A multiple-observations model for response latency and the latencies of correct and incorrect responses in recognition memory

RAY PIKE, LEN DALGLEISH, and JANIS WRIGHT
University of Queensland, St. Lucia, Queensland, Australia 4067

A model for response latency in recognition memory is described which is a strength model incorporating the notion of multiple observations and with the additional assumptions that the variance of the strength distributions increase with set size and that the observer attempts to keep his error rate at a constant level over set size. It is shown that the model can, without recourse to particular parameter values, predict a near linear RT set-size function and, since it is a (TSD) model in its decision aspects, can account for errors and hence error latencies in the recognition task. After the model is described, two experiments are performed which test the prediction that correct mean latency is generally shorter than incorrect mean latency. The prediction is confirmed and this feature is discussed in general, the model being compared with that of Juola, Fischler, Wood, and Atkinson (1971) in this respect. Some possible modifications to the latter model are also considered.

This paper is concerned with two aspects of performance in the item recognition task (e.g., Nickerson, 1966; Sternberg, 1969), the approximate linear relationship between reaction time (RT) and set size, and the relative magnitudes of correct and incorrect mean response latency. A multiple-observations strength model is presented which predicts that the RT set-size function is near linear and that correct mean latencies are shorter than incorrect mean latencies. The term "multiple observations" is used because more than one observation may be made during any one trial when the decision process classifies a probe item as positive or negative. The model involves the usual theory of signal detection (TSD) paradigm (e.g., Egan, 1958; Norman & Wickelgren, 1969), but the final decision is based upon a critical count of observations. For any one set size, the model is identical in form to the multiple-observations model for latency in signal detection described by Pike (1973), but now there are additional assumptions which enable a prediction to be made of near linearity in the RT set-size function.

The term "strength" is, in the present context, not intended to be limited to that of "trace" strength; instead, it is used in a more general sense, in that it may denote the interactive output from a holistic store (e.g., Anderson, 1973; Cavanagh, 1976). The purpose of this description of strength is to avoid the limitations inherent in the trace definition (e.g., Anderson & Bower, 1972; Murdock, 1974) and is of little importance for

the general theory to follow, except insofar as it may justify the assumption concerning the variances of the TSD distributions of strength described below. Since the model makes specific predictions concerning correct and incorrect mean latencies, and since there is considerable ambiguity in the literature concerning these latency features, two experiments will later be described which test these predictions. A description of the model and its properties now follows.

The Recruitment Model for Recognition Memory

As already mentioned, for any given set size, the model is the TSD multiple-observations model as described by Pike (1973). That is, instead of a single observation determining the response, a sequence of observations is assumed to occur. Each observation counts as a "yes" or "no" predecisional response and an overt response only occurs when K such observations have been recruited toward one response or the other. The higher the value of K the more accurate the response but the larger the decision latency. A full description of the consequent features of the model in terms of response probabilities and mean latency is given by Pike (1973), and the relevant latency features are outlined later. For the moment, the primary concern is with the covariation of mean latency and set size, and additional assumptions are necessary. These are: (1) The variances of the strength distributions, both positive and negative, increase with positive set size; (2) the observer attempts to keep his error rate constant with a change in set size.

One way to justify Assumption 1 is to maintain that each probe representation, positive or negative, is interacting with a holistic store (e.g., Anderson,

1973; Cavanagh, 1976) which contains representations of the positive set items only. It can then be further assumed that the output from such interactive processing would, if continuous in the strength measure, be described by the conventional TSD paradigm of two normal distributions with equal variances, and that both of these variances would increase with positive set size. Cavanagh (1976) has already proposed a strength model derived from his concept of holographic storage.

The second assumption is justified by the fact that the data indicate that observers do attempt to keep their error rates constant in item recognition tasks and that the experimental instructions are often framed to help this occur. This assumption is realized in the recruitment model by increasing the value of K with set size. The detail of this latter increase is rather untidy, since for a given probability of being correct there is generally no single integer value of the critical recruitment value, K. It will be assumed, therefore, that the observer uses a combination of K values (i.e., combined between trials) to approximate the required error rate and the simplest case of this will be considered here. This is that the observer uses the two proximate values of K which, in some combination, produce that error rate. That is, those proximate values, K and K + 1, are found such that if the error rates with K and K + 1 are E_K and E_{K+1}, respectively, then $uE_K + (1 - u)E_{K+1}$ = the required rate, $0 \leqslant u \leqslant 1$. The latency is then determined for each K value and the combined mean latency is $uL_K + (1 - u)L_{K+1}$. The theoretical curves presented here are based on this assumption. However, other combinations of K values which also produce the required error rate do not alter the general characteristics of the mean decision time as described below, although the latency distribution variances would alter.

The relationship between the strength distribution variances, S^2, and memory set size, N, which produces an approximate linearity of the RT set-size function is that derived from an additive conception of variance, so that in parameter-free form: $S^2 = N$, with each of the N items contributing equal and independent unit components of variance. This interpretation of variance suits the concept of holistic storage if item representations are assumed to become distorted rather than to decay with time (e.g., through replacement of elements—cf. Bower, 1967) and, consequently, without any overall change in their variance contributions. Note that as an alternative to supposing an increase of variance with set size, it is possible to suppose that the two distributions move closer together, so that d' varies inversely with set size. In this case the necessary relationship is $d' = 1/\sqrt{N}$, but this is not so meaningful as the $S^2 = N$ equation.

