases} \frac{s - r}{s}(1-\alpha) & , \text{ for } j = 0 \\ \left\{1 - \frac{s - r}{s}(1-\alpha)\right\} \beta'_{r,j} & , \text{ for } j > 0 \ , \end{cases} \tag{3}$$

where β_0 is the probability that the item in question does not enter the memory buffer in the first place. It should be clear that the above difference equations can be solved by successive substitution, but such a process is lengthy and cumbersome. In practice, numerical solutions are easily obtained using a high-speed computer.

The probability of a correct response to an item tested at lag i can now be written in terms of the β_j's. Let "C_i" represent the occurrence of a correct response to an item tested at lag i. Then

$$\Pr(C_i) = \left[1 - \sum_{k=0}^{i} \beta_k\right] + \left[\sum_{k=0}^{i} \beta_k \rho_{ik}\right] . \tag{4}$$

The first bracketed term is the probability that the item is in the buffer at the time of test. The second bracket contains a sum of probabilities, each term representing the probability of a correct retrieval from LTS of an item which remained in the buffer for exactly k trials and was then lost.

Experiment I

The first experiment was carried out to determine whether reasonable predictions could be made assuming that the parameters of the model (r, α, θ, and τ) are independent of the number of stimuli the subject is trying to remember. Three experimental conditions were run: $s = 4$, 6, and 8.

Method

Subjects. The subjects were 9 students from Stanford University who received $2 per experimental session. Each subject participated in approximately 10 sessions.

Apparatus. The experiment was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by computer programs running in a modified PDP-1 computer manufactured by the Digital Equipment Corporation, and under control of a time-sharing system. The subject was seated at a cathode-ray-tube display terminal; there were six terminals each located in a separate 7 × 8 ft. sound-shielded room. Stimuli were displayed on the face of the cathode ray tube (CRT); responses were made on an electric typewriter keyboard located immediately below the lower edge of the CRT.

Stimuli and responses. The stimuli were two-digit numbers randomly selected for each subject and session from the set of all two-digit numbers between 00 and 99. Once a set of stimuli was selected for a given session, it was used throughout the session. Responses were letters of the alphabet, thus fixing the guessing probability of a correct response at 1/26.

Procedure. For each session the subject was assigned to one of the three experimental conditions (i.e., s was set at either 4, 6, or 8). An attempt was made to assign subjects to each condition once in consecutive three-session blocks. Every session began with a series of study trials: one study trial for each stimulus to be used in the session. On a study trial the word "study" appeared on the upper face of the CRT. Beneath the word "study" one of the stimuli appeared along with a randomly-slected letter from the alphabet. Subjects were instructed to try to remember the

association between the stimulus-response pairs. Each of these initial study trials lasted for 3 sec. with a 3-sec. intertrial interval. As soon as there had been an initial study trial for each stimulus to be used in the session, the session proper began.

Each subsequent trial involved a fixed series of events. (1) The word test appeared on the upper face of the CRT. Beneath the word test a randomly selected member of the stimulus set appeared. Subjects were instructed that when the word test and a stimulus appeared on the CRT, they were to respond with the last response that had been associated with that stimulus, guessing if necessary. This test portion of a trial lasted for 3 sec. (2) The CRT was blacked out for 2 sec. (3) The word study appeared on the upper face of the CRT for 3 sec. Below the word study a stimulus-response pair appeared. The stimulus was the same one used in the preceding test portion of the trial. The response was randomly selected from the letters of the alphabet, with the stipulation that it be different from the immediately preceding response assigned to that stimulus. (4) There was a 3-sec. intertrial interval before the next trial. Thus a complete trial (test plus study) took 11 sec. A subject was run for 220 such trials during each experimental session.

Results

In order to examine the data for habituation or learning-to-learn effects, the overall probability of a correct response for each stimulus condition ($s = 4$, 6, and 8) was plotted in consecutive 25-trial blocks. It was found that after a brief rise at the start of each daily session, the curves appeared to level off at three distinct values. Due to this brief initial warm-up effect, subsequent analyses will not include data from

the first 25 trials of each session. Furthermore, the first session for each subject will not be used.

Figure 2 presents the probability of a correct response as a function of lag for each of the three stimulus set sizes examined. It can be seen that the smaller the stimulus set size, the better the overall performance. It is important to note that the theory presented in the earlier part of this paper predicts such a difference on the following basis: the larger the size of the stimulus set, the more often an N-item will be presented; and the more often N-items are presented, the more often items in the buffer will be knocked out. Recall that only N-items can knock items from the buffer; O-items merely replace themselves.

It can be seen that performance is almost perfect for lag O in all three conditions. This might be expected because lag O means that the item was tested immediately following its study. The curves drop sharply at first and slowly thereafter, but have not yet reached the chance level at lag 17, the largest lag plotted. The chance level should be 1/26 since there were 26 response alternatives.

It is of interest to examine the type of errors occurring at various lags in the three experimental conditions. There are two categories of errors that are of special interest to us. The first category is composed of errors which occur when the immediately preceding correct response to a stimulus is given, instead of the present correct response. The proportions of errors of this type were calculated for each lag and each condition. The proportions were found to be quite stable over lags with mean values of .065, .068, and .073 for the 4, 6, and 8 stimulus conditions, respectively. If the previously correct response to an item is randomly generated on any

