Psychological Review
1976, Vol. 83, No. 3, 190–214

Retrieval Processes in Recognition Memory

Roger Ratcliff and Bennet B. Murdock, Jr.
University of Toronto, Toronto, Canada

A method of analyzing reaction time data in recognition memory is presented, which uses an explicit model of latency distributions. This distributional method allows us to distinguish between processes in a way that the traditional measure, mean latency, can not. The behavior of latency distributions is described, and four experiments are reported that show how recognition accuracy and latency vary with independent variables such as study and test position, rate of presentation, and list length. These data are used to develop and test the empirical model. The resulting analyses, together with functional relationships derived from the experimental data, are used to test several theories of recognition memory. The theories examined all show problems in light of these stringent tests, and general properties required by a model to account for the data are suggested. As well as arguing for distributional analyses of reaction time data, this paper presents a wide range of phenomena that any theory of recognition memory must explain.

Over the last few years, researchers have been developing theories of recognition memory based not only on accuracy measures but also on latency measures. In this article, we consider latency measures in recognition memory. Results from four experiments are presented, and an empirical model for latency distributions is developed. Latency distributions are shown to provide much more information than can be obtained from mean latency, the most common dependent variable in reaction time measurements. From this, a strong case is made for the study of distributional properties by showing how some current theories are inadequate or wrong when examined in the light of distributional analyses. These recent theories are further evaluated using functional relationships extracted from results of the four experiments presented.

This research was supported by Research Grants APA 146 from the National Research Council of Canada and OMHF 164 from the Ontario Mental Health Foundation.

We would like to thank Khalid El-Ayat and Howard Kaplan for assistance in computer programming.

The senior author is responsible for the theoretical analysis and for the material development; the junior author contributed to some of the experimental work (and paid the computer bills for the parameter estimations).

Requests for reprints should be addressed to B. B. Murdock, Department of Psychology, University of Toronto, Toronto, Ontario, M5S 1A1, Canada.

Perhaps one of the more important advances in the theory of recognition memory came with the application by Egan (Note 1) of the decision mechanism of signal detection theory. Before this, threshold theories prevailed, but Egan (Note 1) demonstrated that this class of theory is inadequate (see Murdock, 1974, chap. 2). The decision mechanism of signal detection theory is present in some form or another in most current theories of recognition memory and is probably a major factor in the success of many of these theories. Figure 1 is an attempt to order theories of recognition memory both historically and in terms of antecedents, which are not necessarily causal.

The strength theory of Norman and Wickelgren (1969) and Wickelgren and Norman (1966) assumes that, along with the trace, some continuous variable is stored that provides a measure of the subject's familiarity with the test item. The Bower (1967) and Norman and Rumelhart (1970) attribute models use a discrete ensemble of features to represent the memory trace, and the finite state models of Bernbach (1967) and Kintsch (1967) suppose that an item is in one of two or three memory states, respectively. Each of these three classes of model does reasonably well in dealing with a substantial body of data, but it has been argued (Anderson & Bower, 1972) that this success is not so much due to the memory representation employed

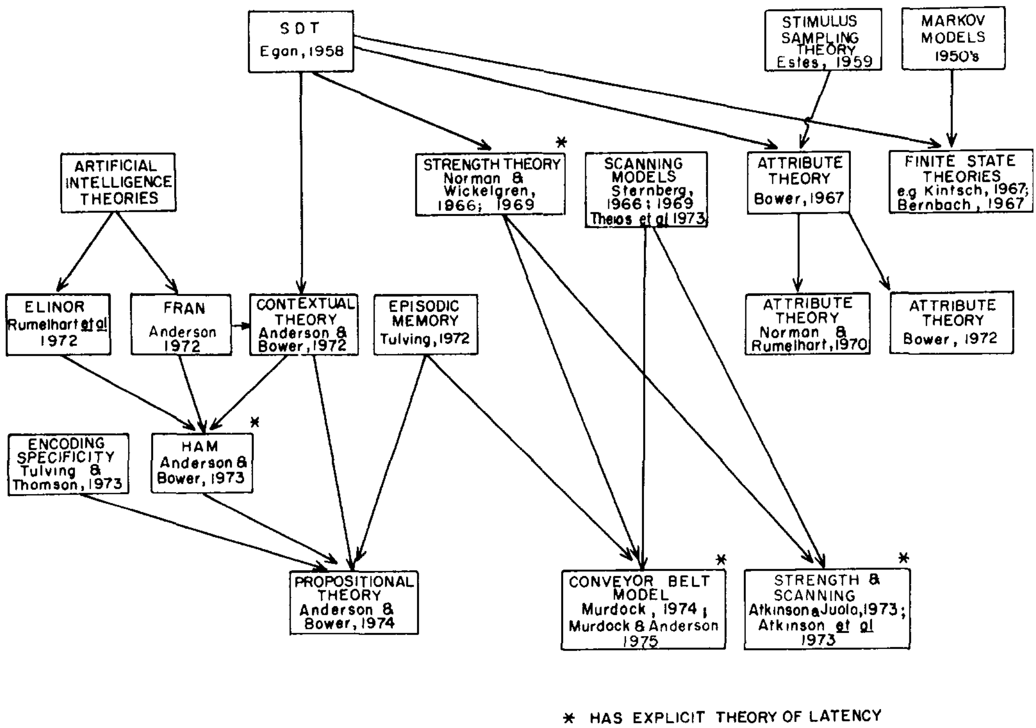

FIGURE 1. Schematic representation of theoretical development in the field of recognition memory. (SDT = Signal Detection Theory; HAM = Human Associative Memory; FRAN = Free Recall in an Associative Net.)

as to the use of signal detection theory. Anderson and Bower (1972) mounted a major attack on these classes of theory, the main thrust of their argument being aimed at the ahistorical nature of the memory trace. They found that in list discrimination experiments, subjects were capable of much finer discriminations than the earlier theories allowed. A theory was developed stating that subjects associate "list markers" with study items in an all-or-none process and that the recognition process, as before, utilizes signal detection theory.

Following hard on the heels of this work and the FRAN model of Anderson (1972) came the comprehensive book, *Human Associative Memory* (HAM; Anderson & Bower, 1973). This work provides a major integration of many different areas of cognitive psychology under one theoretical structure. Later, a comparison will be made between HAM's predictions and our empirical findings. Anderson and Bower (1974) modified their 1972 position to produce a propositional representation of contextual information and to account for context effects in recognition memory (Tulving & Thomson, 1973; Watkins & Tulving, 1975). Some of the modifications seem to weaken the position taken in HAM, but they appear necessary to deal with empirical results. This will be discussed in detail later.

Another research program that has generated considerable interest and inquiry is the list-scanning work of Sternberg (1966, 1969a). An excellent summary of current issues in this research is contained in Sternberg (1975). This investigation is primarily aimed at dealing with short lists (within memory span), but it has provided part of the impetus for the Atkinson and Juola model (1973) and the conveyor belt model of Murdock (1974), both of which deal with longer supraspan lists. The Atkinson and Juola model stores item information in two memory systems, the conceptual store and the event knowledge store. Retrieval consists of a decision, using signal detection theory, based on familiarity (or strength) of the item in the conceptual store, with either a fast response being emitted or an extended search of the event knowledge store taking place. The conveyor belt model (see also Murdock & Anderson, 1975) supposes that items are encoded in some kind of temporal format such as an episodic store (Tulving, 1972) and that retrieval consists of a very rapid 5 msec/item, backward-serial scan through the list, so that when a match is found a response is initiated. This quick reveiw of recognition memory theory can by no means be considered exhaustive. Rather, it is an attempt, at least, to note the major theories and their ancestors.

In comparison, distributional properties of reaction times have received little attention in recent years. There was an awareness in the early days of reaction time research that mean reaction time was not sufficient and that some note of distributional properties should be taken. For example, in 1868, Donders (1969) reported minimum reaction times, as well as means to support his subtractive method. More recently, Woodworth (1938, chap. 14) presented an excellent review of research in reaction time, which we feel has real importance even today. He presented several examples in which arguments are based on the behavior of latency distributions and not simply mean latency. For example, he cited the work of Johanson (1922), which demonstrated that simple reaction time could be speeded up using incentives or punishment. If mean reaction time alone had been reported, we would not have known whether overall reaction time decreased or whether slower responses were just speeded up. The distributions, though, showed that overall reaction time decreased. However, none of this early work appears to have attempted to deal with the mathematical properties of distributions.

Perhaps the most explicit and authoritative review of mechanisms, mathematical models, and applications is still the review of McGill (1963). Since that review, relatively little use of distributional analysis has been made (an important exception being Metzler & Shepherd, 1974, p. 182), although its importance has been noted at times (Broadbent, 1971, p. 320; Townsend, 1972). What research there has been seems to have developed along two different lines.

First, Sternberg (1969b; Note 2) has been concerned with the additive-factor method that he used to establish the existence of processing stages in his paradigm. This powerful method uses the properties of cumulants (essentially moments—mean, variance, etc.) in testing hypotheses about stages. Distributional properties enter through the cumulants, and so explicit forms of the distributions need not be assumed. Sternberg (1969b) used this method to identify four stages in item recognition—stimulus encoding, comparison, decision, and translation plus response organization. However, Aubé and Murdock (1974) found results in a Sternberg procedure that argued against an additive stage model. It is somewhat difficult to see where to proceed from this point. Sternberg (Note 3) introduced some further properties of reaction time distributions, based on cumulative distribution functions, to argue against self-terminating models of memory search; again, these did not require explicit forms of the distributions (see also Sternberg, 1975).

Second, researchers such as McGill (1963), McGill and Gibbon (1965), Hohle (1965), and Snodgrass, Luce, and Galanter (1967) have taken a rather more extreme approach in which distributions underlying stages are specified. From these, the overall distribution can be deduced by convolution if it is assumed that the underlying stages are independent. This convolution is fitted to the empirical data, and variation in reaction time produced by changes in experimental conditions should show up in changes in the parameters of the stage affected by the experimental manipulation. However, as pointed out by Sternberg (1969b), if such an approach fails, it is difficult to decide which of the several assumptions is at fault. Also, as we shall see later, rather different assumptions about processing stages and their distributions can give rise to similar latency distributions, and so there is a problem in deciding between models.

We shall attempt to remedy this lack of consideration of distributional properties by applying distributional analyses to several recognition memory studies. From this, it will become obvious that such an approach is extremely useful in evaluating models and theories. The method we shall promote represents distributions in an explicit form (which is amenable to an additive stage analysis) but does not directly identify stages of processing with component mathematical stages.

Experiments: Procedures and Results

Four experiments are reported and the data serve two main purposes: first, to test an empirical model for response time distributions, and second, to provide functional relationships between accuracy and various experimental variables such as list length, rate of presentation, and output position. These functional relationships, along with the distributional analyses, are used to test various theories of recognition memory. Experiments 2 and 3 were reported by Murdock and Anderson (1975, Experiments III and V, respectively), but several additional analyses are presented here.