The theoretical functions of decision latency with set size for N = 1 to 9, are depicted in Figure 1. The curves were calculated by use of tables of the normal function and the incomplete beta function,[1] the latter function fully describing the recruitment model (Pike,
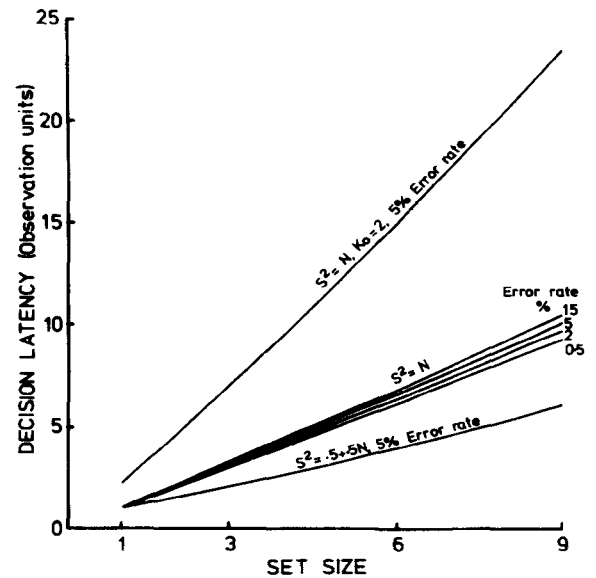

Figure 1. Theoretical RT set-size functions for the recruitment model, for various overall error rates.

1973). The curves are linear to a very good approximation for all reasonable error rates and initial K (K_0) values (i.e., the value of K for N = 1). The curves are almost perfectly linear for a small range of set sizes, only appearing obviously positively accelerated for a large range of set sizes and for the larger error rates. Indeed, the relationship is almost perfectly linear for a very large range of set sizes as the error rate approaches zero. For the case of $K_0 = 1$, the decision time is then approximately given as RT = N, where N denotes both the set size and the number of observations. The virtual linearity of the function for the usual range of set sizes employed in short-term memory experiments is thus a notable feature of the model, a feature which does not depend upon any particular parameter values.

The slope of the RT function increases with the initial K value which may thus be usefully employed as a parameter. Another parameter affecting slope may be introduced by assuming that the decision variances are the more general linear functions of N, $S^2 = a + (1 - a)N$, $0 \leqslant a < 1$, thus dividing the distribution variance into a constant part and a part varying with N. Figure 2 depicts the set-size function for certain values of a: For values of a less than about .6 the curves remain near linear, but as a increases beyond that value a distortion appears in the curve, becoming quite prominent for values of a beyond about .8, especially for small error rates. Thus, a comparative increase in the constant variance component and the consequent initial deformity in the RT function may be one explanation why the function becomes more negatively curved for small set sizes and may decrease in slope with practice (e.g., Corballis, Roldan, & Zbrodoff, 1974, Experiment 3). However, a more general explanation of negative curvature in the RT set-size
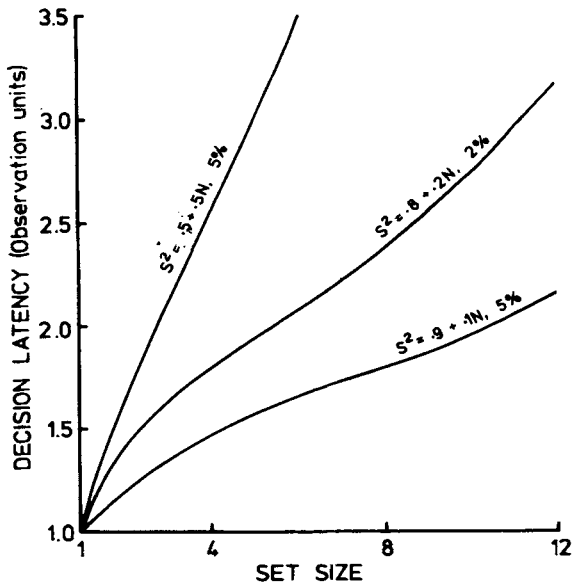
Figure 2. Some examples of the theoretical RT set-size functions for the recruitment model, differing in values of the parameter a and error rates.

function is given later in terms of an increase in error rate.

It is necessary to examine the time scale of events predicted from a recognition memory model, to insure that it appears realistic in terms of neural function. The recruitment model is capable of wide variation in this respect. Thus, assuming the data show a time increase per memory item of T msec, then, for a = 0, observation times may vary from approximately T msec (with K_0 = 1) to considerably smaller values for K_0 > 1. However, as the value of a increases, larger observation times are possible, so that for T = 20, K_0 = 1, and a = .9, an observation time of about 150 msec is realized. In this respect the significance of the parameter a may be seen with regard to the interpretation of long-term memory recognition tasks, with fixed sets. In the usual experiment (e.g., Juola, Fischler, Wood, & Atkinson, 1971), the increase of RT per memory set item may only be in the region of a few milliseconds, but it is unrealistic to suppose that an observation is as short as or less than this: Instead, by the use of a large value of a, more sensible estimates can be made (e.g., for T = 5 msec and a = .9, an observation takes about 40 msec). With the larger memory sets of the long-term situation, the prediction of the model is again that of an approximate linear set-size function, since the initial deformity in the curve would hardly be detectable, and the rate of positive curvature does not appear to increase with set size.

In the case of the short-term fixed-set experiment, data occur where T is of the same order as in the varied-set case (e.g., Sternberg, 1969), but this does not appear to be very general, at least for intercept times of around

400 msec: For example, the data of Okada and Burrows (1974) have T values of around 15 msec and such values, combined with a = .8, say, would give an observation time of about 50 msec. In any case, higher T values can be accommodated by supposing that K_0 > 1. This facility of the model to fit a reasonable time scale to the data is, of course, of no advantage in establishing its validity, depending as it does upon the introduction of the two parameters K_0 and a; but it may be useful in many respects. Thus it may be possible to fit constant observation times for similar situations.