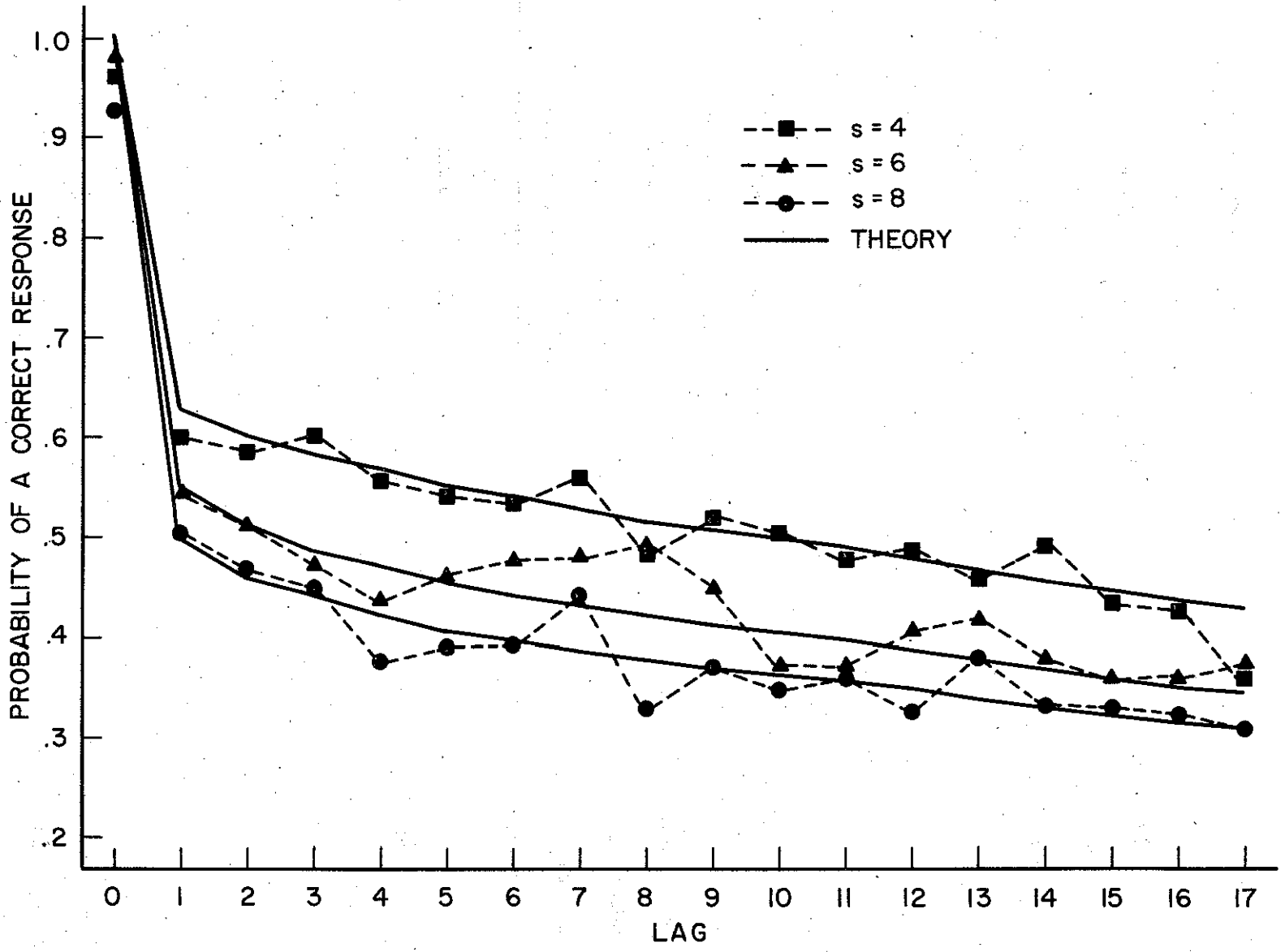

Figure 2. Observed and theoretical probabilities of a correct response as a function of lag (Experiment I).

given error, these values should not differ significantly from 1/25 = .04. The mean proportion for this type of error was computed for each subject and each condition. In both the $s = 4$ and $s = 6$ conditions 7 of the 9 subjects had mean values above chance; in the $s = 8$ condition 8 of the 9 subjects were above chance. A second category of errors of interest to us is composed of those responses that are members of the current set of responses being remembered, but are not the correct response. The proportions of this type of error were calculated for each lag in each of the three experimental conditions. Again, the proportions were found to be quite stable over lags. The mean values were .23, .28, and .35 for the 4, 6, and 8 stimulus conditions, respectively; on the basis of chance these values would have to be bounded below .12, .20, and .28, respectively. No statistical tests were run, but again the values appear to be above those expected by chance. While a detailed examination of the implications of these conditional error results is not a purpose of this paper, it should be pointed out that this type of analysis may yield pertinent information regarding the nature of the LTS retrieval process.

There are two other lag curves that prove interesting. We shall call these the "all-same" and the "all-different" curves. In the all-same conditions, we compute the probability of a correct response as a function of the lag, when all of the intervening items between study and test involve the same stimulus. The model predicts that once the intervening stimulus enters the buffer, there will be no further chance of any other item being knocked out (i.e., once the intervening item enters the buffer, each succeeding presentation is an 0-item). Hence, these curves should drop at a slower rate than the unconditional lag curves presented in Fig. 2. The all-same

curves are plotted in Fig. 3. The points for lag 0 and lag 1 are, of course, the same as in the unconditional lag plots of Fig. 2. It can be seen that the curves indeed drop at a slower rate in this condition.

The all-different condition refers to the probability of a correct response as a function of lag, when the intervening items between study and test all involve different stimuli. For this reason the maximum lag which can be examined is one less than the size of the stimulus set. It should be clear that the all-different condition maximizes the expected number of intervening N-items at a given lag. This lag curve should therefore have a faster drop than the unconditional lag curves presented in Fig. 2. The data are shown in Fig. 4. While it is difficult to make a decision by inspection in this condition because the data are quite unstable, it does seem that the curves drop faster than the corresponding ones in Fig. 2. Note that here, also, the points for lag 0 and lag 1 are of necessity the same as in the previous conditions.

The results that have been presented to this point have been group data. It is of interest to see whether individual subjects perform in a fashion similar to the group curves. Table 1 presents the lag curves for the three experimental conditions for individual subjects. The lag curves have been collapsed into three-lag blocks to minimize variability. An examination of these individual curves indicates that all subjects, except for subject 8, appear to be performing in a manner very similar to the group data.

A final remark should be made regarding the number of observations taken at each point on these lag curves. Because of the random procedure used to select the stimuli from trial to trial, the number of observations going into successive points on the lag curves decrease geometrically. For the group