General Method

A discrete trials standard recognition memory procedure was employed. A list of study items was presented, followed by a test list containing all the study items and an equal number of new items, in random order. The lists were random samples from the University of Toronto wordpool, a collection of 1,024 two-syllable common English words not more than eight letters long, with homophones, contractions, and proper nouns excluded. Each trial consisted of a random selection from the wordpool. In some experiments, repetitions were prohibited until two successive lists had intervened; in others, there were no repetitions per session. List generation, display, and response recording were controlled by a

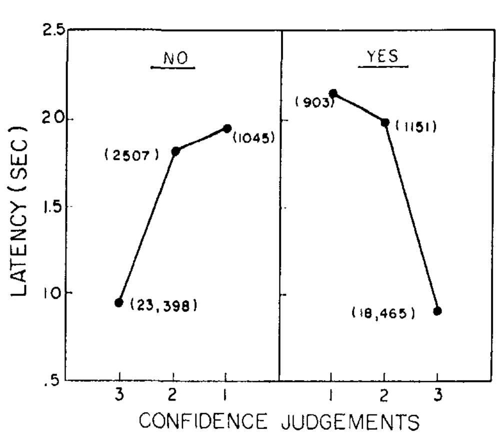

FIGURE 2. Plot of mean latency versus confidence judgment for Experiment 1. (The number of observations per plotted point is shown in parentheses next to the point.)

PDP-12A laboratory computer. The presentation time was typically 1 sec per item, and the list was terminated by an instruction asking the subject to press a response key to start the test phase. The test phase was self-paced, and items stayed in view until a response was made. A confidence judgment procedure was employed, and the subjects had to respond on a 6-point scale from −−− ("sure new") to +++ ("sure old") by pressing the appropriate response key. For each item tested, input and output position (output position only for new items), confidence judgment (i.e., key pressed), and the latency (stimulus onset to response key depression) were recorded. In some experiments, a 25 msec time base was used; in others, a 5 msec time base was used. Subjects were undergraduates in psychology at the University of Toronto and were paid $30 for the 12 sessions. Each experiment used four subjects except Experiment 3, which used five.

Experiment 1

Method

The study list consisted of 16 items and the test list of 32 (16 old and 16 new). There was 1 practice session and 12 experimental sessions, with no repetitions of study or test items in the 32-trial session. The time base used was 5 msec, and latencies up to 12 sec were recorded.

Experiment 1, besides replicating earlier experiments (Murdock, 1974; Murdock & Anderson, 1975), served to show that results in these earlier experiments were not an artifact of display graphics (a much clearer set was used here), time base (a faster time base was employed), randomization process (previously, items were tested in their wordpool order), or repeated items between lists (no repetitions per session in this experiment).

Results

Variation of mean latency with confidence judgment. Murdock and Dufty (1972) found that mean latency decreased with increasing confidence. Figure 2 both demonstrates and replicates that finding.

Mandler and Boeck (1974) conducted an experiment using a recognition memory study–test procedure. Their subjects were required to make a yes/no response followed by an unpaced confidence judgment. Mean latency plotted against confidence judgment showed the same trends as shown in the present Experiment 1. This indicates that these effects are not artifacts of the six-key (as opposed to two-key) response procedure.

Later it will be argued that lower confidence judgments involve more complex processing such as a second memory interrogation or a second decision after some alteration in criteria, so, in line with Murdock (1974), only high-confidence results are examined here.

Accuracy functions. Murdock and Anderson (1975) present strong arguments for the use of lag as the independent variable in these tasks. However, because this obliterates the distinction between input and output interference and because several of the analyses performed maintain this distinction, a two-way classification of input and output position is employed. Figure 3 graphs the change in proportion of high-confidence hits, misses, correct rejections, and false alarms with input (for hits and misses only) and output position. It can be seen that the decrease in the number of hits with output position is more pronounced than the decrease in the number of correct rejections (see also Murdock & Anderson, 1975). If a signal detection theory (or strength theory) approach is employed, then for increase in output position, the "signal" and "noise" distributions move closer together but the criterion remains more or less in the same place relative to the noise distribution. For a fixed output position, the signal and noise distributions are closer together for earlier input positions.

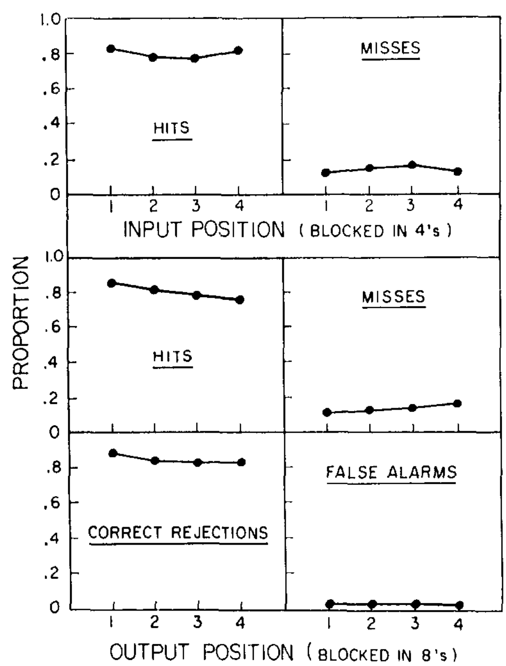

FIGURE 3. Proportion of high-confidence hits, misses, correct rejections, and false alarms for a range of input and output positions for Experiment 1.

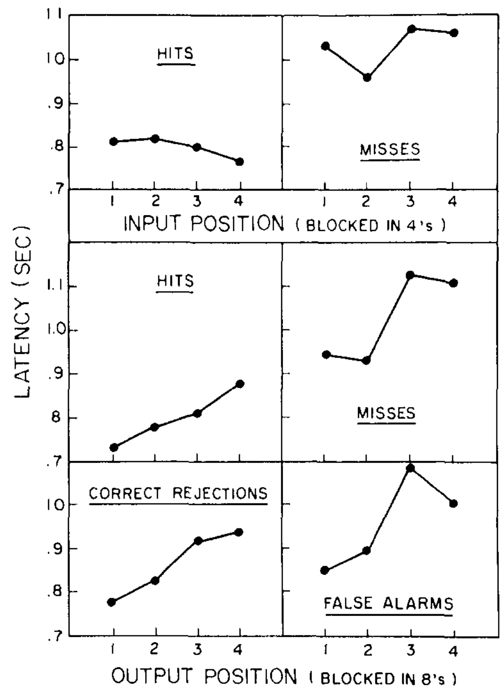

FIGURE 4. Mean latency of high-confidence hits, misses, correct rejections, and false alarms for a range of input and output positions for Experiment 1.

Latency functions. Graphs of latency plotted against input (for hits and misses only) and output position for high-confidence hits, misses, correct rejections, and false alarms are shown in Figure 4. Some of these results replicate those presented in Murdock (1974) and Murdock and Anderson (1975), which provided the main impetus for the backward-serial scanning, or conveyor belt, model. An important point to note is that for some subjects, high-confidence correct rejections are almost as fast as high-confidence hits, compared with the spread in response times (see histograms later)—for example, Subject 3 had a mean hit latency of .757 sec and a mean correct rejection latency of .823 sec. Figures 3 and 4 show a primacy effect for latency and accuracy of hits versus input position, an effect that has important consequences for some of the models discussed later.

Experiment 2

Method

The study list consisted of 15 items, and the test list, of 15 old and 15 new items. Two intervening lists were required before any word could reappear. The subjects were given 1 practice session (32 trials) and were then tested for 12 further sessions, four sessions at each of three presentation rates. Two subjects, who had d' values of about 2.0, had rates of .6, .9, and 1.5 sec/item, and the other two subjects, who had d' values of 3 or greater, had rates of .6, .8 and 1.2 sec/item (see Murdock, 1974, p. 273).

This experiment was reported by Murdock (1974). It is presented here because rate of presentation effects are of particular interest and because further analyses have been carried out.

Results

Latency and accuracy plotted against input and output position show trends similar to those found in Experiment 1 and are therefore not presented here. The effect of rate of presentation on both accuracy and latency is shown in Figure 5. It is interesting to note that as rate of presentation decreases, accuracy increases and there is a small increase in latency. Also, Murdock (1974) reported that slopes of the linear lag and output-position/latency functions change little with rate of presentation.

Experiment 3

Method

On each trial a list of L items was presented, and then $2L$ items were tested (half old items and half new). The number of lists per session was inversely proportional to L; so for $L = 4$, 8, 16, 32, and 64 (a between-sessions variable), the number of lists per session was 128, 64, 32, 16, and 8, respectively. The five subjects were each given a practice session followed by four sessions per condition, 20 sessions per subject in all. No word could be repeated until at least two lists had intervened, and the presentation rate was 1.2 sec per item.

Results

Since our main concern is with supraspan lists, results from list lengths 16, 32, and 64 only are considered. Figure 6 shows mean latency for high-confidence hits and correct rejections as a function of output position for the three list lengths. The main result is that as list length increases, the slope of the latency/output-position function decreases and mean latency increases.

Experiment 4

Method

In this experiment there were two types of task. The first was a standard study–test task having 24 study items—8 presented once (1P) and 8 presented twice (2P) and thus 24 tokens and 16 types—and 32 test items (half old and half new). The second type of task was a word–nonword discrimination task. A list of 24 study items with the same composition as in the study–test task was presented to the subject. A displayed message then told the subject whether each trial was a memory or word–nonword task. For the word–nonword task, 32 test items were presented—half were the stimulus words and half were scrambled versions of new words from the wordpool. The subject's task was to press the high-confidence new key for a nonword and the high-confidence old key for a word. There were 16 memory and 16 word–non-

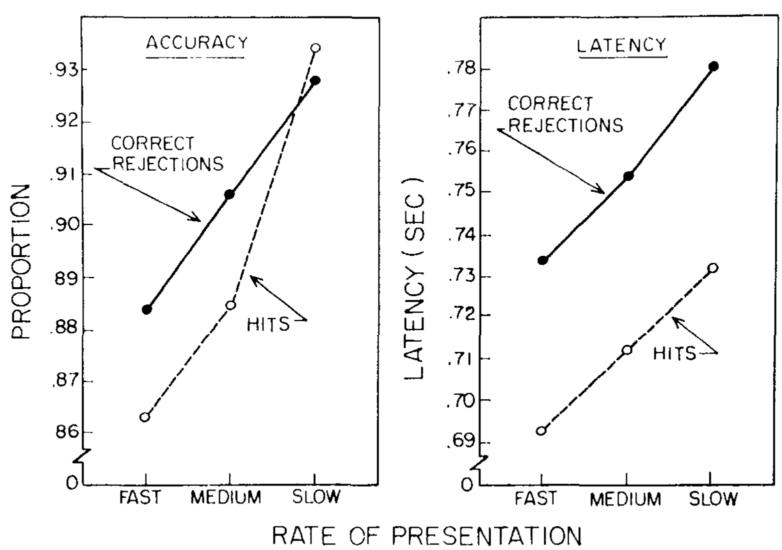

FIGURE 5. Proportion and mean latency of hits and correct rejections plotted against rate of presentation for Experiment 2.

word randomly ordered trials per session, with 8 sessions preceded by one practice session.

The word-nonword task was used as a check on encoding facilitation (Kirsner & Craik, 1971); thus, if there was any encoding facilitation at test because of the second presentation of the word, it would show up in this task.

Results

Figure 7 shows both accuracy and latency data for the memory test. An important result to notice is that for 2P items, accuracy is higher and latency lower than for 1P items. In Figure 8, latency and accuracy are shown as a function of output position for the word–nonword task. It is interesting to notice that this task exhibits a slope for latency plotted against output position; this is examined later. There is almost no difference in latency for 1P and 2P words in the word-nonword task. This suggests that there is little encoding facilitation and that the difference between 1P and 2P items in the memory task arises from memory or decision processes.