The variance of the decision times based upon the recruitment model are fairly small: For example, in the case of S^2 = N and K_0 = 1, an error rate of 5% and a slope of 35 msec, the variance for Set Sizes 1, 2, 4, and 9 are approximately 0, .15, 1.2, and 3.9 in observation units, respectively, which figures need to be multiplied by the square of the observation times (in this case, $35^2 \doteq 1,200$) to be given in units of milliseconds squared. These assume that each observation takes a constant time interval, but there are two obvious modifications which can be made, neither of which has any noticeable influence upon the mean time features, although they do increase variance. First, the range of K values which combine to produce the required error rate at any set size may be broadened from the two proximate values used in the calculations so far; second, and more important, it is probably unrealistic to assume that each observation time is constant (in the same way as it is unrealistic to assume the constancy of the comparison time in the search models—cf. Sternberg, 1975). It can be assumed, therefore, that the time for each observation has a suitable distribution and as a convenient example the exponential distribution may be chosen, a choice which has met with some favor in the description of hypothetical neural events (e.g., the "timing" model of Luce & Green, 1972). It is particularly convenient in the present case because, as shown by Pike (1973), it is a fairly simple matter to derive the moments of the recruitment model latency distribution, given this assumption. The effect upon the variances is, as expected, to increase them from the values for the case of discrete observations, the figures given above now becoming 1, 2.2, 5.5, and 14 observations, respectively. These values yield predicted variances of the order of magnitude described by Naus, Glucksberg, and Ornstein (1972) and Sternberg (1975), after subtraction of intercept values for extradecisional variances. The values will increase or decrease according to the calculated observation time value or, more basically, the values of K_0 and a.

RT Distribution

A brief note on the form of the recruitment model latency distribution: The form is beta for the case of discrete observations and gamma for the case of exponentially distributed observations (Pike, 1973). Hence, one possibility for the overall RT distribution

is that it is a convolution of the gamma with the normal distribution, a form which should work very well (cf. Ratcliff & Murdock, 1976). However, this conjecture regarding the RT distribution will not be tested in the experiments reported here, as there are quite clearly many different possibilities depending upon the form of the latency distribution for each observation and the residual latency distribution.

Errors and Speed-Accuracy Tradeoff

One of the advantages of strength models is that they deal with the occurrence of error as an instrinsic part of the system, that is, as a statistical decision effect. In contrast, the lack of a theory of errors for scanning models leads to the neglect of error latency analysis, a neglect which is increasingly being questioned with demands for the specification of speed-accuracy tradeoff mechanisms in recognition memory (e.g., Aubé & Murdock, 1974; Wickelgren, 1975). The recruitment model, on the other hand, provides full scope for such specification, as follows.

The main effect of variation in overall error rate (see Figure 1) is to slightly increase the positive curvature, but not so much as would be detectable for small set sizes. The slope is hardly affected (see Figure 1), certainly for the case of $K_0 = 1$, although the slope increases rather more when $K_0 > 1$; for example, for $K_0 = 6$, there is a 10% increase of slope as the error rate changes from 2% to 10%. Variation in overall error rate is, however, a result of sensitivity (d') variation rather than a speed-accuracy tradeoff; it reflects the initial d' measure of the underlying strength distributions (i.e., the value of d' for $N = 1$). Speed-accuracy tradeoff is brought about in the basic recruitment model by variation in the critical K value; the higher the value the more accurate the response, but the more time is taken. Hence, if the tradeoff were based upon the decision process it would be reflected in differences in slope of the RT function, the more accurate performance having a steeper slope. However, there is strong evidence that in the recognition memory task the tradeoff may occur at some level other than the decision one, this evidence being reflected in differences in intercept values (Swanson & Briggs, 1969). In terms of the present model, and assuming the tradeoff occurs at the encoding level, as suggested by Swanson and Briggs (1969), a more efficient encoding would take a longer time but result in an increased d' for the strength distributions. The slopes would then be virtually unaltered since, as mentioned above, overall error rates hardly affect the slope in the model.

There is also a variation in error rate which has a tradeoff effect, but which may be dictated by the limitations of the processing system, when attempting to cope with the higher K values necessary to maintain the error rate for larger set sizes. If the system fails in this respect, consequently operating on a lower K value than that required to maintain the error rate,

then an increase of errors will occur but at a shorter overall mean time than would otherwise be the case. The general effect may be to bring about marked nonlinearity in the RT function, although the amount of curvature would depend upon the form of the increase of error rate. The experimental results are indeed equivocal in this regard; although some data do not show any curvature with apparent increase of error rate (e.g., Burrows & Okada, 1974, Experiment 2; Corballis & Miller, 1973; Wingfield & Branca, 1970), other data do so (e.g., Aubé & Murdock, 1974; Banks & Atkinson, 1974; Corballis et al., 1974, Experiment 2; Swanson & Briggs, 1969).

One important aspect of error performance is the amount and direction of the difference between correct and incorrect mean latencies. The recruitment model makes the general prediction that mean error latencies are longer than mean correct latencies (cf. Pike, 1973). However, with the value of K varying between trials, this prediction may be reversed due to the combination of differing speed-accuracy conditions (i.e., because the larger K values give more accurate performance but longer latencies), and this is precisely what is being proposed for the recruitment model as a means of keeping a constant error rate. Consequently, some reversals of the general prediction can be expected, especially for the lowest set sizes, where, because of the lower K values, the variation in K has a greater effect of this kind, there being no reason why variation in K should increase with set size.