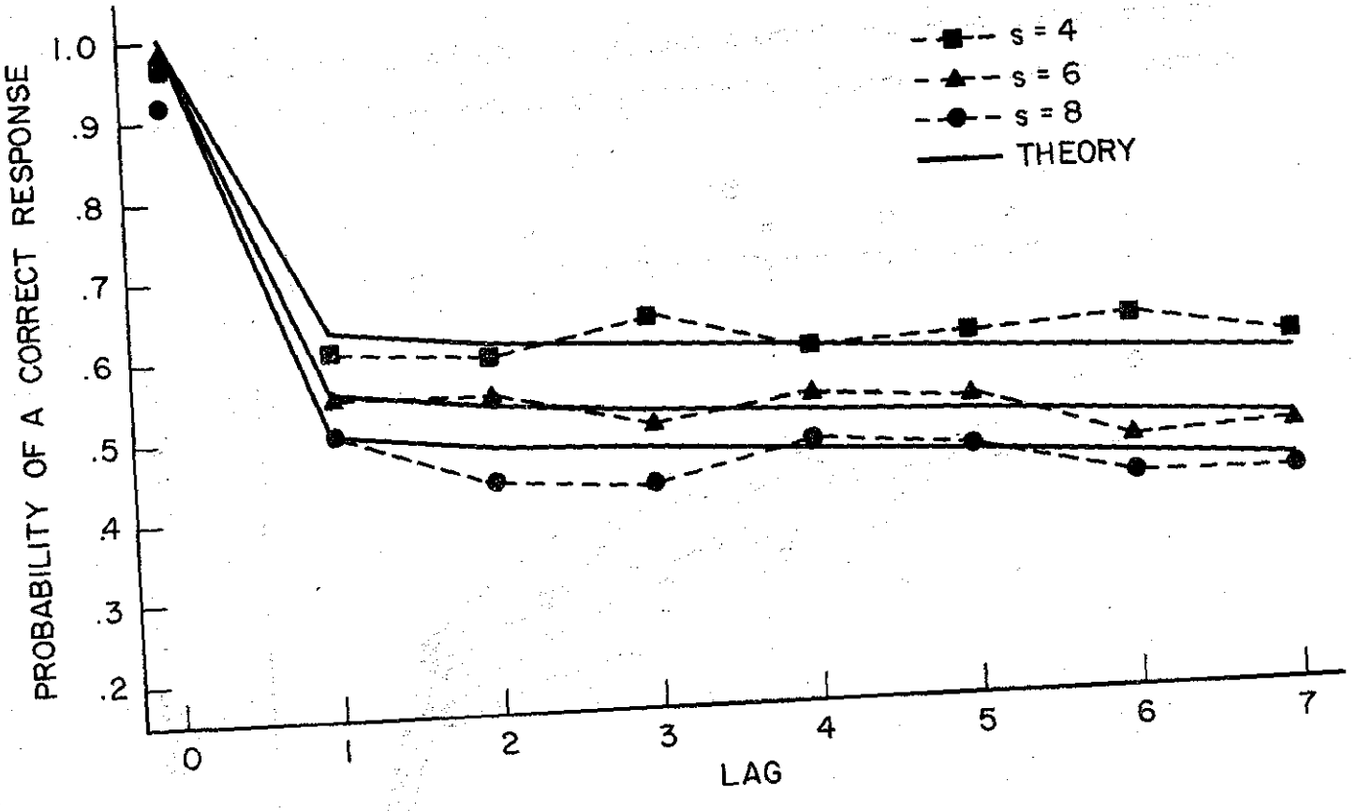

Figure 3. Observed and theoretical probabilities of a correct response as a function of lag for the "all-same" condition (Experiment I).

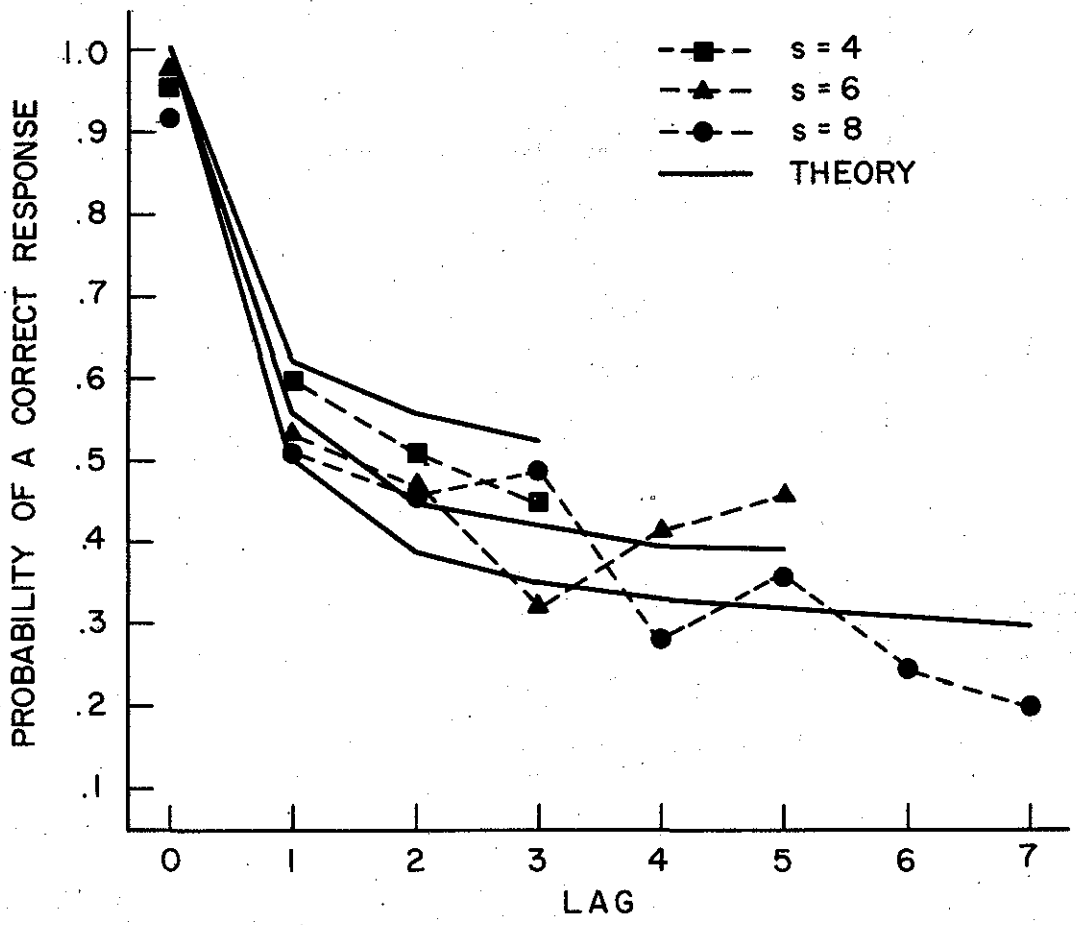

Figure 4. Observed and theoretical probabilities of a correct response as a function of lag for the "all-different" condition (Experiment I).

TABLE I

Observed and predicted probabilities of a correct response as a function of lag for individual subjects. The predicted values are in parentheses and are based on the parameter estimates that give the best fit for that subject; these estimates are presented in the bottom section of the table. The χ^2_I and χ^2_G are computed for each subject using the individually estimated parameters and the group parameters, respectively. Entries in the top section of the table should be read with a leading decimal point (Experiment I).