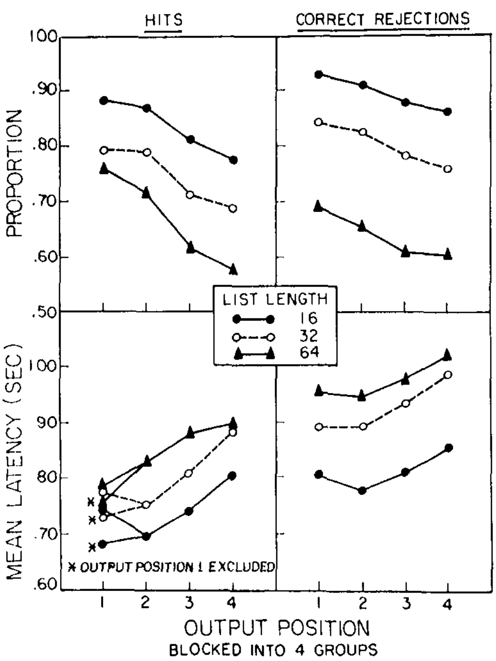

FIGURE 6. Mean latency of high-confidence hits and correct rejections as a function of output position for list lengths 16, 32, and 64 in Experiment 3. (Note that the output-position blocking is different for the three list lengths: for list length 16, blocks of 8; for 32, blocks of 16; and for 64, blocks of 32.)

AN EMPIRICAL MODEL FOR LATENCY DISTRIBUTIONS

In most reaction time studies, the dependent variable is taken to be mean latency. On the basis of the model developed here, it will be shown that in some cases this measure is inadequate and misleading. The alternative we wish to suggest is the consideration of latency distributions, which provide much more information than mean latency (see McGill, 1963).

In this section, three models are considered; the method of fitting the models to the data and estimating parameter variance is described; and results of applying the best of these models to the four experiments are presented.

The Mathematical Models

Gamma Distribution

Gamma distribution has often been used with some success in modeling time-dependent processes (Anderson & Bower, 1973; McGill, 1963; McGill & Gibbon, 1965; Ratcliff, in press; Snodgrass et al., 1967; Townsend, 1972). To fit the experimental data, the displaced gamma distribution must be used because there is a delay in time (c) reflecting encoding of the probe and response output. The expression for this distribution is given by

$$f(t) = c + \frac{\rho(\rho t)^{\alpha-1}e^{-\rho t}}{\Gamma(\alpha - 1)} \qquad (1)$$

For the parameter α taking integer values, the second term is the convolution of α independent, exponentially distributed, random variables each with rate parameter ρ. (For further properties and discussion see Cox, 1962.)

Lognormal Distribution

This distribution has been used implicitly in many reaction time analyses in psychology, because a lognormal distribution is transformed to a normal distribution by a logarithmic transformation. This transformation is used in the analysis of variance of reaction time data to make the variance independent of the mean.

Because the lower limit of this distribution is pegged to zero, a displaced lognormal distribution (by c) must be used:

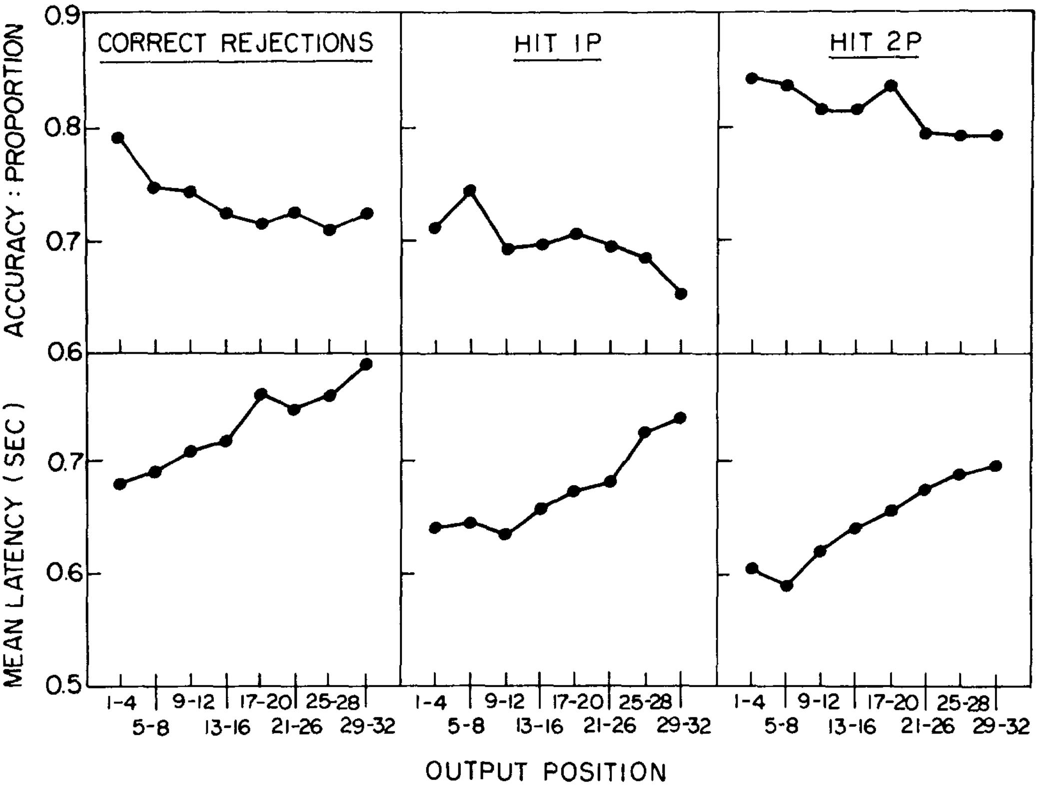

FIGURE 7. Proportion and mean latency of hits and correct rejections plotted against output position for the memory test, Experiment 4. (1P = presented once; 2P = presented twice.)

$$g(t) = \frac{\tau}{(c-t)\sqrt{2\pi}} e^{-\frac{1}{2}\{\tau \ln [(t-c)/t_0]\}^2} \quad (2)$$

where $\ln(t-c)$ is normally distributed with a mean of $-\ln(1/t_0)$ and a variance of $1/\tau$.

Convolution of Exponential and Normal Distributions

This distribution was used by Hohle (1965) to fit choice reaction time distributions, but it seems to have been used little since. This distribution represents the situation in which several of the processes have normally distributed latency functions (the convolution of two normal distributions is another normal distribution) and one process has an exponential distribution. The expression for the convolution is

$$h(t) = \frac{e^{-[(t-\mu)/\tau]+\sigma^2/\tau^2}}{\tau\sqrt{2\pi}} \int_{-\infty}^{[(t-\mu)/\sigma]-\sigma/\tau} e^{-y^2/2} dy \quad (3)$$

where μ and σ^2 are the mean and variance, respectively, of the normal distribution and τ is the parameter (and mean) of the exponential distribution.

Fitting the Models and Goodness-of-Fit

The method of maximum likelihood was used to fit the three models to the data and obtain parameter estimates.

Let $L(\underset{\sim}{\theta}) = f(x_1, \ldots, x_n, \underset{\sim}{\theta})$ be the joint probability density function of the data sample $x_1, \ldots, x_n$ with function parameters $\underset{\sim}{\theta}$ (note that $L(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(x_i, \underset{\sim}{\theta})$ for a random sample). Then, the maximum likelihood estimates $\hat{\underset{\sim}{\theta}}$ are obtained by maximizing $L(\underset{\sim}{\theta})$, or equivalently $\ln L(\underset{\sim}{\theta})$, with respect to $\underset{\sim}{\theta}$. As a check, this was done two ways for a few sets of data: first by directly maximizing $\ln L(\underset{\sim}{\theta})$ using the SIMPLEX routine (Nelder & Mead, 1965) and second by numerically solving the

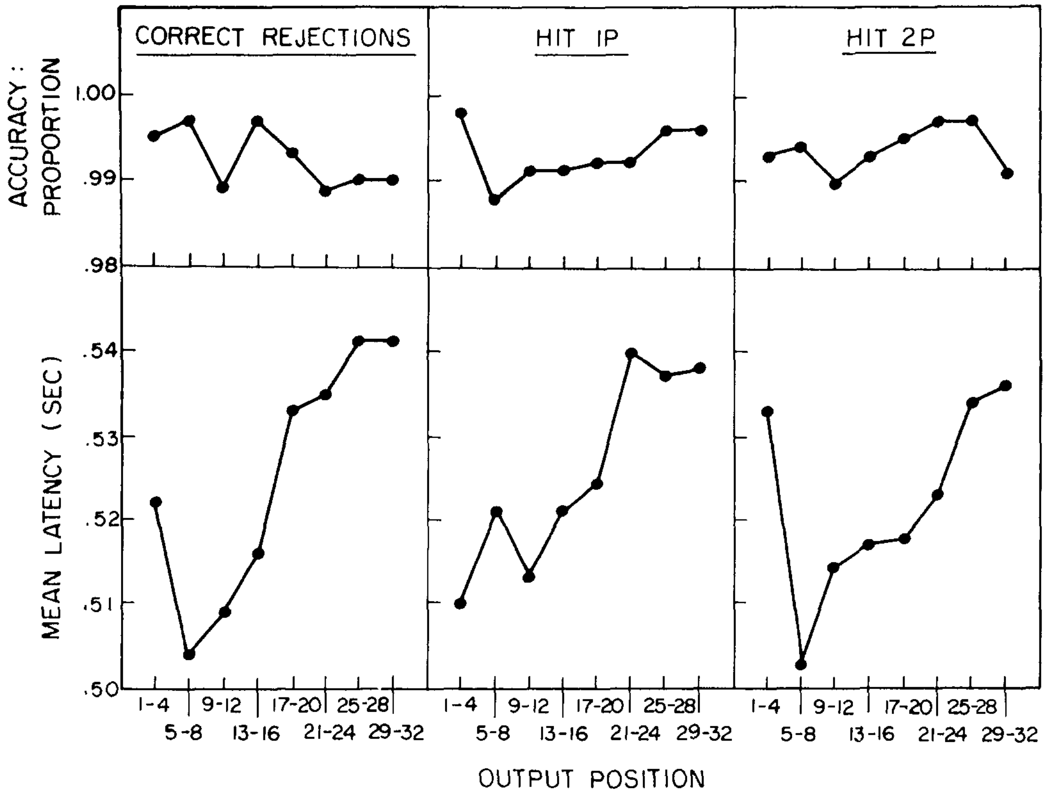

FIGURE 8. Proportion and mean latency of hits and correct rejections plotted against output position for the word–nonword test, Experiment 4. (1P = presented once; 2P = presented twice.)

set of nonlinear equations (obtained by setting the first derivatives of ln $L(\underset{\sim}{\theta})$ with respect to $\underset{\sim}{\theta}$ equal to 0) using ZSYSTM (Brown, 1969). The two sets of estimates were identical, although the SIMPLEX method was preferred because it had fewer convergence problems and was cheaper than the ZSYSTM method.

Only high-confidence responses were fitted, and these were blocked by input and output position so that no less than about 300 responses were used for parameter estimation, except in the case of high-confidence misses and false alarms. Responses longer than 2.5 sec were eliminated, although a check carried out showed that the parameter estimates were altered little and that parameter trends were maintained (though goodness-of-fit became a little worse with those responses included). Individual subjects' data were fitted, and averages over subjects were calculated on the results of the fits to the individual data. This is most important, because combining distributions results in a distribution that can be quite different in shape from the original individual distributions.