As already mentioned, the search models for recognition memory are generally formulated in terms of error-free performance, and it has been proposed by Sternberg (1969) that it is necessary to study such performance. The literature suggests, however, that subjects always make more or less errors and that, consequently, the notion of error-free performance is only an ideal which is rarely obtained in recognition memory experiments. Earlier, the work of Briggs and his co-workers (e.g., Swanson & Briggs, 1969) and of Norman and Wickelgren (1969) was very concerned with error performance, particularly as it is related to the speed-accuracy tradeoff (e.g., Swanson & Briggs, 1969), and currently, a renewed interest in this aspect of performance is beginning to appear (e.g., Aubé & Murdock, 1974; Reed, 1976). The scanning model of Murdock (1974) takes account of errors by supposing that multiple scans occur with changing criteria. The prediction of error latencies has not yet been fully developed for the model, but it would appear that the error rate must increase with the number of "observations" (i.e., scans) that have to be taken. Consequently, incorrect mean latencies should be longer than correct mean latencies, as in the recruitment model.

The current research findings on the correct-incorrect mean latency difference in recognition memory are by no means clear. Error latencies have been discussed

largely in connection with the Juola et al. (1971) model. The prediction of that model is that error latencies, being due to fast "familiarity" responses, will be shorter than correct response latencies and should not increase with memory set size (Banks & Atkinson, 1974; Juola et al., 1971), although there has been little detailed analysis of error latency (e.g., Juola et al., 1971). The data of Banks and Atkinson (1974) are the best presented in this regard, and there the prediction of the model appears to be fulfilled except that error latencies increase with set size. However, error responses in that study arise in a condition where there is an extreme bias to speed performance, one which is ideal for producing fast, inaccurate responses, so whether these fast responses are based on strength of "familiarity" or are fast guessing responses is open to question. The predictions from the Juola et al. (1971) model are not borne out in the studies of Banks and Fariello (1974), Biederman and Stacy (1974), Homa and Fish (1974), and Murdock and Dufty (1972). The difficulty in interpreting the data may be illustrated by the Biederman and Stacy (1974) study where correct mean times were less than incorrect mean times in 17 out of 26 cases, but where there was a strong bias to respond "No," a condition which can, according to the recruitment model, artifactually induce faster error mean times (see Pike, 1973). The data of Egeth and Smith (1967) also demonstrated error latencies decidedly faster than correct response latencies in a fixed-set task, but again there was a bias to fast responding (brought about by a monetary reward), with the probable consequence of greater inaccuracy for the faster responses.

The experiments reported here sought additional information on the issue of correct and incorrect mean latency when subjects are not biased to fast responding. The experiments are, therefore, largely a comparison between the Juola et al. (1971) model and the recruitment model in this respect. In order to obtain a reasonable error rate and hence reliable error mean times, geometrical forms were used as the memory items in the first experiment. In the second experiment, more conventional items were used, but with a fairly large total ensemble.

Before describing the experiments, a word is necessary on the artifacts that may arise in the assessment of correct-incorrect mean latency differences. These artifacts arise from heterogeneity in the performance of the individual subjects. There are two possibilities: Either the less accurate performance will yield trials with faster latencies, including some fast guess responses; or latencies will be slower owing to an overall lessening of efficiency. In the first case, the averaging of responses may produce faster incorrect mean times, as was suggested to be the case for some studies discussed above. In the second case, slower incorrect mean times may be produced. The second case has been documented in studies of recognition memory by Homa and Fish (1974) and Juola et al.

(1971). Some attempt was made in the second experiment to control for these possible artifacts, since they can give rise to results biased in favor of one model or the other.

There are, of course, ways in which the occurrence of errors in the data may interfere with the analysis of correct responses, mainly by the inclusion within the correct responses of those which are due to a guessing process. Since the following experiments are concerned with differences in magnitude between correct and incorrect responses, such contamination of the correct responses is not a crucial concern, its occurrence being conservative with respect to tests of differences between the latencies of the two response categories.

EXPERIMENT 1

Method

Subjects. Subjects were 18 first-year undergraduate students at the University of Queensland, between the ages of 18 and 26, who participated as a course requirement.

Stimuli and apparatus. Memory items were outlined geometrical forms, a few of which would be classed as familiar (circle, square, etc.), but most of which were generally unfamiliar. The forms were regular and not random; they were easily distinguishable, but chosen so as not to make any particular item too distinctive. Each memory set of six items was contained on one slide and projected onto a glass screen by means of a Kodak projector. The size of each item when projected was such as to be contained in a 7.5-cm square with 2.5 cm between the squares. The memory set items were presented as two rows of three items. The subject sat at a distance of about 75 cm from the screen. Two response buttons in front of the subject were labeled "Yes" and "No." These labels were interchanged between subjects.

Procedure. All subjects were given 144 experimental trials in the varied-set condition and 532 trials in the fixed-set condition (it was anticipated that the error rate would be less in the latter condition). In the fixed-set condition, there were three distinct target sets, the 18 positive stimuli being chosen from a complete set of 36, leaving 18 negative items. In the varied-set condition, each set was randomly chosen from the 18 positive items (i.e., the positive sets were "consistent," Kristoffersen, 1972). Subjects were allocated to the two orders of the two conditions (fixed and varied set) in equal numbers. Particular care was taken in this experiment to stress the importance of accuracy in performance. Instructions emphasized that speed in response was important but not so much as accuracy. Subjects received three sessions of 144 trials in the fixed-set condition and one session of 144 trials in the varied-set condition.