Experimental Condition	Lag	Subjects								
		1	2	3	4	5	6	7	8	9
s = 4	0-1-2	72 (70)	80 (85)	86 (86)	66 (67)	73 (75)	86 (84)	75 (73)	60 (60)	89 (84)
	3-4-5	51 (50)	74 (72)	76 (78)	52 (56)	62 (64)	65 (73)	50 (50)	61 (58)	70 (69)
	6-7-8	49 (48)	72 (69)	74 (76)	61 (52)	52 (58)	71 (68)	42 (44)	59 (56)	64 (63)
	9-10-11	50 (47)	62 (68)	72 (74)	59 (50)	49 (54)	67 (67)	41 (42)	60 (54)	66 (62)
	12-13-14	51 (46)	64 (66)	70 (73)	50 (49)	44 (49)	55 (66)	41 (40)	58 (52)	61 (61)
s = 6	0-1-2	63 (62)	82 (77)	78 (75)	56 (55)	56 (53)	72 (71)	61 (58)	49 (46)	68 (67)
	3-4-5	43 (42)	74 (66)	71 (69)	56 (50)	52 (46)	68 (63)	53 (49)	48 (42)	59 (58)
	6-7-8	41 (41)	72 (63)	65 (65)	45 (43)	41 (39)	63 (60)	35 (37)	40 (41)	53 (53)
	9-10-11	48 (40)	62 (61)	66 (62)	53 (42)	43 (37)	58 (59)	37 (35)	40 (40)	55 (52)
	12-13-14	51 (39)	64 (60)	60 (60)	54 (41)	45 (36)	60 (58)	38 (34)	44 (39)	54 (51)
s = 8	0-1-2	59 (57)	69 (73)	73 (70)	47 (49)	43 (45)	63 (63)	50 (51)	35 (38)	58 (57)
	3-4-5	40 (40)	67 (62)	63 (62)	46 (43)	47 (38)	60 (54)	45 (41)	56 (35)	52 (48)
	6-7-8	35 (36)	53 (60)	61 (60)	31 (39)	39 (31)	54 (52)	30 (32)	52 (32)	44 (45)
	9-10-11	30 (30)	56 (59)	58 (58)	41 (37)	45 (29)	53 (51)	31 (30)	44 (31)	45 (42)
	12-13-14	29 (29)	57 (58)	51 (57)	38 (36)	36 (28)	51 (50)	29 (29)	37 (30)	49 (41)
Parameter Estimates										
	r	2	3	3	2	2	2	2	2	2
	α	.56	.66	.59	.50	.50	.32	.65	.38	.31
	θ	.70	1.22	2.09	.39	1.12	1.84	.62	1.37	.82
	τ	.84	.95	.93	.89	.80	.82	.78	.99	.94
	χ^2_I	23.6	29.2	24.8	31.2	38.2	26.4	31.2	67.2	13.6
	χ^2_G	29.2	48.3	29.4	33.6	47.6	32.6	34.7	89.3	15.2

data there are over 1000 observations at lag 0 and slightly more than 100 at lag 17 for each of the three experimental conditions. Of course, the exact form of the distribution of data points varies as a function of the experimental condition, with more short lags occurring in the $s = 4$ condition and more long lags occurring in the $s = 8$ condition.

Model Predictions

In order to estimate parameters and evaluate the goodness-of-fit of the theory to the data, we define the following χ^2 function:

$$\chi^2 = \sum_i \left\{ \frac{1}{N_i Pr(C_i)} + \frac{1}{N_i - N_i Pr(C_i)} \right\} \left\{ N_i Pr(C_i) - O_i \right\}^2 \qquad (5)$$

where the sum is taken over all data points i which are being evaluated. The observed number of correct responses for the i^{th} point is denoted by O_i; N_i is the total number of responses for the i^{th} point; and $Pr(C_i)$ is the theoretical probability of a correct response which depends on r, α, θ, and τ. Thus $N_i Pr(C_i)$, the predicted number of correct responses for the i^{th} point, should be close to O_i if the theory is accurate.

We first analyze the lag curves displayed in Fig. 2. The set of parameter values r, α, θ, and τ that minimizes the above χ^2 function over the $3 \times 17 = 51$ data points in Fig. 2 will be taken to be the best fit of the model.[7] In order to minimize χ^2 we resorted to a numerical routine using a computer. The routine involved selecting tentative values for r, α, θ, and τ, computing the $Pr(C_i)$'s and the related χ^2, repeating the procedure with another set of parameter values, and continuing thusly until

[7] The lag 0 point in this and subsequent analyses is not included in the χ^2 since its predicted probability value is one.

the space of possible parameter values has been systematically explored. The parameter values yielding the smallest χ^2 are then used as the estimates. When enough points in the parameter space are scanned, the method yields a close approximation to the true minimum.[8]

The predictions for $Pr(C_i)$ could be derived using Eqs. 3, but it was decided to set the $\kappa_i = 1/r$, in which case the equations simplify greatly. In a study by Phillips, Shiffrin and Atkinson (1966) it was found that the assumption $\kappa_i = 1/r$ was not tenable; in that experiment, however, there were strong reasons for expecting that the subject would tend to eliminate the oldest items from the buffer first. In the current experiment there is a continuous display of items and there seemed to be no compelling reason to believe that the subject would not discard items from the buffer in a random fashion. For this reason κ_i was set equal to $1/r$ for every buffer position. Under this assumption it is immaterial what position an item occupies in the buffer. Thus $\beta'_{i,k} = \beta'_{j,k}$ for all i and j; hence (as can be easily verified) every line of Eq. 3 can be rewritten as follows:

$$\beta'_{i,k} = \left\{1 - \frac{s - r}{s - 1}\,\alpha\,\frac{1}{r}\right\}\beta'_{i,k-1} \quad . \tag{6}$$

Let the term in brackets be denoted by $1 - X$. Then we have $\beta_0 = (1-\alpha)(s-r)/s$ which is the probability that the item will not enter the buffer, and

$$\beta_k = (1-\beta_0)X(1-X)^{k-1} \quad . \tag{7}$$

---

[8] For a discussion of the minimum χ^2 method see Holland (1965) or Atkinson, Bower, and Crothers (1965).

It is easy to verify this equation if we note that X is the probability that an intervening item will enter the buffer and knock out the item of interest. For the item of interest to be knocked out of the buffer by exactly the k^{th} following item, it is necessary that the following conditions hold: (1) the item must enter the buffer in the first place; (2) the next $k - 1$ intervening items must not knock it out; (3) the k^{th} item must knock out the item of interest. These considerations lead directly to Eq. 7.

Given β_k we can calculate the predicted lag curves for each set of parameters considered using Eq. 4. The χ^2 procedure described earlier was applied simultaneously to all three curves displayed in Fig. 2 and the values of the parameters that gave the minimum χ^2 were as follows: $r = 2$, $\alpha = .39$, $\theta = .40$, and $\tau = .93$. The theoretical lag curves generated by these parameters are shown in Fig. 2. It can be seen that the observed data and the predictions from the model are in close agreement; the minimum χ^2 value is 43.67 based on 47 degrees of freedom ($17 \times 3 = 51$ data points minus four estimated parameters).[9] It should be emphasized that the three curves are fit simultaneously using the same parameter values, and the differences between the curves depend only on the value of s

---

[9] In this and all subsequent minimizations reported in this paper, r was permitted to take on only integer values. Better fits can be obtained by removing this constraint (e.g., in this case the minimum χ^2 is 40.36 when $r = 2.1$, $\alpha = .37$, $\theta = .44$, and $\tau = .91$), but we prefer to evaluate the model assuming r is fixed for all subjects.

used. The predicted probabilities of a correct response weighted and summed over all lag positions are .562, .469, and .426 for s equal to 4, 6, and 8, respectively; the observed values are .548, .472, and .421.

