MEMORY SCANNING:
NEW FINDINGS AND CURRENT CONTROVERSIES*

SAUL STERNBERG

Bell Laboratories, Murray Hill, New Jersey 07974, U.S.A.

Contents

1. Introduction: The reaction-time method in memory research

Human memory has traditionally been studied by examining how and when it fails—by considering the frequency and pattern of errors in recall or recognition. These errors may result from failures of learning, retention, or retrieval, and one difficulty in the traditional approach is the disentangling of these alternative sources of error.

During the past decade a complementary approach to the study of memory has become increasingly popular. Here memory is examined under conditions in which it functions successfully and produces performance that is virtually errorless. By applying time pressure to the subject under these conditions, the experimenter can induce some of the mechanisms at work to reveal themselves, not by how they

fail, but by how much time they need in order to succeed. The questions addressed by such reaction-time (RT) studies have focused on mechanisms of memory retrieval for information in both short-term and long-term memory, but the approach is also being widely used to confront issues such as what information is stored and how it is coded and organized (see, e.g. Landauer and Meyer, 1972). All this is part of a general revival in the use of RT methods to infer the organization of perceptual and cognitive processes. Our current desire to analyse the processing of information into its functional components (particularly when combined with the hypothesis that component processes are arranged in stages) leads naturally to RT methods and to an interest in the temporal structure of processing. The power of these methods compared to others lies partly in the fact that the appropriate scale of measurement can be specified on the basis of relatively weak assumptions (Sternberg, 1969b). This means that quantitative aspects of data, such as additivity and linearity of effects, come to have powerful implications.

In this paper I shall focus on studies of the recognition of items in relatively short memorized lists. I start by briefly reviewing some of the experiments, now about 10 years old, that led me to infer the existence of a particularly simple but curious process of internal scanning (Sternberg, 1963, 1966, 1969a). Then I discuss, on the one hand, some of the extensions and generalizations of the early findings, obtained in a number of laboratories, and, on the other hand, several of the findings that appear to create serious difficulties for the scanning model. These findings have led others to propose alternative models; I consider three contrasting examples of alternative models, and point out some of their strengths and weaknesses. Finally I review several new findings from various laboratories—findings that particularly intrigue me and that bear on the elaboration or extension of the original scanning model.

2. The item-recognition paradigm

In the *item-recognition paradigm* we shall be considering, the *stimulus ensemble* consists of all the items that might appear as test stimuli. In the early experiments, for example, the ensemble was made up of one-digit numerals. From among the ensemble a set of elements is selected arbitrarily and is defined as the *positive set*; these items are presented as a list for the subject to memorize. The remaining items are called the *negative set*. In general, positive and negative sets are not distinguished by any simple physical or semantic feature. When a *test stimulus* is presented the subject must decide whether it is a member of the positive set. If it is, he makes a *positive response*, for example, by pressing one of two buttons. If it is not, he makes a *negative response* by pressing the other button. The RT is measured from the onset of the test stimulus to the response. The task is easy enough so that without time pressure, performance would be virtually perfect. Even with time pressure, the error rate can be held to 1 or 2% by paying subjects in such a way as to penalize errors heavily while rewarding speed, although in some experiments the error rate is somewhat higher.