Results and Discussion

Table 1 gives the mean times and response probabilities for the two experimental conditions. A 2 by 4 repeated measures analysis of variance (conditions by response category) yielded significant F ratios for the conditions factor [F(1,17) = 25.9, MSe = 123,400, p < .001] and response category [F(3,51) = 18.9, MSe = 18,600, p < .001], but no significant interaction. A significant contrast occurred with correct vs. incorrect (i.e., hits + correct rejections vs. misses + false alarms) [F(1,51) = 50.7, MSe = 18,600, p < .001].

Table 1
Mean Latencies (ML) and Probabilities (p) for the Four
Response Categories in Experiment 1, With a
Set Size of Six

	Hit	Miss	False Alarm	Correct Rejection
Varied Set				
ML (msec)	1031	1226	1130	986
p	.782	.218	.168	.832
Fixed Set				
ML (msec)	714	944	834	733
p	.952	.048	.029	.971

Note—Values of p were calculated by combining number of trials over all subjects.

Individual differences in the direction of the correct-incorrect latency difference were as follows. In the fixed-set results, 7 of the 18 subjects had smaller mean incorrect times in the case of negative probes; 2 of the 18 with positive probes. In the varied-set condition, this result only occurred twice in the whole of the 36 cases. Within the context of the recruitment model, this difference between the two conditions may indicate that between-trials variation in critical count values is stronger in the fixed-set case.

The results of this simple experiment confirm the prediction of the recruitment model in terms of the latency difference and are contrary to the prediction of the Juola et al. (1971) model, as it is usually conceived. However, in a more elaborate version (Atkinson & Juola, 1973), it is proposed that the latency of the fast familiarity responses should vary inversely with the distance of an observation from the detection criterion, and this would result in correct latencies being shorter than incorrect latencies. Whether this modification could result in the larger overall mean incorrect times presents a problem, since the exhaustive scan process would then have to be comparatively rapid, which would question the necessity for "familiarity" responding. In any case, the correct mean latencies should nevertheless increase with set size (since they are, of course, subject to the scan processing) and eventually become larger than incorrect latencies. This trend should be apparent in the second experiment which contains a range of set sizes.

An alternative explanation is that serial scanning is not occurring in the present task because of the difficulty in representing geometrical forms in a serial list and that some form of strength interpretation, based on features or attributes, is then valid. Whether it is then conceivable that there could be a scanning process for familiar items (e.g., letters or digits) and a strength process for geometrical forms is another matter.

EXPERIMENT 2

Method

Subjects. Six first-year psychology students (four females,

two males) took part in the experiment. Each subject completed a total of 10 h, 3 h as a course requirement and 7 h paid at the rate of $2/h.

Apparatus. Stimuli were presented on a Burroughs Self-Scan Panel Display Unit controlled by an Interdata 5 computer. The stimulus ensemble consisted of 52 different Teletype characters, including 25 letters, 9 digits, and 18 ASCII characters. Responses were made by pressing either of the two buttons labeled "Yes" and "No," and were recorded automatically, including the response time, on paper tape.

Procedure. Memory set items were chosen randomly from the stimulus ensemble, but the sequence of trials was preprogrammed. The experiment was conducted as a 2 by 4 repeated measures design, with two levels of presentation (fixed and varied) and four levels of set size (2, 4, 7, and 11). The memory set items were presented simultaneously, in a horizontal row. Each subject was presented with a total of 40 series of 60 trials, 20 varied set and 20 fixed set (i.e., there were five series for each set size at each condition). The trials were completed, in most cases, over five experimental sessions, each session containing about eight series, there being some variations between subjects. Trial series, each consisting of 60 trials, were arranged in blocks of four with each such block being either a fixed- or varied-set condition and containing each set size in an order which was balanced across subjects. Fixed- and varied-set blocks were alternated, three subjects commencing with fixed and three with variable condition.

Each series of 60 trials for one set size was preceded by 12 practice trials. In the fixed-set condition, the same set of memory items was tested in each of the 60 trials and the set was re-presented every 12 trials. The fixed memory sets were displayed for a time determined by the set size, at .5 sec per item at the first and subsequent presentations. Varied memory sets were presented at .5 sec per item on each trial. Under both conditions a constant 2-sec interval separated the stimulus set from the following probe stimulus, which remained on until the subject responded or until 5 sec elapsed. The intertrial interval was 3.5 sec in the varied-set case. Positive and negative probes occurred equally often and in a "random" arrangement. Each serial position was tested equally often. In one half of the trials for each subject, the right hand signaled "Yes," and in one half "No," with practice for both conditions. Instructions again emphasized accuracy rather than speed of performance.

Results and Discussion

A three-way analysis of variance of the mean response times yielded significant F ratios for fixed vs. varied [$F(1,5) = 34.64$, MSe = 155,700, $p < .01$], set size [$F(3,15) = 12.64$, MSe = 150,400, $p < .001$], and response category: hit, miss, false alarm, or correct rejection [$F(3,15) = 16.1$, MSe = 49,800, $p < .001$]. No interaction was significant. The only significant contrast within the response category factor was that of correct vs. incorrect [H + C vs. M + F; $F(1,15) = 42.38$, MSe = 49,800, $p < .001$], this contrast being significant at both levels of the presentation condition. Table 2 gives the mean time and probabilities for each response category at each set-size level.