The estimated value of α indicates that only 39 percent of the N-items presented actually enter the buffer (remember that O-items always enter the buffer). At first glance this percentage may seem low, but a good deal of mental effort may be involved in keeping an item in the buffer via rehearsal, and the subject might be reluctant to discard an item which he has been rehearsing before it is tested. Actually, if there were no long-term storage, the subject's overall probability of a correct response would be independent of α. Thus it might be expected that α would be higher the greater the effectiveness of long-term storage in an experiment. The estimate of θ found does not have a readily discernable interpretation, but the value of $\tau = .93$ indicates that the decay in LTS is extremely slow. It is not necessary to assume that any actual decay occurs--several alternative processes are possible. For example, the subject could search LTS backwards along a temporal dimension, sometimes stopping the search before the information relevant to the tested item is found.[10]

Next we examine the lag curves for the all-same condition. As indicated earlier these curves should be less steep than the unconditional lag curves. This would be expected because, in the all-same conditions (where the intervening trials all involve the same stimulus), once an intervening item enters the buffer, every succeeding item will be an O-item and will

---

[10]The high value of τ might suggest that a reasonable fit could be obtained setting $\tau = 1$. When this was done, however, the minimum χ^2 was 62.74 with parameter estimates $r = 2$, $\alpha = .42$, $\theta = .24$.

replace itself. Indeed, if $\alpha = 1$ and there is no LTS storage, the all-same lag curves would be level from lag 1 onward. The model applies directly to this case. Define β_j^* as the probability that an item resides in the buffer for exactly j trials and is then knocked out, given that all the intervening trials involve the same stimulus. Then

$$\beta_j^* = \begin{cases} \frac{s - r}{s}(1-\alpha) & , \text{ for } j = 0 \\ (1-\beta_0^*)\left[\frac{s - r}{s \ominus 1}(1-\alpha)^{j-1}\frac{\alpha}{r}\right] & , \text{ for } j > 0 . \end{cases} \tag{8}$$

It can easily be seen that the β_j^* have the above form. For an item to be knocked out by the j^{th} succeeding item it is necessary that the following holds: (1) the item enters the buffer initially; (2) the following items must be new items and must not enter the buffer for $j - 1$ trials (clearly, if the first intermediate item is an N-item, then in the all-same condition each succeeding item has to be an N-item until one of the items enters the buffer); (3) the j^{th} following item enters the buffer and knocks out the item of interest. The predicted lag curves for the all-same condition may be calculated substituting β_j^* for β_j in Eq. 4. The parameters found in fitting the unconditional lag curves in Fig. 2 were used to generate predictions for the all-same condition, and the predicted lag curves are presented in Fig. 3. The fit is excellent as indicated by a χ^2 of 26.8 based on 21 degrees of freedom.

Next we turn to the lag curves for the all-different condition. Considerations similar to those presented in the discussion of the all-same data lead to the prediction that the all-different lag curves will be steeper than the unconditional lag curves. Unfortunately there were

relatively few observations in this condition and the data is fairly unstable. Nevertheless we shall apply the model to these data in large part because the mathematical techniques involved are rather interesting. Define

β_j^{**} = the probability that an item will reside in the buffer for exactly j trials, given that the intervening stimuli are all different.

It can be quickly demonstrated that an attempt to develop the β_j^{**} equations directly does not succeed, primarily because the probability of presenting an N-item changes from trial to trial. The solution is to view the process an an inhomogeneous Markov chain with $r + 1$ states. The first state will correspond to the event that the item of interest is currently not in the buffer. The other r states will denote the conditions in which the item of interest is in the buffer and m ($m = 0$ to $r - 1$) of the remaining places in the buffer are filled with items that have already been presented in the sequence of all-different items. For the sake of simplicity we shall develop the process for the case where $r = 2$ since the all-different curves will be fit using the parameters estimated from the unconditional lag curves. It is easy to see how to generalize the method to larger values of r.

To start with, define $\overline{B}$ as the state in which the item of interest is not currently in the buffer. Define BA as the state where the item of interest is in the buffer and the other slot of the buffer is occupied by an item which has already been presented in the sequence of all-different items. Define $B\overline{A}$ as the state in which the item of interest is in the buffer and the other slot of the buffer is not occupied by an item which

has already been presented in the sequence of all-different intervening items. Then the following matrix describes transitions from intervening trial k to intervening trial $k+1$:

$$\begin{array}{c} \\ \text{Trial } k \\ \\ \\ \end{array} \begin{array}{c} \\ \overline{B} \\ BA \\ B\overline{A} \end{array} \begin{array}{c} \text{Trial } k+1 \\ \begin{array}{ccc} \overline{B} & BA & B\overline{A} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ \frac{\alpha}{2} & 1-\frac{\alpha}{2} & 0 \\ \frac{\alpha}{2}\left[\frac{s-(k+2)}{s-(k+1)}\right] & 1-\left[1-\frac{\alpha}{2}\right]\left[\frac{s-(k+2)}{s-(k+1)}\right] & [1-\alpha]\left[\frac{s-(k+2)}{s-(k+1)}\right] \end{bmatrix} \end{array} . \tag{9}$$

The starting vector at $k = 0$ is as follows:

$$\begin{array}{ccc} \overline{B} & BA & B\overline{A} \\ \left[\frac{s-2}{s}(1-\alpha) \right. & 0 & \left. 1-\frac{s-2}{s}(1-\alpha)\right] . \end{array}$$

Let the probability of being in state $\overline{B}$ on intervening trial k be $p_k(\overline{B})$. Then

$$\beta_j^{**} = \begin{cases} p_j(\overline{B}) - p_{j-1}(\overline{B}) & \text{, for } j > 0 \\ p_0(\overline{B}) & \text{, for } j = 0 \end{cases} \tag{10}$$

where $p_0(\overline{B}) = (1-\alpha)(s-2)/s$.

In order to determine $p_k(\overline{B})$ we used a computer to multiply the starting vector by the transition matrix the appropriate number of times. This was done using the parameter values from the fit of the unconditional lag curves. The β_j^{**} were then computed and the lag curves generated as before. The predicted curves are shown in Fig. 4. Considering the lack of stability in the data, the fit is not too bad. The χ^2 was 64.8 based on 15 degrees of freedom.

The model is not explicit regarding the likelihood of the previously correct response being incorrectly emitted at the time of test. Nevertheless, the interpretation of the LTS retrieval process which postulates a temporal search of stored items suggests that the previously correct information may be accidentally found during retrieval, thus heightening the probability that the prior response will be given. A slight anomaly here is that in the data this probability appears to be independent of lag which might not be predicted from the preceding argument.