Two of the procedures that have been used within the item-recognition paradigm
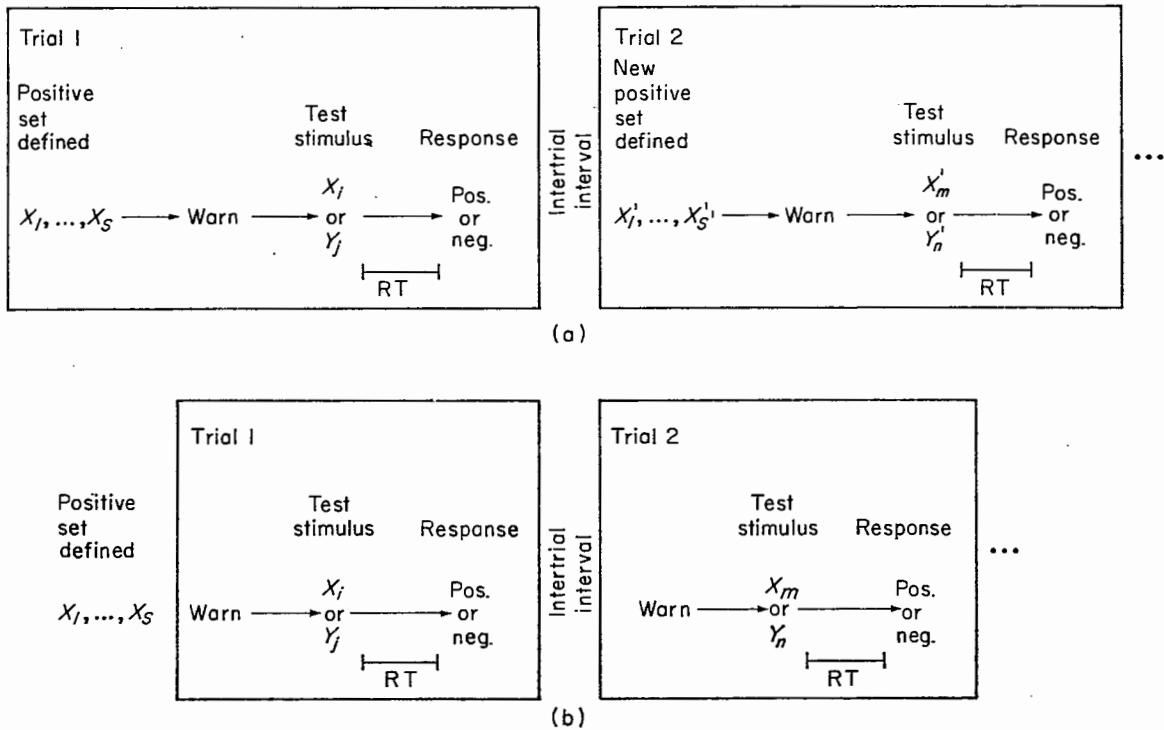
FIG. 1. Varied-set (a) and fixed-set (b) procedures in item recognition. A Y represents an item in the negative set. Primes are used in representing trial 2 of the varied-set procedure to show that both the items in the positive set $(X_1, \ldots, X_s)$ and its size (s) may change from trial to trial.

are shown in Fig. 1. In the *varied-set procedure*, the subject must memorize a new positive set on each trial. The set might be presented serially, followed by a retention interval of 2 or 3 s during which the subject is free to rehearse if he wishes, then a warning signal, and then a test stimulus. In the *fixed-set procedure*, the same positive set is used for a long series of trials, and a trial consists only of warning signal, test stimulus, and response.

3. Early findings and interpretations: the exhaustive search model

I and many others have studied the effects of a large number of experimental factors on performance in the item-recognition paradigm. The focus of most attention is the effect of varying the size of the positive set, while keeping constant the relative frequency with which positive and negative responses are required. The effect of this factor in an experiment using the varied-set procedure is shown in Fig. 2(a). The ensemble contained the 10 digits, and the positive set could contain from 1 to 6 different digits. Mean RT is plotted as a function of the size of the positive set. Four features of these data should be noted: first, mean RT increases approximately linearly with set size. Second, the rate of increase is the same for positive and negative responses. Third, the rate of increase is about 38 ms for each item in the positive set. And finally, the zero intercept is about 400 ms.