Analyses on the error proportions show that there are significant differences between conditions [$F(1,5) = 9.8$, MSe = 62.137, $p < .05$] and between set sizes [$F(3,15) = 16.9$, MSe = 155.63, $p < .001$], the latter result confirming the increase of error rate with set size. Similarly, for the error latency, for conditions [$F(1,4) = 55.02$, MSe = 40,783, $p < .01$], and for set size [$F(3,12) = 6.68$, MSe = 83,438, $p < .01$]

Table 2
Mean Latencies (ML) and Probabilities (p) for the Four
Response Categories in Experiment 2

Set Size		Hit	Miss	False Alarm	Correct Rejection
		Fixed Set			
2	ML	500 (492)	520	639	685 (600)
	p	.975	.025	.033	.967
4	ML	548 (581)	558	954	735 (667)
	p	.962	.038	.051	.949
7	ML	787 (700)	964	1051	791 (761)
	p	.937	.063	.047	.953
11	ML	845 (810)	1048	1029	968 (888)
	p	.854	.146	.071	.929
		Varied Set			
2	ML	704	1059	1122	809
	p	.961	.039	.027	.973
4	ML	838	1155	1153	912
	p	.935	.065	.046	.954
7	ML	1031	1482	1471	1149
	p	.893	.107	.113	.887
11	ML	1083	1354	1296	1255
	p	.805	.195	.214	.786

Note—Values of p were calculated by combining number of trials over all subjects. Figures shown are adjusted figures for fixed-set correct results; unadjusted figures are in parentheses.
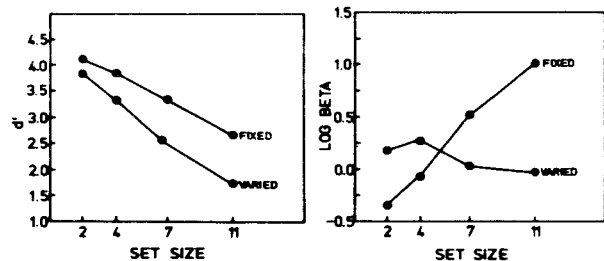

Figure 3. Sensitivity (d') and bias (log beta) variation as a function of set size in Experiment 2.

(one subject had no errors on Set Size 2 and his data were omitted from the analysis). Figure 3 gives the d' and β values for the four set sizes.

The difference between correct and incorrect latencies in the fixed-set case was further assessed by comparing the error latencies with those correct latencies in their immediate vicinity within the sequence of trials, as follows. For each error trial, the immediately preceding and following correct response was ignored, but the two correct response latencies which otherwise immediately preceded and followed the error response were recorded. This was done for correct responses to the same stimulus as the incorrect response and for correct responses which were the same category of response as the error response. The means of the correct response latencies were then taken as the ones to be compared with the mean incorrect latencies. The procedure enabled a comparison of correct and incorrect

mean times with an adjustment for fluctuations in speed and accuracy of performance over sessions or the experiment as a whole (cf. Homa & Fish, 1974; Juola et al., 1971). The results of this analysis in terms of means are displayed in Figure 4, where the adjusted mean times for the case of the same stimulus and the same response are combined (it turned out that there were no appreciable differences between these cases).

In the varied-set case the mean times were also calculated after eliminating the first series at each set size, again in case of initial impairment in performance. Results of this analysis will not be given as the mean times are, to all intents and purposes, unaltered.

GENERAL DISCUSSION

It should first be mentioned that because the ensemble of memory items consisted of three different categories (ASCII characters, letters, and digits) the reaction times may well have been affected, probably reduced (cf. Homa, 1973; Naus, 1974; Okada &· Burrows, 1973). An overall reduction is not important for our present investigation, but if there are differential latency and accuracy effects between the three categories, then artifactual differences between the overall correct and incorrect mean latencies may be present. The data of individual subjects were examined in this respect and there are, indeed, some differential effects present, particularly in the fixed-set data. However, the effects are too small to produce such marked changes in the overall mean times for correct and incorrect responses for artifactual results to occur and, moreover, the differential effects are of a mixed variety, so that the more accurate category of stimulus item sometimes produces longer latencies and sometimes shorter latencies. In general, there are considerable individual differences in the category effect.
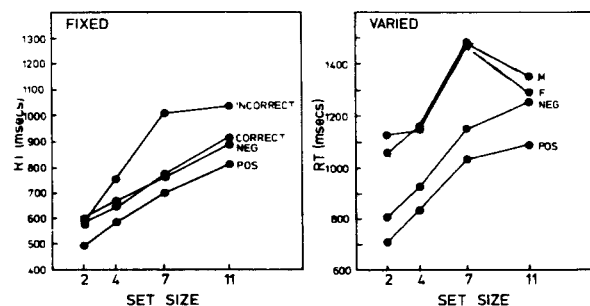

Figure 4. Mean latency as a function of set size in Experiment 2. (a) Fixed-set case: Here the unadjusted positive and negative correct response times are shown, together with the combined adjusted correct mean times and the combined incorrect mean times. Note that the extent of the adjustment may be seen from the fact that the combined unadjusted mean correct times should be the average of the positive and negative cases. (b) Varied-set case, with the four detection categories of response (F = false positive, M = miss).

On the whole, it is again shown that mean correct times tend to be shorter than mean incorrect times in the present task. The adjustments made in the fixed-set case have indeed brought about some increase in the calculated correct times, but not so much as to cast doubt on this conclusion. As for the varied-set case, there does not appear to be any marked practice effect, since the mean times remain approximately the same when initial trial blocks are removed from the analysis. The linear relationship between correct RT and set size is well shown in the adjusted and unadjusted fixed-set data, with positive and negative functions parallel in the latter case and with negative RT greater than positive (Figure 4), so that the task appears to produce results which conform to the usual recognition memory pattern.