Similarly, the model does not make predictions concerning the probability that a response in the current response set will be given as an error. However, there will be overlap between the current response set and the items stored in the buffer; it does not seem unreasonable that subjects who cannot find the correct response in their search of the buffer and LTS might tend to guess by favoring a response currently in the buffer. The data indicate that this tendency is above the chance level. This suggests that our assumption of a guessing level of 1/26 could be slightly inaccurate. In future work it may prove necessary to postulate a changing guessing level which declines toward the reciprocal of the number of responses only as the lag tends toward infinity.[11]

We now consider the implicit assumption involved in fitting curves for group data--namely that the subjects are homogeneous. A direct approach would be to fit the model to each subject's data separately. This was done

---

[11]Our use of the term "guessing level" in this context is itself misleading because it seems clear that the subject is using stored information concerning recent responses while "guessing."

under the restriction that three adjacent lags be lumped into a single point (there were not enough observations to guarantee stable lag curves from individual subjects without lumping adjacent points). Thus the model was fit independently to the data from each subject in the same manner that the group data was fit (naturally, for each set of parameter values considered, the predicted lag curves were lumped in the same manner as the observed data). The predictions of the model yielding minimum χ^2's for each subject are presented in Table 1 along with the observed data. Also given are the minimum χ^2 values and the parameter estimates for each subject. It is somewhat difficult at this point to decide the question of homogeneity of the subjects. In order to do so, the lag curves for each subject were predicted using a single set of parameters; namely those values estimated from the group data. When this was done the sum of the χ^2 values over subjects was 359.9 with 131 degrees of freedom. The sum of the χ^2 when each subject was fit with a separate best set of parameters was 285.4 with 99 degrees of freedom. The ratio of the two χ^2's, each divided by its respective degrees of freedom, is 1.05. This suggests that the assumption of homogeneity of subjects is not unreasonable.

Experiment II

Experiment II was identical to Experiment I in all respects except the following. In Experiment I the set of s stimuli was the same throughout an experimental session, with only the associated responses being changed on each trial, while in Experiment II all 100 stimuli were available for use in each session. In fact, every stimulus was effectively an N-item since the stimulus for each study trial was selected randomly from the set of all 100 stimuli under the restriction that no stimulus could be

used if it had been tested or studied in the previous fifty trials. There were still three experimental conditions with s equal to 4, 6, or 8 denoting the number of items that the subject was required to try to remember at any point in time. Thus a session began with either 4, 6, or 8 study trials on different randomly selected stimuli each of which was paired with a randomly selected response (from the 26 letters). On each trial a stimulus in the current to-be-remembered set was presented for test. After the subject made his response he was instructed to forget the item he had just been tested on, since he would not be tested on it again. Following the test a new stimulus was selected (one that had not appeared for at least fifty trials) and randomly paired with a response for the subject to study. This procedure is quite different from Experiment I where the study stimulus was always the one just tested.

Denote an item presented for study on a trial as an O-item (old item) if the item just tested was at the moment of test in the buffer. Denote an item presented for study as an N-item (new item) if the item just tested was not in the buffer. This terminology conforms precisely to that used to describe Experiment I. If an O-item is presented there will be at least one spot in the buffer occupied by a useless item (the one just tested). If an N-item is presented, the buffer will be filled with information of the same value as that before the test. If we assume that an N-item has probability α of entering the buffer, and that an O-item will always enter the buffer and knock out the item just made useless, then the theory used to analyze Experiment I will apply here with no change whatsoever. In this case we again expect that the lag curves for $s = 4$, 6, and 8 would be separated. In fact, given the same parameter values, exactly the same

predicted curves would be expected in Experiment II as in Experiment I.

We may have some doubt, however, that the assumptions regarding N-items and O-items will still hold for Experiment II. In Experiment I the stimulus just tested was re-paired with a new response, virtually forcing the subject to replace the old response with a new one if the item was in the buffer. To put this another way, if an item is in the buffer when tested, only a minor change need be made in the buffer to enter the succeeding study item: a single response is replaced by another. In Experiment II, however, a greater change needs to be made in order to enter an O-item; both a stimulus and a response member have to be replaced. Thus an alternative hypothesis which could be entertained holds that every entering item (whether an N-item or an O-item) has the same probability α of entering the buffer, and will knock out any item currently in the buffer with equal likelihood. In this case there will be no predicted differences among the lag curves for the $s = 4$, 6, and 8 conditions.

Results

The observed lag curves for Experiment II are displayed in Fig. 5. The number of observations at each point range from 1069 for lag 0 in condition $s = 4$ to 145 for lag 17 in condition $s = 8$. It should be emphasized that except for the procedural changes described above and the fact that a new sample of subjects was used in Experiment II, the experimental conditions and operations were identical in the two experiments. The important point of interest in this data is that lag curves for the three conditions appear