A χ^2 goodness-of-fit statistic was obtained by blocking latencies into 50-msec steps and forming

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where k is the number of frequency classes with expected value E_i greater than 5 (for E_i less than 5, the frequency classes were grouped and degrees of freedom reduced appropriately) and O_i is the observed number of responses in class i. The number of degrees of freedom was $k-4$, as three parameters were fitted for each model. Table 1 contains values of χ^2 from a nonsystematic sample of fits of the three models to data from Experiments 1 and 4.

It can be seen that the gamma distribution is clearly inferior to the convolution and lognormal distributions, with the convolution being slightly better than the lognormal. Even though the fits of the convolution are better than the fits of the other two distributions, they often give statistically significant χ^2 values. Three things can be said about the χ^2 values being significant. First, if the sample size is reduced to about 100, as for misses and false alarms, then nonsignificant values of χ^2 would be obtained, but parameter trends would be more noisy. Second, blocking over a range of input and output positions means that the resulting distribution is a combination of latencies with differing parameters, and so an inflated value of χ^2 may be expected. Third, average response time decreases over the course of testing (Burrows & Murdock, 1969), which again leads to an inflated value of χ^2 because the distribution will be a combination of latencies with different parameters.

Maximum likelihood estimators have nice asymptotic (large sample) properties. There is a theorem (Wilkes, 1962) that states that $\hat{\underset{\sim}{\theta}}$ (the estimates of $\underset{\sim}{\theta}$) are asymptotically distributed $N(\underset{\sim}{\theta}, I^{-1})$ under certain regularity conditions, where I is the information matrix with elements

$$I_{ij} = -E\left(\frac{\partial^2 \ln L(\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j}\right).$$

For the convolution model, the diagonal elements of I^{-1} provide variance estimates of μ, σ, and τ, and these estimates allow statistical comparisons to be carried out.

Table 2 contains a comprehensive sample of standard deviation estimates for values of μ, σ, and τ obtained in Experiments 1 through 4.

TABLE 1

COMPARISON OF χ^2 VALUES BETWEEN MODELS

Response	Input position[a]	Output position[a]	Convolution χ^2	Convolution df	Lognormal χ^2	Lognormal df	Gamma χ^2	Gamma df
Experiment 1, Subject 1								
Hit	1	1	38.9	11	35.1	11	124.9	12
	1	2	40.7	12	13.6	11	61.5	11
	2	1	20.9	9	28.5	9	43.9	9
	2	2	37.9	11	44.0	10	44.1	10
Miss	1	3	5.3	6	5.0	5	7.1	6
	1	4	5.2	3	3.0	3	6.2	3
	2	3	3.9	6	3.9	6	6.5	5
	2	4	4.4	5	2.0	5	7.4	5
Experiment 4, Subject 3								
Hit (1P)[b]		1	18.9	8	13.5	7	43.2	8
		2	14.2	8	14.1	8	38.0	8
		3	25.0	9	47.8	10	53.0	9
		4	50.4	11	56.4	12	89.0	12
Correct rejection		1	56.9	13	37.0	11	84.2	12
		2	26.8	13	13.4	12	46.4	13
		3	36.5	14	24.5	13	60.2	13
		4	35.2	16	39.2	15	74.0	16
Word, hit (1P)		1	2.6	7	36.4	7	11.9	7
		2	19.7	7	49.2	7	27.7	7
		3	19.4	8	25.6	8	39.8	8
		4	28.2	8	32.0	8	38.3	8
Nonword, correct rejection		1	17.3	9	21.7	9	69.3	9
		2	24.4	8	30.9	8	60.3	8
		3	49.2	10	51.5	10	122.6	10
		4	42.0	10	41.1	10	81.3	10

[a] Input and output position are blocked by eights.
[b] 1P indicates item was presented once.

TABLE 2

Asymptotic Standard Deviation Estimates[a] for a Range of Parameter Values with Sample Size 300[b]

μ	s_μ	σ	s_σ	τ	s_τ
.500	.005	.030	.004	.100	.007
.500	.007	.030	.006	.200	.013
.500	.008	.030	.007	.300	.019
.500	.006	.040	.005	.100	.008
.500	.008	.040	.007	.200	.014
.500	.009	.040	.009	.300	.019
.500	.007	.050	.006	.100	.009
.500	.009	.050	.008	.200	.014
.500	.010	.050	.009	.300	.020
.600	.008	.040	.006	.200	.014
.600	.009	.040	.008	.300	.019
.600	.010	.040	.009	.400	.025

Note. An independent check to see if the variance estimates were reasonable was carried out. If τ were made large relative to σ, then the distribution would approximate an exponential, and $s_\tau = \tau$. For $\tau = 1.0$ ($\sigma = .03$, $N = 1$), $s_\tau = 1.027$, and for $\tau = 1.5$ ($\sigma = .03$, $N = 1$), $s_\tau = 1.527$ (where N is the sample size).

[a] These estimates are square roots of the asymptotic variance estimates.

[b] To obtain values for sample size N, multiply s by $\sqrt{300/N}$.

Results of Applying the Convolution to the Experimental Data

In this section, only high-confidence hit, miss, correct rejection, and false alarm latencies are analyzed, and unless otherwise noted, average parameter trends for the group of subjects are representative of each individual. Note that any statistical tests on μ, σ, and τ can be carried out using the standard deviations presented in Table 2.

Experiment 1

In Figure 9 are shown histograms of latency distributions plus fits of the convolution for hits, misses, and correct rejections for Subject 4. (Hits and misses are blocked in two groups of input positions. Hits, misses, and correct rejections are all blocked in four groups of output positions.) There were too few false alarms to permit fitting, although for other subjects, the trends were the same for

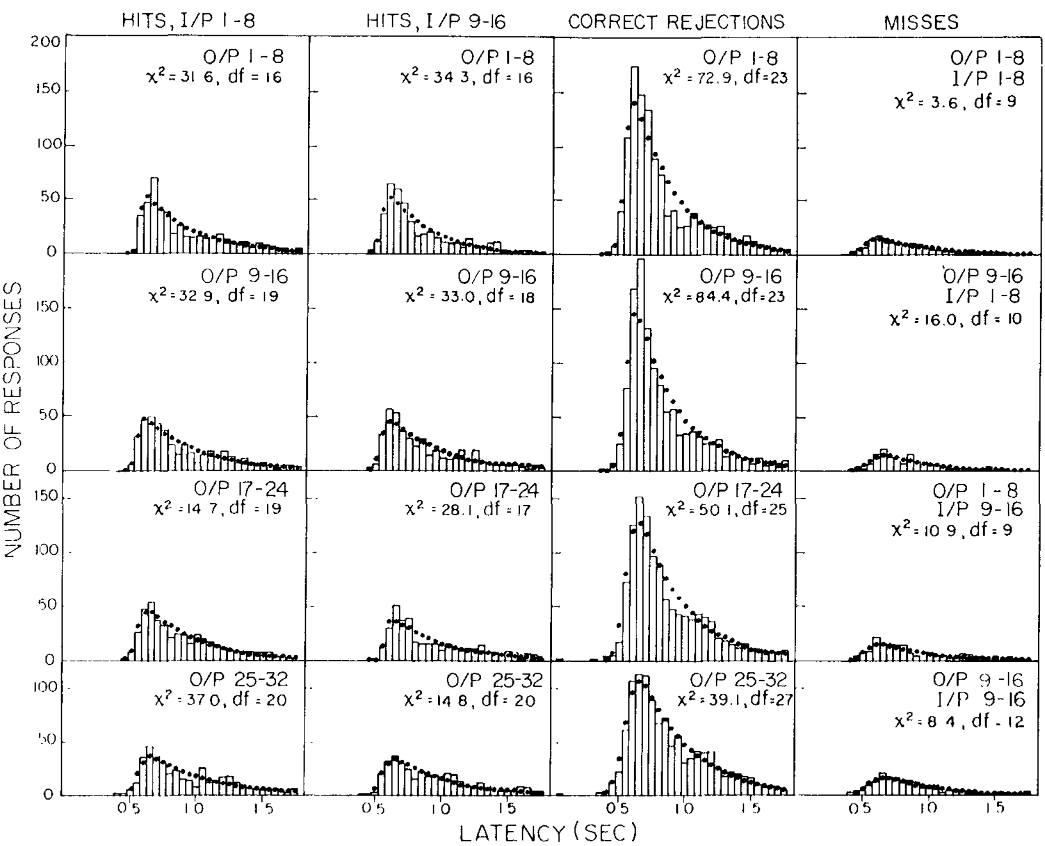

FIGURE 9. Empirical and fitted latency distributions for Subject 4, Experiment 1. (I/P = input position; O/P = output position.)

hits, misses, and correct rejections as shown in Figure 9. These results show that the leading edge and mode of the distribution change little with output position, and that change in mean latency (Figure 4) comes from a greater number of slower responses. In Figure 10, it can be seen that these two effects are mirrored in group average parameters of the convolution. The mean of the normal μ increases a little with output position, and this corresponds to the behavior of the leading edge and mode, whereas the exponential parameter τ gives the main contribution to the change in mean latency. Table 3 shows values of χ^2 for Subjects 1, 2, and 3 for hits, misses, and correct rejections over the range of input and output position blocks. These parameter trends, which are present in data from all the experiments, provide much of the justification for use of the convolution. Also, it is these trends that turn out to be very important in criticizing various models of the underlying processes.

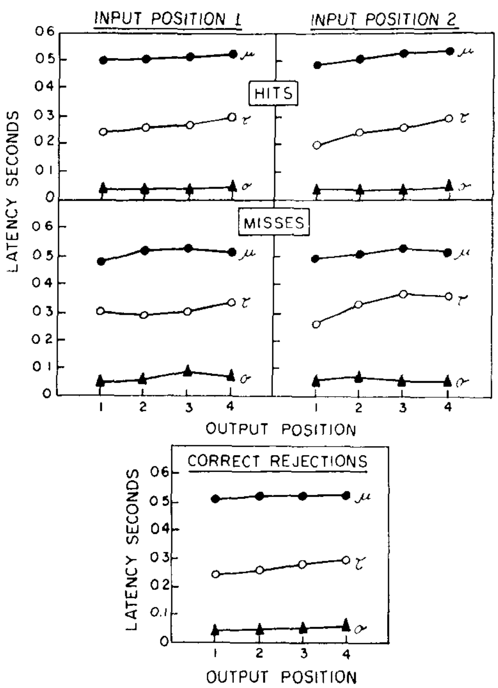

FIGURE 10. Distribution parameters for hits, misses, and correct rejections as a function of input and output position for Experiment 1. (Note that there is a weak effect of output position on μ; the average slope of a linear least squares fit on μ for hits and correct rejections combined is 1.1 msec/item.)

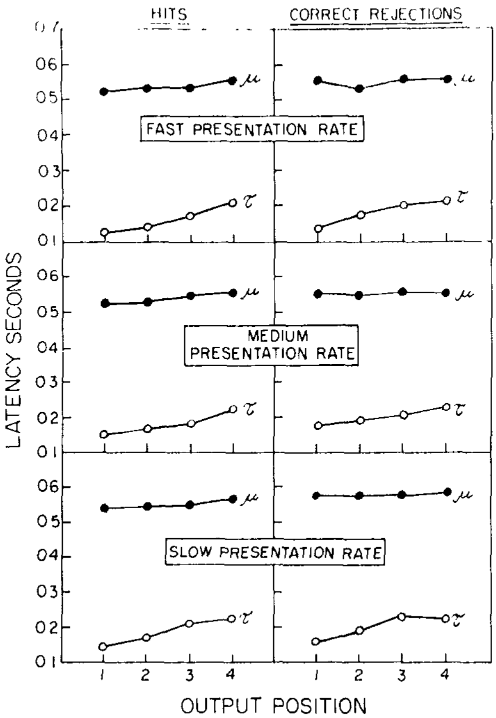

FIGURE 11. Distribution parameters for hits and correct rejections as a function of rate of presentation and output position for Experiment 2. (Note that the average slope of a linear least squares fit on μ as a function of output position for hits and correct rejections combined is .9 msec/item.)