The fact that the obtained incorrect mean latencies are, in general, decidedly longer than correct mean latencies is probably due to the overall longer times taken for responding by our subjects than is usually the case in recognition memory tasks. The instructional emphasis on accuracy would appear to have resulted in the avoidance of much responding of the faster guessing type, thus, reducing heterogeneity of performance, in that direction at least.

The increase of errors with set size (Table 2 and the decreasing d' of Figure 3) does not seem to have affected the linear relationship in the fixed-set case, but may be the cause of the negative curvature of the variable-set case, such data indicating that curvature is related to inconstancy of error rates as suggested by the recruitment model. However, a large part of the nonlinearity is due to the Set Size 11 latencies, and this set size could be regarded as being too large for serial-scan processing, in the present instance at least. The increase of error rate for the fixed-set case, unaccompanied by any apparent departure from linearity in the RT functions is, clearly, not a favorable result for the recruitment model because of the relationship between the assumption that the observer attempts to keep a constant error rate, and the predicted near linearity. As mentioned earlier, other experimental results are ambiguous in this regard, and our results have not improved that situation. However, the individual fixed-set RT functions are not nearly so tidy, in terms of linearity, and it could be argued that, within the context of the recruitment model, when an observer fails to keep a constant error rate over set size, there is no guarantee that the consequent RT function will have a well defined positive or negative curvature; it may be quite irregular, with the result that the averaged function may still indicate considerable linearity. For the moment, we will rest our case upon that argument, but it is realized that this is an important point which needs further investigation.

The crossover effect obtained in the β values (Figure 3) presumably reflects differences in subject strategy in the two types of task, subjects having a tendency to be increasingly biased to "Yes" in the variable-set case, as set size increases, and a more marked tendency in the opposite direction in the case of fixed sets. If this crossover effect can be replicated in other studies, it may require explanation by theories of processing, these data thus indicating a further way in which the study of errors can be important in memory tasks. There is a similar decrease of β with set size in the variable-set experiment of Banks and Atkinson (1974).

The model described in this paper is reasonably well supported by the results of the present experiments, insofar as the differences between the mean latency for correct and incorrect responses are concerned. Memory recognition tasks similar to that of the single probe case also produce larger incorrect mean latencies. Thus, Murdock and Dufty (1972), using their study and test list procedure, found that correct mean latencies for individual subjects are smaller in 13 out of 16 cases. In the continuous memory task (e.g., Hintzmann, 1969; Okada, 1971), larger error latencies appear to occur. The latter task is clearly one in which the strength explanation is superior to that of serial scanning: Okada (1971) has demonstrated and argued convincingly for a trace strength process rather than a scan process to explain the results in this task, although an alternative theoretical description of processing for the task has been presented by Juola, Taylor, and Young (1974), using the Juola et al. (1971) combined strength and scan model. In the task where the subject is asked to recall adjacent items in the memory list, error latencies are, apparently, larger than correct latencies according to the results of Sternberg (1969, Experiment 6).

We conclude, then, that the recruitment model shows signs of being able to describe the extant data on mean latencies of correct and incorrect responses better than the Juola et al. (1971) model. One possible modification of the latter model has already been discussed, but there is little evidence of the prediction from that modification (correct mean latency becoming larger than incorrect mean latency with increase of set size) in the present data.

Another assumption which would enable the Juola et al. (1971) model to explain larger incorrect means is that the so-called familiarity responses only occur when the subject is under speed pressure, and that, otherwise, an exhaustive scan will occur. However, it is then necessary to conceive a viable error theory for the exhaustive-scan process, and this may not be easy. For example, one approach is to assume that miss responses occur because of omissions in the memory list and false alarm responses occur as a result of intrusions in the list. However, this predicts that miss responses would generally be shorter than correct mean latencies, whereas false alarms would be longer. This pattern of mean latency results occurs in the Biederman and Stacy (1974) data, but, as mentioned above, that

experiment involved a task biased toward "no" responding, and in this case such a pattern is also predictable by the recruitment model (Pike, 1973). The present data also support any multiple-scan model (cf. Murdock, 1974) which assumes that the criteria for attainment of a match between memory and probe representations becomes less stringent with each additional scan. The Sternberg (1969) exhaustive-scan model could, presumably, be modified in this way. Only one point will be mentioned in this regard. The time scale of events is already, we feel, less realistic within the serial-scan models than within the recruitment model. It would be even less realistic if the multiple-scan concept is employed, since the greater the number of scans the smaller the scan rate for individual items within each scan. The problem that already exists in the time scale, when interpreting many experimental results in terms of serial-scan models, would clearly be exacerbated (cf. Anderson, 1973; Murdock & Anderson, 1975).

Finally, it is to be hoped that the emphasis given here to incorrect response times and their relationship to mean correct response times will encourage analysis of error data in the recognition memory task to a greater extent than previously. For, although the serial-scan models have stimulated considerable research, it is unlikely that a complete evaluation of the models for recognition memory can be made until a satisfactory theory of error performance is demonstrated. The study of Shevell and Atkinson (1974) is a start in this direction, but one which does not deal explicitly with error latency.

REFERENCES

ANDERSON, J. A. A theory for the recognition of items from short memorized lists. *Psychological Review*, 1973, **80**, 417-438.