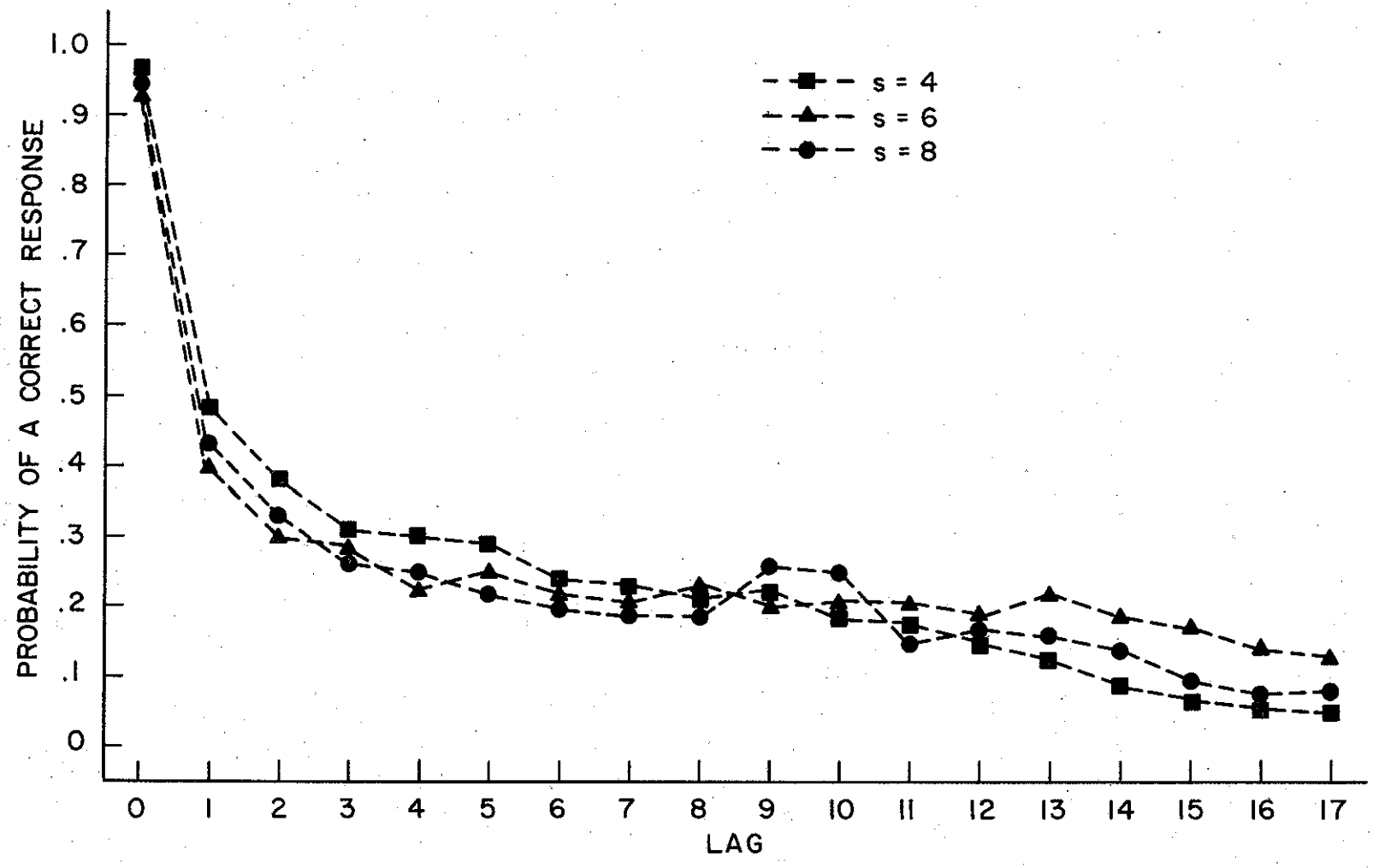

Figure 5. Observed and theoretical probabilities of a correct response as a function of lag. (Experiment II).

to overlap each other.[12] For this reason we lump the three curves to form the single lag curve displayed in Fig. 6.

Theoretical Analysis

Because the lag curves for the three conditions are not separated we assume that every item has an independent probability, α, of entering the buffer. If an item does enter, it randomly knocks out any one of the items already there. Under these assumptions we define

β°_{j} = probability that an item will be knocked out of the buffer by exactly the j^{th} succeeding item.

For this event to happen the following must hold: (1) the item must enter the buffer initially; (2) the item must not be knocked out for $j - 1$ trials; (3) the item must be knocked out by the j^{th} following item. Therefore

$$\beta^{\circ}_{j} = \begin{cases} 1 - \alpha & , \text{ for } j = 0 \\ (1 - \beta^{\circ}_{0})(1 - \frac{\alpha}{r})^{j-1} \frac{\alpha}{r} & , \text{ for } j > 0 \end{cases} \qquad (11)$$

where α/r is the probability that an intervening item will knock out the item of interest.

The curve in Fig. 6 was then fit using the minimum χ^2 technique. The parameter estimates were $r = 2$, $\alpha = .52$, $\theta = .17$, and $\tau = .90$; the minimum χ^2 value was 14.62 based on 13 degrees of freedom. It can be seen that the fit is excellent. Except for r, the parameters differ somewhat from those found in Experiment I. This result is not too

[12] To determine whether the three curves in Fig. 5 differ reliably, the proportions correct for each subject and condition calculated and then ranked. An analysis of variance for correlated means did not yield significant effects ($F = 2.67$, $df = 2/16$, $p > .05$).

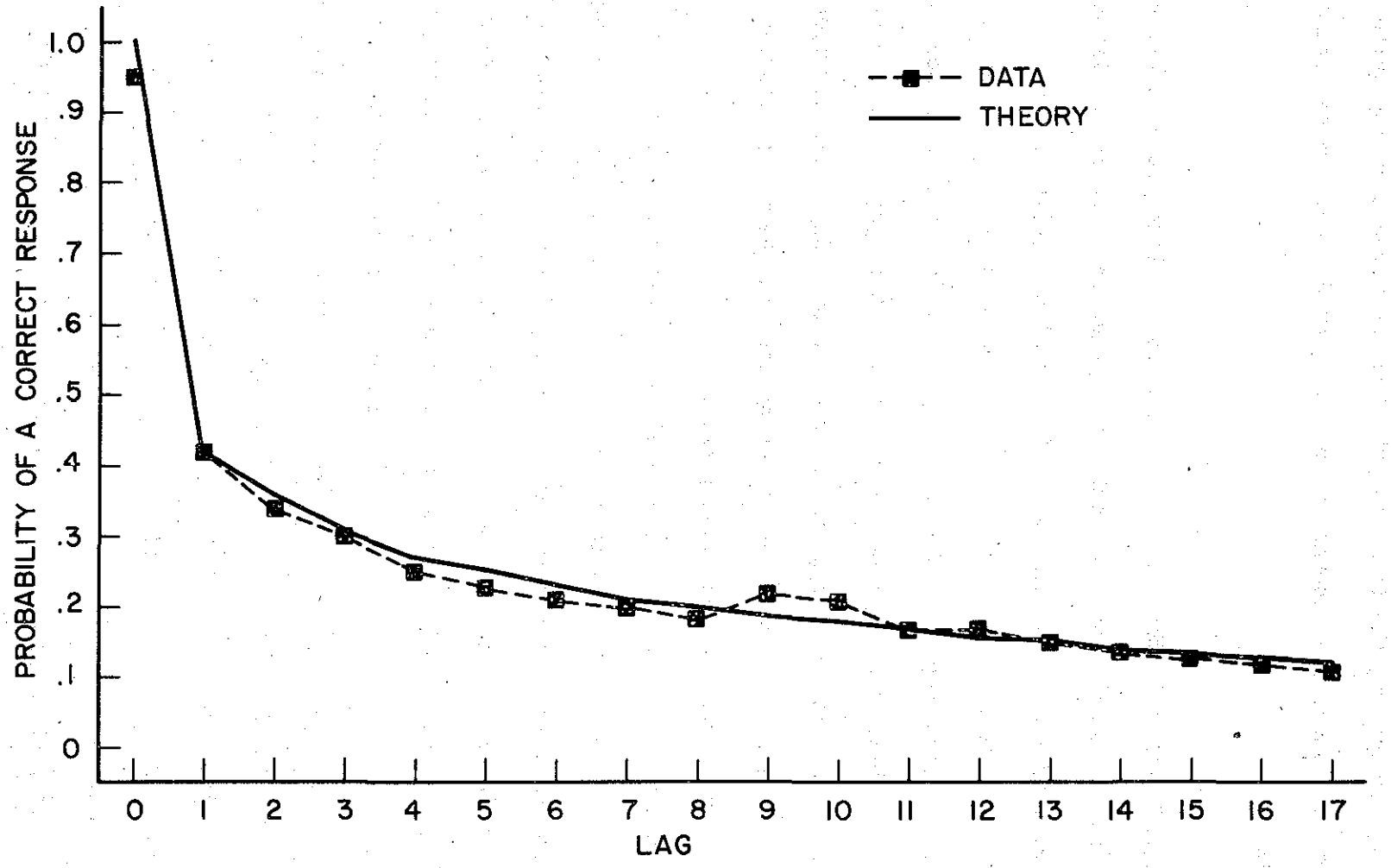

Figure 6. Observed and theoretical probabilities of a correct response as a function of lag. Data from the $s = 4, 6,$ and 8 conditions have been pooled to obtain the observed curve (Experiment II).

surprising considering the fact that the two experiments employ quite different procedures even though on logical grounds they can be regarded as equivalent.