Experiment 2

As in Experiment 1, change in the exponential parameter τ with output position is mainly responsible for the change in mean latency. Figure 11 shows that both μ and τ increase a little as rate of presentation decreases.

Experiment 3

Figure 12 shows that changes in list length produce changes in both μ and τ, although the rate of change of τ with output position decreases as list length increases. As before, change in mean latency with output position is mainly attributable to changes in τ.

Experiment 4

A sample of fits of the convolution and histograms of the latency distributions for both memory and word–nonword trials are shown in Figure 13. These indicate that the convolution is adequate for both 2P memory

TABLE 3

χ^2 Goodness-of-Fit Statistics for Subjects 1, 2, and 3 of Experiment 1, by Input and Output Position

Response	Input position[a]	Output position[a]	Subject 1			Subject 2			Subject 3		
			χ^2	*df*	*N*	χ^2	*df*	*N*	χ^2	*df*	*N*
Hit	1	1	38.9	11	615	17.0	12	472	66.7	14	615
	1	2	40.7	12	642	32.2	13	470	39.9	14	700
	1	3	55.1	13	633	17.4	15	433	81.2	15	739
	1	4	27.2	13	648	16.3	14	410	60.5	16	738
	2	1	20.9	9	609	17.9	11	516	40.1	13	675
	2	2	37.9	11	714	16.6	13	451	97.0	15	738
	2	3	41.0	13	661	25.7	13	368	55.0	14	745
	2	4	63.5	14	636	28.9	13	353	58.4	16	688
Correct rejection	—	1	100.6	16	1207	31.6	15	1121	45.1	19	1223
	—	2	50.3	17	1273	83.8	17	1284	85.7	21	1399
	—	3	116.7	19	1252	72.9	20	1240	52.2	20	1368
	—	4	80.7	20	1246	97.1	21	1225	78.9	21	1421
Miss	1	1	6.8	3	66	7.4	5	102	—	—	12
	1	2	2.4	4	78	7.1	9	152	—	—	12
	1	3	5.3	6	92	7.9	10	173	—	—	8
	1	4	3.9	6	97	11.9	11	202	—	—	14
	2	1	—	—	33	10.6	5	87	—	—	4
	2	2	—	—	52	7.7	9	157	—	—	4
	2	3	5.2	3	75	16.3	10	190	—	—	6
	2	4	4.4	5	95	37.5	13	232	—	—	17
False alarm	—	1	4.8	5	98	8.2	3	61	—	—	8
	—	2	3.9	8	134	—	—	44	—	—	12
	—	3	14.4	8	134	.2	1	64	—	—	7
	—	4	4.9	7	111	—	—	49	—	—	2

Note. χ^2 values for Subject 4 can be found with the histograms in Figure 9.
[a] Input and output position are blocked by eights.

responses and word–nonword responses. Figure 14 contains graphs of convolution parameters plotted against output position. These show that 2P items are responded to more quickly than 1P items in the memory task, with the difference being mainly in τ.

Also, 2P words are responded to more quickly than 1P words in the word–nonword task—again, with the difference being mainly in τ. Most of the increase in mean latency with output position is reflected in the change in τ in this latter task. Thus we should not assume that all of the 5 msec/item slope in the memory task comes solely from changes in memory or decision processes.

Summary of Results and Empirical Method

In the first section, standard analyses of results from four experiments were presented; in the second section, an empirical model for latency distributions was developed and the four experiments were then analyzed using the model. In this section, a summary of the experimental results in terms of both standard analyses and the model is presented. This summary takes the form of a series of graphs (see Figure 15) of the various independent variables, as well as the following functional relationships:

1. In general, hits are faster than correct rejections, with the difference being much less than the spread of the distributions.
2. Changes in latency with input and output position are mainly changes in τ in the model.
3. Changes in latency between 1P and 2P items (Experiment 2) are mainly changes in τ in the model.
4. Changes in latency with list length are changes in μ and τ in the model.

We wish to propose that the convolution of normal and exponential distributions (where the fit is reasonable) provides an excellent summary of reaction time distributions for three reasons. First, the parameters μ and τ

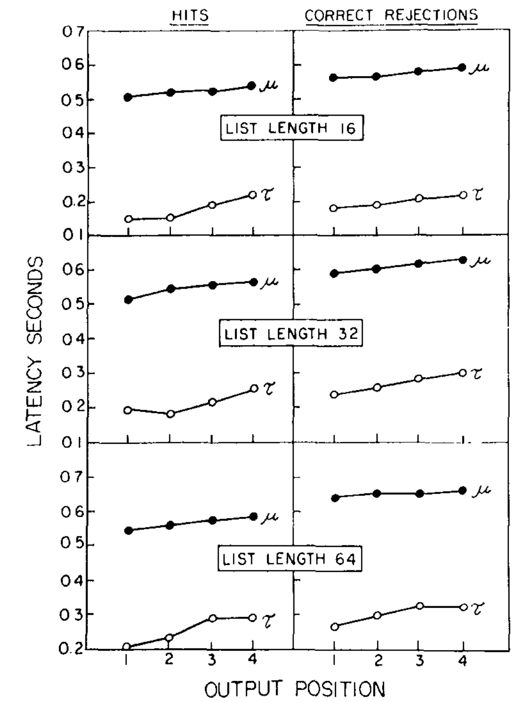

FIGURE 12. Distribution parameters for hits and correct rejections as a function of rate of presentation and output position for Experiment 3. (Note that, as before, output-position blocking is different for the three list lengths—larger blocks for the longer lists.)

mirror properties of distributions that seem important in evaluating models and theories—namely, leading edge (or minimum reaction time) and spread (or variance). For example, a change in μ is consistent with a stage insertion model, and a change in τ (but not μ) is consistent with models that postulate a proportion of responses being slowed. Second, any required properties of the data, such as mean, variance, or mode, can be calculated quickly and cheaply (albeit approximately), and thus rapid assessments can be made. This is particularly easy for moments, for it is well known that the mean and variance of a sum of independently distributed random variables (the distribution function being the convolution of the individual distribution functions) are equal to the sums of the means and variances of the component variables. Third, if a viable model of the processes under consideration is developed, then it is a simple matter to fit the convolution to that model and check whether the convolution parameters from the model behave in the same way as do parameters from the data.

Theories of Recognition Memory Latency

In this section, several theories of recognition memory that deal with latency are examined using the functional relationships summarized in the last section.

In much of the work on recognition-memory latency, it is not often stated explicitly which theoretical processes produce observed differential latency effects. By differential latency effects we mean changes in latency as a function of experimental variables such as input or output position, list length, or rate of presentation. To highlight how the theoretical processes produce such empirical effects, each theory will be analyzed in terms of a simple stage model, as shown in Figure 16. It should be noted that such a model with independent stages is more than likely a gross oversimplification, but on the basis of our criticisms of existing models, it provides a useful framework for comparison of theories.

Conveyor Belt Model

In the conveyor belt model (Murdock, 1974; Murdock & Anderson, 1975), the consecutively presented stimulus items are encoded and stored with their serial order maintained. Upon presentation, the probe is encoded and compared with the most recent item in the list. If the match is not successful, then the next most recent item is examined. This process continues until a match is found, in which case a *yes* response is made, or until the end of the list is reached, whereupon a *no* response is produced. Note that both study and test items are scanned. To incorporate confidence judgments (and very slow responses), it is assumed that the result of the best match to date is stored in a register. If no match has produced a result above a positive response criterion (c_u) leading to a high-confidence *yes* response, then the contents of the register are examined; if the result in the register is below a negative criterion (c_l), then a high-confidence *no* response is made. Otherwise, the criteria are made less strict (c_l raised and c_u lowered) and a second search

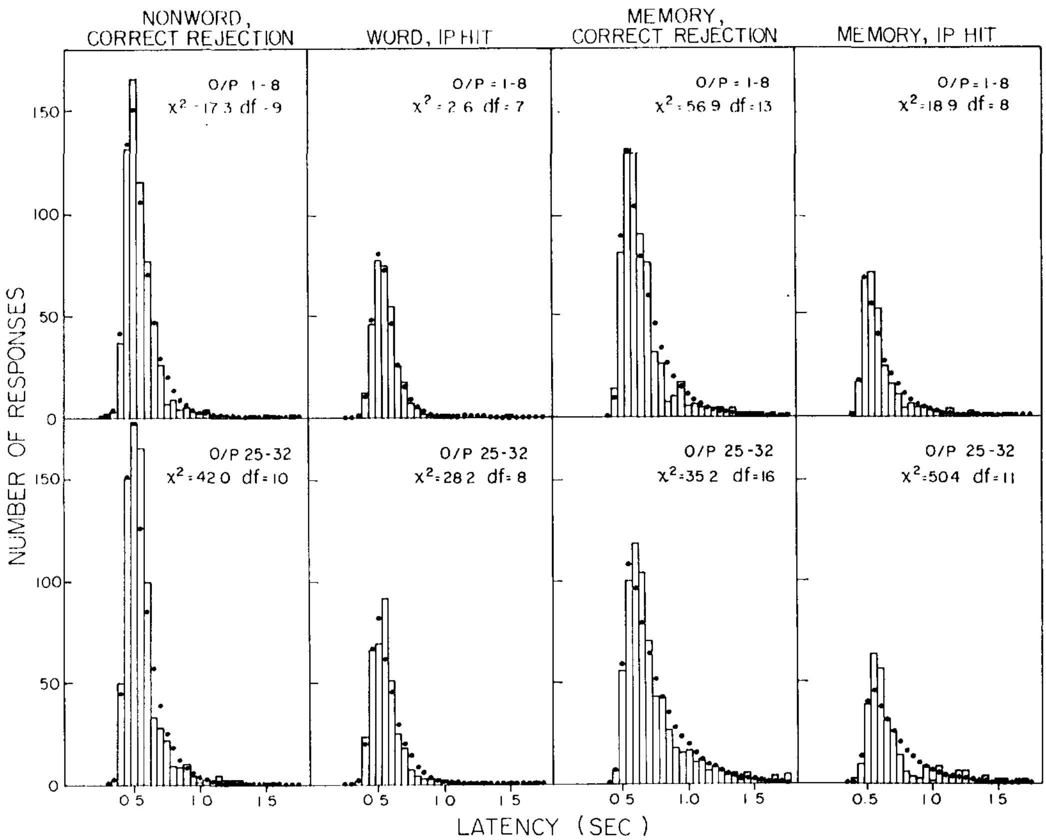

FIGURE 13. A sample of empirical and fitted latency distributions for Subject 3, Experiment 4. (O/P = output position; 1P = presented once.)

takes place, which gives rise to intermediate-confidence responses. Thus, accuracy enters through the matching process: the greater the difference between the stored trace and the probe, the more likely a low-confidence or a *no* response.