ANDERSON, J. R., & BOWER, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97-123.

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

BANKS, W. P., & ATKINSON, R. C. Accuracy and speed strategies in scanning active memory. *Memory & Cognition*, 1974, **2**, 629-636.

BANKS, W. P., & FARIELLO, G. R. Memory load and latency in recognition of pictures. *Memory & Cognition*, 1974, **2**, 144-148.

BIEDERMAN, I., & STACY, E. W. Stimulus probability and stimulus set size in memory scanning. *Journal of Experimental Psychology*, 1974, **102**, 1100-1107.

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

BURROWS, D., & OKADA, R. Scanning temporarily structured lists: Evidence for dual retrieval processes. *Memory & Cognition*, 1974, **2**, 441-446.

CAVANAGH, P. Holographic and trace strength models of rehearsal effects in the item recognition task. *Memory & Cognition*, 1976, **4**, 186-199.

CORBALLIS, M. C., & MILLER, A. Scanning and decision processes in recognition memory. *Journal of Experimental Psychology*, 1973, **98**, 379-386.

CORBALLIS, M. C., ROLDAN, C. E., & ZBRODOFF, J. Response set effects in recognition memory. *Memory & Cognition*, 1974, **2**, 501-508.

EGAN, J. P. *Recognition memory and the operating characteristics*. Tech. Note AFCRC-TN-58-51. Bloomington: Indiana University Hearing and Communication Laboratory, 1958.

EGETH, H., & SMITH, E. E. On the nature of errors in a choice reaction task. *Psychonomic Science*, 1967, **8**, 345-346.

HINTZMANN, D. L. Recognition time: Effects of recency, frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **79**, 192-194.

HOMA, D. Organization and long-term memory search. *Memory & Cognition*, 1973, **1**, 369-379.

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*, 1974, **104**, 71-80.

JUOLA, J. F., FISCHLER, I., WOOD, C. T., & ATKINSON, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, **10**, 8-14.

JUOLA, J. F., TAYLOR, G. A., & YOUNG, M. E. Stimulus encoding and decision processes in recognition memory. *Journal of Experimental Psychology*, 1974, **102**, 1108-1115.

KRISTOFFERSON, M. W. When item recognition and visual search functions are similar. *Perception & Psychophysics*, 1972, **12**, 349-384.

LUCE, R., & GREEN, D. M. A theory of psychophysics and response times based on inter-arrival times of "neural" pulses. *Psychological Review*, 1972, **79**, 14-57.

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

MURDOCK, B. B., & 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., & DUFTY, P. O. Strength theory and recognition memory. *Journal of Experimental Psychology*, 1972, **94**, 284-290.

NAUS, M. J. Memory search of categorized lists: A consideration of alternative self-terminating categories. *Journal of Experimental Psychology*, 1974, **102**, 992-1000.

NAUS, M. J., GLUCKSBERG, S., & ORNSTEIN, P. Taxonomic word categories and memory search. *Cognitive Psychology*, 1972, **3**, 643-654.

NICKERSON, R. S. Response times with a memory dependent decision task. *Journal of Experimental Psychology*, 1966, **72**, 761-769.

NORMAN, D. A., & WICKELGREN, W. A. Strength theory and decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, 1969, **6**, 192-208.

OKADA, R. Decision latencies in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 27-32.

OKADA, R., & BURROWS, D. Organizational factors in high-speed scanning. *Journal of Experimental Psychology*, 1973, **101**, 77-81.

OKADA, R., & BURROWS, D. Divided attention and high-speed memory search. *Journal of Experimental Psychology*, 1974, **103**, 191-195.

PIKE, R. Response latency models for signal detection. *Psychological Review*, 1973, **80**, 53-68.

RATCLIFF, R., & MURDOCK, B. B. Retrieval processes in

recognition memory. *Psychological Review*, 1976, **83**, 190-214.

REED, A. V. List length and the time course of recognition in immediate memory. *Memory & Cognition*, 1976, **4**, 16-30.

SHEVELL, S. K., & ATKINSON, R. C. A theoretical comparison of list scanning models. *Journal of Mathematical Psychology*, 1974, **11**, 79-106.

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421-457.

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

SWANSON, J. M., & BRIGGS, G. E. Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, 1969, **81**, 223-229.

WICKELGREN, W. A. Dynamics of retrieval. In D. Deutsch & J. A. DEUTSCH (EDS.), *Short-term memory*. New York: Academic Press, 1975.

WINGFIELD, A., & BRANCA, A. A. Strategy in high-speed memory scanning. *Journal of Experimental Psychology*, 1970, **83**, 63-67.

NOTE

1. For a detailed description of the mathematical expressions involved in the recruitment model, see Pike (1973). Briefly, for the response with observation probability p and critical recruitment value K, the probability of response as against the alternative response with the same critical count and observation probability $q = 1 - p$ is $P = I_p(K,K)$, where $I_x(p,q)$ is the standard notation for the incomplete beta function ratio. Similarly, the mean time is given by:

$$L = \frac{K}{p} \frac{I_p(K+1, K)}{I_p(K, K)}.$$

To obtain the functions of Figure 1, it was first assumed that $S^2 = 1$ for $N = 1$ and that the initial K value is 1. A given error rate may now be chosen and the probabilities of the observations are thus determined. As N increases, the variance is, by assumption, increased and hence the new probabilities for a single observation can be obtained from normal curve tables. Given these probabilities and the incomplete beta function tables, those two proximate values of K which span the required error rate can then be found and hence the proportion u such that $uE_K + (1 - u)E_{K+1} =$ the required rate.