Discussion

The difference in the effects of stimulus set size found in Experiments I and II suggests that the subject engages in an active decision process as each item is presented. This decision involves whether or not to enter the item into the memory buffer. The subject may also engage in a related decision regarding whether or not to transfer information on a given item to LTS. The experiments reported in this paper do not bear on this second point, but this type of decision undoubtedly would be important in studies of learning where each entering item may have been studied before as in the typical paired-associate paradigm.

An extended discussion of the relation of this model to other theories of memory may be found in Atkinson and Shiffrin (1965). The following points, however, are worth brief mention here. The model contains both all-or-none and incremental components: retrieval from the buffer is all-or-none and the buildup and decline of information in LTS in incremental. It is possible, however, to view LTS in a more discrete fashion than was done in this paper. For example, the transfer process might involve making partial copies of items in the buffer and then placing them in LTS. The number of copies made, of course, could depend on the length of time the item resided in the buffer. With one such copy the subject may be able to make a correct recognition response, whereas multiple copies would be needed for a correct recall response. Retroactive interference effects are also represented in the model. A sharp retroactive interference effect occurs

in the buffer caused by the knockout process; a weaker effect occurs in LTS which is represented by the decay process. While proactive interference effects are not explicitly handled in the present paper, the general statement of the model includes a representation of them (Phillips, Shiffrin, and Atkinson, 1966). In the present study it is assumed that interference caused by preceding items in the sequence averages out at each lag. Finally, we note that other writers, in particular Broadbent (1963), Bower (1966), and Estes (1966), have presented theoretical models which mesh nicely with the conceptualization presented here.

Appendix

Throughout this paper it has been assumed that information is transferred to LTS at a constant rate, θ, during the entire period that an item resides in the buffer. Thus, if an item remains in the buffer for j trials, $j\theta$ is the amount of information transferred to LTS. Although this process seems reasonable to us, alternative schemes can be proposed. In particular, it can be assumed that an amount of information equal to θ is transferred to LTS at the time an item enters the memory buffer, and that this ends the transfer process for that item independent of any further time that it stays in the buffer. Thus any item that enters the memory buffer would have the same amount of information transferred to LTS. Two versions of this new model now come to mind: the information in LTS may start decaying at once, or the information may not start decaying until the item is knocked out of the memory buffer. These two versions are represented by the following retrieval functions:

$$\rho_{ij}^{(A)} = 1 - (1-g)\exp[-\theta\tau^{i}] \qquad (A)$$

$$\rho_{ij}^{(B)} = 1 - (1-g)\exp[-\theta\tau^{i-j}] \; . \qquad (B)$$

In order to make predictions from these models $\rho_{ij}^{(A)}$ and $\rho_{ij}^{(B)}$ were substituted for ρ_{ij} in Eq. 4. These two models were then fit to the unconditional lag curves from Experiment I using the same method as before; i.e., a minimum χ^2 estimate of the four parameters was obtained. For Model A the minimum χ^2 was 51.47 and the parameter estimates were $r = 2$, $\alpha = .30$, $\theta = .90$, and $\tau = 1.0$. For Model B the χ^2 procedure also yielded a best fit when $\tau = 1.0$. Since the Models A and B are identical when $\tau = 1.0$ the χ^2 and the parameter estimates are the same for both models.

Because the minimum χ^2's for Models A and B were somewhat larger than that for the version in the body of the paper, and because the earlier version seemed more reasonable, we have relegated these two models to an appendix. It should be noted, however, that these models do not require the assumption of a decay process. More precisely, the assumption of a decay process does not improve the fit of Models A and B (i.e., when τ equals one the models predict no decay in LTS). These alternative models are of interest also because they represent various branches of the general family of multi-process memory models formulated by Atkinson and Shiffrin (1965). There remain many other branches, however, that are as yet unexplored. In this regard, it is interesting to speculate that a model postulating a larger amount of information transfer when an item first enters the buffer, with smaller amounts thereafter, might fit the data as well as the version in this paper without requiring an LTS decay process.

References

Atkinson, R. C., Bower, G. H., and Crothers, E. J. An introduction to mathematical learning theory. New York: Wiley, 1965.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Tech. Rep. 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in The anatomy of memory, vol. 3, proceedings of the third conference on learning, remembering and forgetting, edited by D. P. Kimble, Palo Alto, California: Science and Behavior Books and Company, 1966.)

Bower, G. H. Notes on a descriptive theory of memory. In D. A. Kimble (Ed.) Learning, remembering, and forgetting: proceedings of the second conference. New York: New York Acad. Sci., 1966.

Brelsford, J. W., Jr., Keller, L., Shiffrin, R. M., and Atkinson, R. C. Short-term recall of paired-associates as a function of the number of interpolated pairs. Psychonomic Science, 1966, 4, 73-74.

Broadbent, D. E. Flow of information within the organism. Journal of Verbal Learning and Verbal Behavior, 1963, 4, 34-39.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

Estes, W. K. Learning and memory. In E. F. Beckenbach (Ed.) Conceptual bases and applications of the communications sciences, 1966, in press.

Holland, P. A variation on the minimum chi-square test. Tech. Rep. 85, Institute for Mathematical Studies in the Social Sciences, Stanford Univer., 1965.

Katz, L. A technique for the study of steady-state short-term memory. Psychonomic Science, 1966, 4, 361-362.

Peterson, L. R., and Peterson, Margaret. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

Phillips, J. L., Shiffrin, R. M., and Atkinson, R. C. The effects of list length on short-term memory. Journal of Verbal Learning and Verbal Behavior, 1966, in press.

Wickelgren, W. A. Acoustic similarity and retroactive interference in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 53-61.

Wickelgren, W. A., and Norman, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, in press.

Yntema, D. B., and Mueser, G. D. Keeping track of variables that have few or many states. Journal of Experimental Psychology, 1962, 63, 391-395.