In terms of a stage model, all changes in latency with input and output position are a result of the number of times the memory-decision loop is executed (Figure 16). Thus, accuracy and latency effects arise from different stages in the process. The conveyor belt model is successful in dealing with relationships between latency and confidence judgments (Figure 15a), with the linear latency-input and latency-output position functions (Figure 15b), and with the relationship between latency and rate of presentation of the input list (Figure 15d). Also, the conveyor belt model does a good job in dealing with a wide range of other results—for example, judgments of recency (Murdock, 1974), subspan lists (Murdock & Anderson, Note 4), and forced choice data (Murdock & Anderson, 1975). Thus, this model provides a framework that accommodates a large body of results. However, some of the functional relationships shown in Figure 15 do present problems for the model. For example, the model provides no reason why the slopes of the latency-input and latency-output position functions decrease with increasing list length (Figure 15f). Also, the model is not designed to deal with stimulus items presented more than once (Figure 15e), and it is incapable of explaining the primacy effect (Figure 15c). Furthermore, we run into very serious problems when attempting to apply the conveyor belt model to properties of latency distributions, as shown in the following paragraphs.

Fixed rate scanning. In this subclass of models, it is assumed that scanning rate is constant at, for example, about 5 msec per item for 16-item lists. It has been shown (through the functional relationships and in Figure 9) that the fastest responses at all input and output positions are approximately as fast. Consider the latency distributions for lags 5 and 45: The fastest responses at lag 5 must be 200 msec faster than the fastest responses at lag 45 (40 items scanned × 5 msec/item). This contradicts the empirical findings.

Variable rate scanning. To attempt to overcome this contradiction, let us allow the scanning rate to be variable by giving the time for an individual scan a probability distribution, so that some sequences of scans can be very fast and produce rapid responses. A simple (and perhaps extreme) distribution for scanning time is the exponential distribution with a mean of 5 msec. In scanning n items, it is reasonable to assume that as soon as each scan ends, the next begins, so that the resulting probability distribution function is the convolution of n exponential distributions. This results in a gamma distribution with parameter n and rate parameter .2/msec. For lag 5, the gamma distribution has a mean of 25 msec and a standard deviation of 11.2 msec; for lag 45, the gamma distribution is approximately normal, with a mean of 225 msec and a standard deviation of 33.5 msec. Thus, at lag 45, the fastest response (for sample size 500, say) would be 125 msec (225 msec − [3 × 33.5 msec]). Therefore, the variable rate scanning model reduces the difference in leading edges of latency distributions at large and small lags, but not by enough to be consistent with the empirical results.

Thus, we have shown that an analysis of latency distributions poses very serious problems for a class of otherwise acceptable models, namely, serial scanning models.

Strength Theory

Strength theory (Norman & Wickelgren, 1969; Wickelgren & Norman, 1966) was primarily developed to deal with accuracy data and to provide measures that would eliminate response criterion biases. Information in mem-

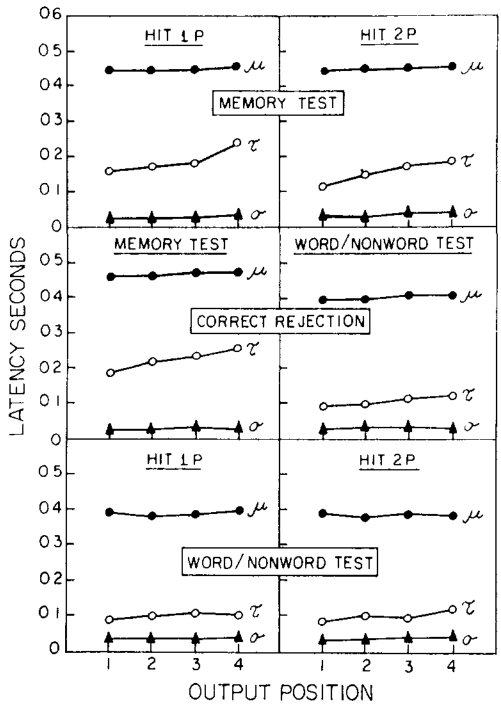

FIGURE 14. Distribution parameters for hits and correct rejections as a function of rate of presentation and output position for Experiment 4. (1P = presented once; 2P = presented twice.)

ory is represented by normal distributions, with old items being stronger than new items on the average. In recognition, the subject has automatic access to the test item and considers only the memory trace strength. This strength is then compared to a criterion, a *yes* response being produced if the criterion is exceeded and a *no* response otherwise. The decision rule comes originally from signal detection theory (see Figure 17). As time passes or as interference builds up, forgetting occurs, and the old item distribution gradually drops back to its initial resting level.

Strength theory can handle many results from simple experiments on recognition memory, from yes/no and confidence judgment procedures to m-alternative forced choice procedures. However, as argued earlier, this is not a strong test of strength theory because the same results can be predicted from a variety of underlying trace distributions (Lockhart & Murdock, 1970) and also from completely different representations of information in memory (Bernbach, 1967; Bower, 1972;

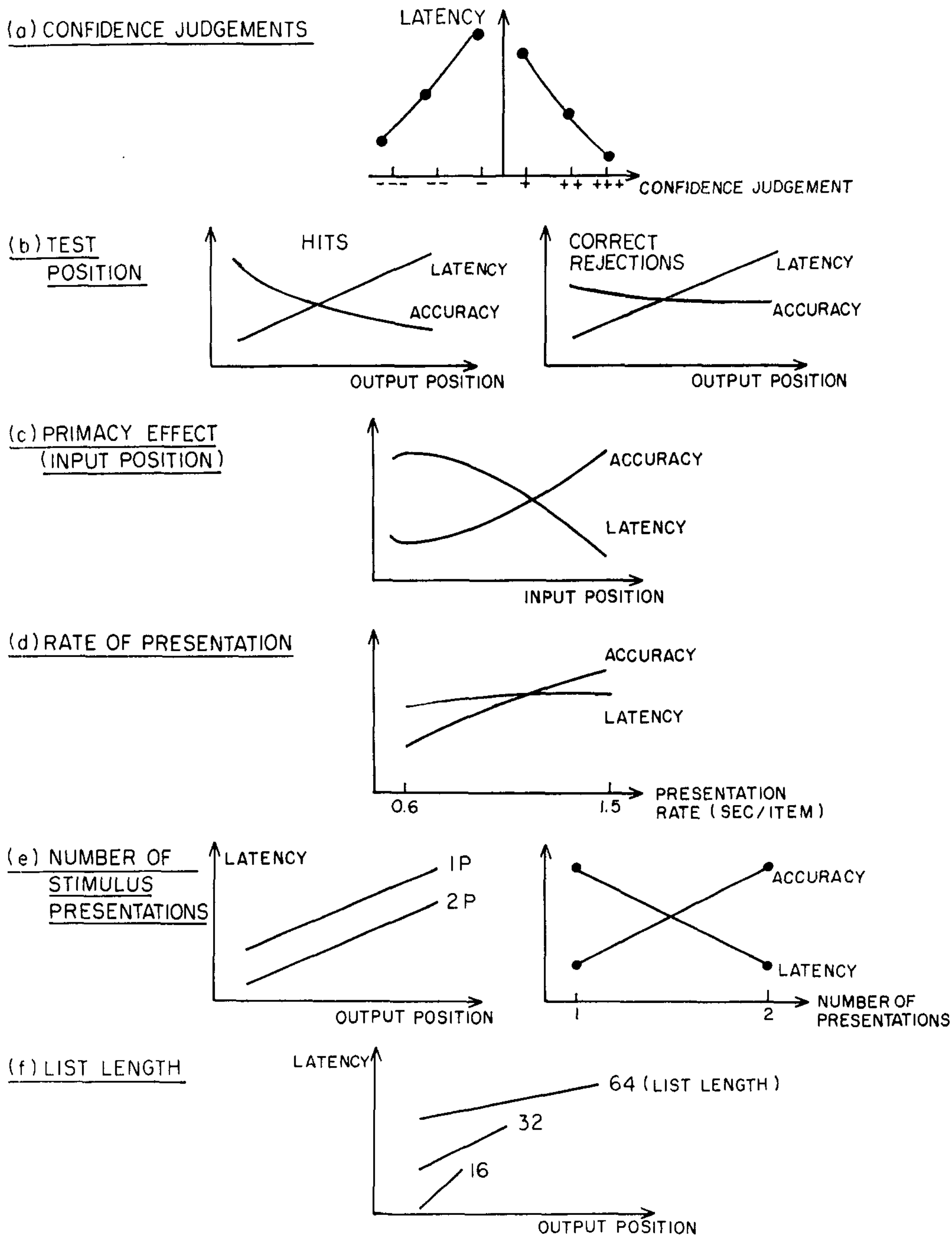

FIGURE 15. Functional relationships derived from Experiments 1–4. (Note that relationships (b) through (f) are for high-confidence responses only.)

Kintsch, 1967). Therefore, it appears that much of the success of strength theory comes from the use of signal detection theory.

Several types of data are not adequately predicted by strength theory. Judgments of frequency and judgments of recency are two examples that have been examined in some detail by Wells (1974). A major problem with strength theory is that it is ahistorical, that is, it is not based on the past history of occur-

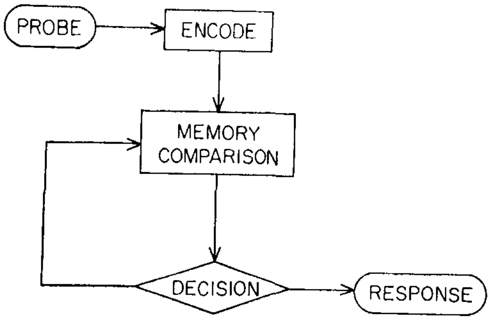

FIGURE 16. A simple stage model indicating the memory-comparison process and decision.

rences of items. Anderson and Bower (1972) raised serious doubts about this theory by showing that memory is capable of much finer discriminations in list identification and discrimination experiments than such a simple strength theory would allow.

A major finding that relates "strength" to latency is the result that latency of response decreases as confidence (which reflects strength) increases. Thus it is assumed that extreme values of strength give rise to short latencies (see Figure 17), and so in terms of our stage model (Figure 16), all differential latency effects result from processes occurring in the decision stage. The assumption that latency depends on the difference between the criterion and the strength of the item causes serious problems when applied to the accuracy and latency-input and latency-output position functions (Figure 15b).

To map from accuracy (strength) to latency for high-confidence hits, an approximately logarithmic transformation is required. (Hits appear to decrease roughly exponentially, and latency is roughly linear with input and output position; see Figure 15b.) If this transformation is applied to the nearly flat accuracy function for high-confidence correct rejections, then a linear latency-output position function with the same slope as for hits is not obtained. Evidence from confidence judgment procedures implies that this transformation should be nearly symmetric (Murdock & Dufty, 1972; Norman & Wickelgren, 1969). This implies that simple strength theory cannot account for the linear latency-output position functions. Even if the theory could account for these functions, it would still run into problems with latency distributions. Suppose, following Murdock and Dufty (1972), that a logarithmic transformation maps strength to latency; it is difficult, if not impossible, to get reasonable looking distributions for both error and correct responses with the same parameter values.

As rate of presentation of study items decreases, then the strength of those items increases, and this is mirrored in accuracy increasing. If the strength–latency relation described above holds (distance from the criterion), then latency should decrease as rate of presentation decreases. Experimentally, however, this is not found to occur (see Figure 15d).

Murdock and Dufty (1972) performed a yes/no recognition memory experiment in order to test the prediction of strength theory that the variability in latency of error responses should be less than the variability in latency of correct responses. This prediction failed—correct responses were less variable in latency than error responses.

Therefore, a simple strength model of recognition latency seems untenable.

Atkinson and Juola Model

It must be stressed from the outset that the Atkinson and Juola model (Atkinson & Juola, 1973; Atkinson, Herrmann, & Wescourt, 1974) was developed to account for results from an experimental paradigm employing prememorized lists and that this paradigm is somewhat different from the study–test paradigm considered so far.

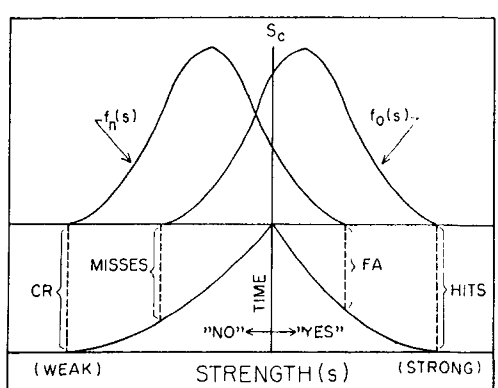

FIGURE 17. The relationship between trace strength and latency in a strength theory.

For response accuracy and latency, this model assumes that the subject makes either a fast response based on the familiarity of the test item or a slower response based on an extended search of memory if the familiarity is neither low nor high. In the best fitting submodel, the latency for a fast response is independent of familiarity, as long as familiarity of the test word is outside the two criteria. Therefore, this submodel cannot deal with latency distributions unless further assumptions are included. In the second best fitting submodel, an exponential function mapping familiarity to latency was used. From this, predictions of the form of the latency distributions can be made, but these turn out to have the same problems as noted for strength theory.

In a footnote, Atkinson et al. (1974, p. 116) state that the latency distributions in their experiments are bimodally distributed, reflecting fast responses and a slow memory search. In the experiments reported here, using the study–test paradigm, all observed high-confidence distributions are unimodal. It seems reasonable to suppose that the lower confidence judgments, which have relatively long response latencies, involve extra processing—such as a slow memory search of the partly learned material. Therefore, high-confidence responses can be assumed to include all fast responses based on familiarity in the Atkinson and Juola model, together with too few slower memory searches to produce bimodality. However, because there is no learned memory set, it is possible that a memory search strategy is not used and that all responses are based on familiarity. If this is so, the version of the Atkinson and Juola model thus derived is isomorphic to the strength model.

In terms of the stage model (Figure 16), there are two memory stores in the Atkinson and Juola model: a conceptual store and an event knowledge store. Information from the conceptual store is fed into the decision system, at which point a fast response is made if familiarity is extreme or else the event knowledge store is searched. Therefore, it can be seen that differential latency effects can be due first, to the temporal properties of the decision system, second, to the relative number of fast responses and memory searches, and third, to the properties of the memory search. Thus, this model has more degrees of freedom than either the conveyor belt model or strength theory.

It seems that Atkinson et al. (1974) would like to have their model apply to lists not prememorized (footnote, p. 113), and so on this basis we shall evaluate the applicability of this model to the study–test paradigm. In the model used by Atkinson et al. (1974), fast responses have constant latency. So if it is assumed that high-confidence responses are fast because of the item's familiarity, then the linear increase in latency with output position for high-confidence responses cannot be explained (Figure 15b; see also Homa & Fish, 1975, for a discussion of lag effects). Following Atkinson et al. (1974, footnote, p. 113), let us suppose that errors can be made in the memory search in dealing with items not prememorized. This would account for errors being slower than correct responses if a greater proportion of errors came from the memory search. In this case, the linear output position-latency function for high-confidence correct rejections could not be predicted at the same time as the relatively flat accuracy function for correct rejections (Figure 15b).

A similar contrast can be found in attempting to explain the primacy effect and the effect of number of stimulus presentations together with the presentation rate effect (Figures 15c, 15e, and 15d, respectively). If only fast responses with constant latency contribute to high-confidence correct responses, then latency should be constant with increasing rate of presentation (familiarity and accuracy decreasing)—but then no primacy effect or decrease in latency with increasing number of stimulus presentations would be expected. If, on the other hand, there are slow memory searches involved, then the primacy effect and the effect of the number of stimulus presentations can be predicted, but constant latency over a range of stimulus presentation rates can not. Thus, the Atkinson and Juola model has problems similar to those of simple strength theory.

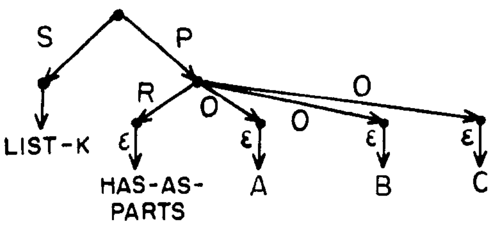

FIGURE 18. The memory representation of the input list, as conceived in the model of Human Associative Memory.

Human Associative Memory

Although Anderson and Bower (1974) have presented a more up to date version of their propositional theory of recognition memory, they did not attempt to account for latency effects in that article; thus the following evaluation deals with their original HAM formulation (Anderson & Bower, 1973). Even this formulation only deals explicitly with prememorized lists, so it is not directly applicable to the study–test paradigm. However, in many of the experiments reported in the present article, the hit rate was 80%–90%, so the study–test paradigm should converge to the prememorized list paradigm.

In HAM, the items in the study list are encoded in propositional form, as shown in Figure 18. For example, presenting the test item C is equivalent to constructing and proffering the probe "List-K, Has-as-parts C?" HAM then attempts an all or none match of that probe to memory. The matching process begins the search simultaneously from the three entry nodes "List-K," "Has-as-parts," and "C." A *no* response is produced if an unsuccessful match from any of the three entry nodes occurs. When the negative probe has not been used previously in the experimental context, instances of it will be linked to other preexperimental propositions. If these are accessed, then immediate falsification occurs. In order to obtain a match from "List-K" or "Has-as-parts," the processes beginning from these nodes have to search the whole object conjunction. If a match is found, then a *yes* response is made. Therefore, this model predicts that *no* responses must take longer than *yes* responses unless encoding decision or response output times are different in the two cases. From "List-K" and "Has-as-parts" there are $5 + 2k$ links to be searched, and from "C" there are 7, k being the memory set size. Thus, the overall search rate parameter for the fastest of a three-way race is $1/(5 + 2k)a + 1/(5 + 2k)a + 1/7a$, where a is the time to search one association. Thus, for a hit response, the mean latency is

$$T_k = K_T + \frac{a}{k}\sum_{i=1}^{k} 7(5 + 2i)/(19 + 2i), \quad (4)$$

and for correct rejections the mean latency is

$$F_k = K_F + a[7(5 + 2k)/(19 + 2k)], \quad (5)$$

where K_F and K_T are constants reflecting encoding, decision, and response output. When Equations 4 and 5 are applied to Atkinson and Juola data, the resulting fits are as good as the fits of the Atkinson and Juola model (Anderson & Bower, 1973, p. 377).

Because of lack of development, the HAM model has several weaknesses when applied to partly learned material. When a node is searched for a particular type of association, then a serial search is carried out on a list of nodes (GET-list), each having the required relationship to the initiating node. The only mechanism for forgetting in HAM is a stop rule associated with depth of search or time of search through the GET-list. Anderson and Bower (1974, p. 408) indicate that this may be inadequate: "However, some of the associations in Figure 2 ["Memory representation of propositions formed upon the appearance of the word dog in List N"] may not have been formed in the time allotted for study in a typical experiment (or may be unavailable at the time of testing)." This implies that sometimes items may not be encoded and sometimes access to a proposition may be lost for reasons other than the association disappearing down the GET-list. Thus, it is probably wise to reserve judgment on the model where criticisms involve these two processes.

In terms of the stage model (Figure 16), all the differential latency effects are localized in the memory stage. Decision appears to be an automatic process occurring along with the search processes.

The HAM model can predict input-position effects if it is assumed that the search from "List-K" and "Has-as-parts" is backward-serial through the object conjunction (Figure 18). This means that Equation 4 has to be modified so that the upper limit of summation is (list length) + 1 − (serial position of the item). Using a value of 46 msec for a (Anderson & Bower, 1973, p. 375), the time to search an association for the Atkinson and Juola data, an input-position effect of 25 − 17 msec/item for input positions 1–16 is found. This effect is not too different from a linear latency/input-position function but is too large by a factor of about 5. If some random entry were allowed, then this figure might be reduced to the 5 msec/item found in the study–test paradigm. If the search processes have exponentially distributed latencies (Anderson, 1974) and if we introduce normal distributions for encoding, decision, and response output, then the response time distribution turns out to be a normal distribution convoluted with an exponential distribution. A very important property of the model formulated this way is that increases in mean latency with input position are reflected in changes in the rate parameter of the exponential alone. Therefore, as was found empirically, the leading edge of the distribution remains in about the same place while mean latency increases.

In both Anderson and Bower (1973, p. 364) and Anderson (1974), it is assumed that the time to traverse, sequentially, n links in a HAM associative structure is exponential with rate parameter $1/na$, where a is the time to search a single link. This is counter to notions of sequential processing (Cox, 1962), as independence is usually assumed and the distribution is gamma with parameter n and rate $1/a$. Anderson (1974) states that the only reason for the assumption of an exponential distribution for the conjunction is mathematical tractability and that this assumption is not critical when dealing with mean reaction time. It seems likely that these two distributional assumptions could be tested using properties of reaction time distributions and that the results of such testing would have implications for our earlier discussion of distributional properties in the HAM formulation of the study–test paradigm.

HAM is also capable of dealing with the following effects: (a) rate of presentation effects—because the time to search the memory structure is independent of encoding time; (b) number of stimulus presentations—because, for an item encoded twice in the list, the process becomes a four-way race with rate increased by $1/7a$, and therefore, twice-presented items should have faster response times than once-presented items; and (c) decrease in slope of the latency-input position function with increasing list length—because if list length increases, there are more elements to search in the list and therefore the relative change in latency with input position is smaller.

Although the HAM model is capable of producing acceptable latency-input position and distributional functions, it is incapable of producing similar output-position functions. For, in this model, the test list is independent of the study list, which means that latency is predicted to be independent of output position. Also, the HAM model has no way of dealing with confidence judgments, for most of the processes in HAM are all or none, with no memory trace strength or equivalent attribute. Thus, it is difficult to see how errors can arise from the processes proposed, but as was noted earlier, extra processes may be added to the model to overcome these difficulties.

Summary of Recognition Memory Latency Theories

Four major models have been examined—namely, simple list scanning, strength theory, a strength and search model, and a propositional model. It seems that each of these has major problems in dealing with latency distributions and with some of the functional relationships shown in Figure 15. These problems are summarized below.

Conveyor belt model. The major problem with this model is that its specific predictions about latency distributions are wrong. Besides this, the model cannot predict the primacy effect, and it is not designed to deal with double presentation of stimulus items. In terms of a stage model (Figure 16), all differential latency effects (input position, output position, and confidence judgment effects) are a result of the number of times the memory-decision loop is executed.

Strength theory. A simple strength model assumes that extreme values of strength give rise to fast responses. As discussed earlier, a model with such a single accuracy-to-latency (i.e., strength-to-latency) mapping can not account for (a) the shape of latency distributions for correct and error responses simultaneously, (b) both high-confidence accuracy and latency results for hits and correct rejections, or (c) rate of presentation effects. In this model, access to memory is automatic, with differential latency effects occurring in the decision stage and no decision-to-memory return loop.

Atkinson and Juola model. In this model, the subject makes a fast response if the familiarity of the item is extreme; otherwise, he makes a slower response based on an extended memory search. As with simple strength theory, high-confidence latency and accuracy results for hits and correct rejections cannot be explained together, and reasonable latency distributions cannot be produced. If the model is adjusted to predict constant latency with rate of presentation, it cannot predict a difference in latency between once- and twice-presented items. In terms of a stage-model representation, there are two separate memory stores. Differential latency effects can enter at the decision stage or in the slow memory search, or they can be related to the ratio of fast responses to slow memory search responses. Therefore, there are many more degrees of freedom in this model than in the other models considered in this section.

Human associative memory. This is perhaps the least well developed model for dealing with latency in a study–test paradigm, but it seems to hold the most promise for accounting for empirical results. It seems that Anderson and Bower (1974, p. 408) wish to include some further mechanism for forgetting, in which case the problems raised here may provide a good test of the adequacy of such an addition. The model predicts input-position effects (serial-position effects) and produces distributions that are of the same shape and vary with input position in the same way as the empirical distributions. Also accounted for are effects on latency of number of stimulus presentations, rate of presentation, and list length. At this stage, the model cannot deal with output-position effects, confidence judgments, or errors, and it predicts that correct rejections should always be significantly slower than hits. In terms of a stage model, all differential latency effects occur in the memory search stage.

Discussion

Attempts to develop models of recognition memory have not utilized the information contained in latency distributions. For any model to be a reasonable approximation to the truth, it should, at least, account for results obtained from considerations of latency distributions. We have shown that a knowledge of distributions is necessary before conclusions can be drawn about serial scanning. We have also warned that even if serial scanning does occur in a paradigm under consideration, there may be other processes going on. Thus, we urge experimenters, at the least, to be aware of how their latency distributions change with the experimental variables, even if no further analysis is carried out. Similarly, we urge theoreticians to look at the distributions predicted by their models in order to avoid developing theories that are contradicted by more rigorous analyses of experimental data.

The method we have proposed uses an explicit probability function to serve as a summary of the empirical data. There are several advantages in using this method. First, the distribution parameters directly represent two properties (namely, minimum response time and spread of the distribution) that seem important in evaluating models and theories. Second, distributional properties such as mean, variance, and mode can be calculated quickly and cheaply. Third, if a theoretical model of the processes is developed, then it is simple and convenient to fit the convolution to that model and compare this fit with the fit of the convolution to the data.

We shall now attempt to bring together the more promising aspects of several of the models discussed earlier and see whether this points the way to development of a model that can deal with latency distributions and with the functional relationships (Figure 15). It seems that most reasonable models of recognition memory employ the decision mechanism of signal detection theory, but the arguments against strength theory suggest that it is likely that more complex processing than simple direct access is occurring, such as memory search processes. There is one important issue we have not considered in this article, and that is the speed/accuracy tradeoff (Pachella, 1974). In the present experiments we believe we have been addressing situations in which forgetting processes affect recognition performance (situations in which both speed and accuracy generally decrease) rather than speed *versus* accuracy situations in which speed is traded off for accuracy.

What can be said about the way in which the stimulus items are stored in memory and the retrieval processes employed? It seems at the moment that there is no clear concensus as to the form of stored information or the types of retrieval processes. Tulving (1972) maintains a distinction between semantic and episodic memory stores, and in such a view the list would be stored in episodic memory. Explicit models for this view have been developed by Murdock (1974) and Kintsch (1974). Anderson and Bower (1974), on the other hand, suggest that episodic memory is simply a tagging of nodes in semantic memory to form some kind of propositional representation. Collins and Quillian (1969), Shiffrin (1970), and Estes (1972) have all developed models that employ the notion of control elements so that items are organized with respect to some superset tag. Rumelhart, Lindsay, and Norman (1972), Norman and Rumelhart (1975), and Schank (1973) have developed important models that all have a propositional memory structure, but these have not been applied to item learning situations, so their suitability cannot be assessed. Foss and Harwood (1975) claim that theories based on association (such as HAM) are fundamentally inadequate and that Gestalt information must be included in any memory model. Smith, Rips, and Shoben (1974) make the even stronger claim that a graph structure is the wrong framework from which to view semantic memory. They see it as more reasonable to suppose that concepts lie in some abstract semantic space and that response times reflect set overlap.

What about retrieval processes? Murdock (1974) has presented the backward-serial scanning model to account for retrieval from episodic memory. Anderson and Bower (1974), Anderson (1974), and Thorndyke and Bower (1974) have all considered possible search processes in sentence memory and have come to the conclusion that simultaneous entry search into a HAM-like structure fits the data best. Foss and Harwood (1975) make a case for conjunctive nodes in which a node acts as a logical AND-gate; that is, the path from the node can only be traversed when the node receives two simultaneous inputs. Smith et al. (1974) use a feature matching version of the Atkinson and Juola model to account for recognition latency in semantic memory, but this is not a necessity for their semantic memory theory.

What general comments can be made about memory structures and search processes? We find the notion of control elements that connect to items in the list, together with a top-down bottom-up search, attractive. If an exponential distribution is used to characterize the search and if the top-down search is serial through the input list, then latency distributions and input-position effects will be consistent with the empirical data. The test list must somehow interact with the memory representation of the list to provide output-position effects for both hits and correct rejections.

The point we wish to make from this discussion is that there is no ready concensus on the form of the memory structure or the form of the retrieval processes used to access information. The data and empirical model presented here provide quite stringent tests of any reasonably explicit model that deals with both accuracy and latency, and we believe these tests should be used by researchers in developing theories of recognition memory.

REFERENCE NOTES

1. Egan, J. P. *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58-51). Bloomington: Indiana University, Hearing and Communication Laboratory, 1958.
2. Sternberg, S. *Estimating the distribution of additive reaction time components.* Paper presented at the meeting of the Psychonomic Society, Niagara Falls, Ontario, Canada, October 1964.
3. Sternberg, S. *Evidence against self-terminating memory search from properties of RT distribu-*

tions. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1973.

4. Murdock, B. B., Jr., & Anderson, R. E. *Retrieval of item information from subspan and supraspan lists.* Paper presented at the meeting of the Psychonomic Society, Boston, November 1974.

REFERENCES

Anderson, J. R. FRAN, a simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 5). New York: Academic Press, 1972.

Anderson, J. R. Retrieval of propositional information from long-term memory. *Cognitive Psychology,* 1974, *6,* 451–474.

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review,* 1972, *79,* 97–123.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: Winston, 1973.

Anderson, J. R., & Bower, G. H. A propositional theory of recognition memory. *Memory and Cognition,* 1974, *2,* 406–412.

Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Potomac, Md.: Erlbaum, 1974.

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press, 1973.

Aubé, M., & Murdock, B. Sensory stores and high-speed scanning. *Memory and Cognition,* 1974, *2,* 27–33.

Bernbach, H. A. Decision processes in memory. *Psychological Review,* 1967, *74,* 462–480.

Bower, G. H. A multicomponent theory of memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1). New York: Academic Press, 1967.

Bower, G. H. Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: Winston, 1972.

Broadbent, D. E. *Decision and stress.* New York: Academic Press, 1971.

Brown, K. M. A quadratically convergent Newton-like method based on Gaussian elimination. *SIAM Journal on Numerical Analysis,* 1969, *6,* 560–569.

Burrows, D., & Murdock, B. B., Jr. Effects of extended practices on high speed scanning. *Journal of Experimental Psychology,* 1969, *82,* 231–237.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 1969, *8,* 240–247.

Cox, D. R. *Renewal theory.* London: Methuen, 1962.

Donders. F. C. [On the speed of mental processes.] In W. G. Koster (Ed. and trans.), *Attention and performance II.* Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica,* 1969, *30;* originally published 1868.)

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2). New York: McGraw-Hill, 1959.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: Winston, 1972.

Foss, D. J., & Harwood, D. A. Memory for sentences: Implications for human associative memory. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 1–16.

Hohle, R. H. Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology,* 1965, *69,* 382–386.

Homa, D., & Fish, R. Recognition reaction time in long-term memory as a function of repetition, lag, and identification of positive and negative search sets. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *104,* 71–80.

Johanson, A. M. The influence of incentive and punishment upon reaction-time. *Archives of Psychology,* 1922, *8,* No. 54.

Kintsch, W. Memory and decision aspects of recognition learning. *Psychological Review,* 1967, *74,* 496–504.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kirsner, K., & Craik, F. I. M. Naming and decision processes in short-term recognition memory. *Journal of Experimental Psychology,* 1971, *88,* 149–157.

Lockhart, R. S., & Murdock, B. B., Jr. Memory and the theory of signal detection. *Psychological Bulletin,* 1970, *74,* 100–109.

Mandler, G., & Boeck, W. J. Retrieval processes in recognition. *Memory and Cognition,* 1974, *2,* 613–615.

McGill, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

McGill, W. J., & Gibbon, J. The general-gamma distribution and reaction times. *Journal of Mathematical Psychology,* 1965, *2,* 1–18.

Meltzer, J., & Shepherd, R. N. Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Potomac, Md.: Erlbaum, 1974.

Murdock, B. B., Jr. *Human memory: Theory and data.* Potomac, Md.: Erlbaum, 1974.

Murdock, B. B., Jr., & Anderson, R. E. Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* Hillsdale, N.J.: Erlbaum, 1975.

Murdock, B. B., Jr., & Dufty, P. O. Strength theory and recognition memory. *Journal of Experimental Psychology,* 1972, *94,* 284–290.

Nelder, J. A., & Mead, R. A SIMPLEX method for function minimization. *Computer Journal,* 1965, *7,* 308–313.

Norman, D. A., & Rumelhart, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

Norman, D. A., & Rumelhart, D. E. *Explorations in cognition.* San Francisco: Freeman, 1975.

Norman, D. A., & Wickelgren, W. A. Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology*, 1969, *6*, 192–208.

Pachella, R. G. An interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Potomac, Md.: Erlbaum, 1974.

Ratcliff, R. A mathematical model for paced serial addition. *Journal of Mathematical Psychology*, in press.

Rumelhart, D. E., Lindsay, P. H., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Schank, R. C. Identification of conceptualizations underlying natural language. In R. C. Schank & K. M. Colby (Eds.), *Computer models of thought and language.* San Francisco: Freeman, 1973.

Shiffrin, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

Smith, E. E., Rips, L. J., & Shoben, E. J. Semantic memory and psychological semantics. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8). New York: Academic Press, 1974.

Snodgrass, J. G., Luce, R. D., & Galanter, E. Some experiments on simple and choice reaction time. *Journal of Experimental Psychology*, 1967, *75*, 1–17.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, *153*, 652–654.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, *57*, 421–457. (a)

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica*, 1969, *30*, 276–315. (b)

Sternberg, S. Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, 1975, *27*, 1–32.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, 1973, *97*, 323–336.

Thorndyke, P. W., & Bower, G. H. Storage and retrieval processes in sentence memory. *Cognitive Psychology*, 1974, *6*, 515–543.

Townsend, J. T. Some results concerning the "identifiability" of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 1972, *25*, 168–199.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, *80*, 352–373.

Watkins, M. J., & Tulving, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, 1975, *104*, 5–29.

Wells, J. E. Strength theory and judgments of recency and frequency. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 378–392.

Wickelgren, W. A., & Norman, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, *3*, 316–347.

Wilkes, S. S. *Mathematical statistics.* New York: Wiley, 1962.

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

(Received June 11, 1975)