Figure 2(b) shows the effect of the same factor in an experiment using the fixed-set procedure where the positive set could contain 1, 2, or 4 different digits, and a

subject worked with each set for 180 trials. The results are similar: a roughly linear increase, at the same rate for positive and negative responses, with the slope of the fitted function about 38 ms/item, and the zero intercept about 400 ms. (In this experiment, the positive response was required on 27% of the trials. When the two responses are equiprobable in the fixed-set procedure, positive responses are produced about 40 ms faster than negatives, at each set size.) The phenomenon seems to be the same, whether the positive set is fixed over a long series of trials or is varied from trial to trial.*

Now, since the ensemble was constant in these experiments, changes in the size of the positive set induced complementary changes in the size of the negative set. Results of two other early experiments (Sternberg, 1963), which permit us to ignore this confounding, are shown in Fig. 3. In each of these experiments, the fixed-set procedure was used, the stimulus ensemble consisted of the 10 digits, the size of the positive set was fixed, and the size of the negative set was varied. In all conditions, subjects were told the members of both positive and negative sets in advance, and the positive response was required on half of the trials. Results are shown at the top of the figure for positive sets of size two, and at the bottom for positive sets of size one. Here, mean RT is plotted as a function of the size of the *negative* set. (In the fixed-set procedure, the only feature that distinguishes the two sets is that the positive set is the smaller. When the two sets are made equal in size this asymmetry disappears. Results for the smallest values of negative set size in each experiment, for which the characteristic that distinguishes positive and negative sets is missing, should therefore be considered separately from the other data.) In neither experiment was there a significant effect on the overall mean RT of the size of the negative set. Notice that this means that absolute stimulus probability *per se* had no effect, and also that ensemble size *per se* had no effect, since both of them varied with negative set size. In short, there is a striking asymmetry between the effects of positive and negative set size, at least when the items are digits. Positive-set size influences the latencies of both positive and negative responses, and in the same way. Negative-set size influences neither. As we shall see later, this finding has new importance in relation to some recent theoretical proposals.

How does a person decide whether the test stimulus is contained in the positive set? The early results discussed above, together with others, suggested a search through the positive set in which the test item is compared serially to each of the memorized items, and each comparison results in either a match or a mismatch. The data indicate two remarkable features of such a scanning process. First, the fact that positive and negative latency functions have equal slopes means that the search is *exhaustive*: even when a match has occurred, scanning continues through the entire positive set. Otherwise, if the scanning process terminated on the occurrence of a match, the positive function would have half the slope of the negative. [This follows from the fact that in *self-terminating* search of a set of s elements, the test item is compared to all s elements before negative responses,
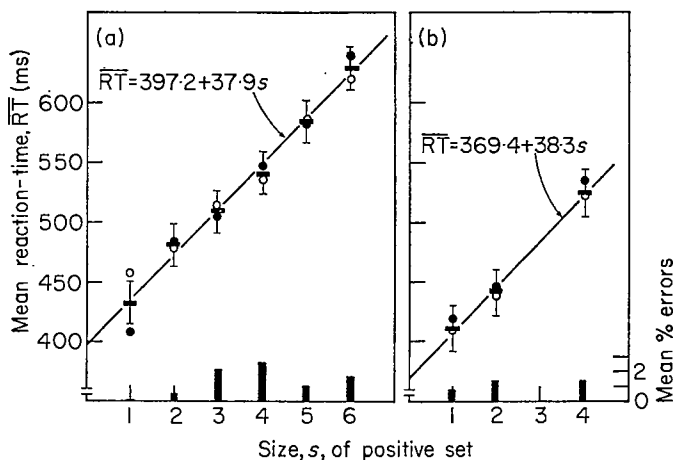
FIG. 2. Results from item-recognition experiments with varied-set [(a) eight subjects] and fixed-set [(b) six subjects] procedures. Mean latencies of positive responses (filled circles) and negative responses (open circles) as functions of size of positive set. Overall means (heavy bars) and estimates of $\pm$s.e. Lines were fitted by least squares to overall means.

whereas it is compared to $(s+1)/2$ elements, on average, before positive responses; the rates at which these two numbers increase with s are in the ratio of 2:1.] The second remarkable feature is the speed of the search process: each additional member of the positive set adds 1 to the number of comparisons producing mismatches that must occur in the course of the search. The time taken by each mismatching comparison is therefore estimated by the slope of the function, which is between 35 and 40 ms per item in the examples we have seen. This implies a scanning rate of about 30 items per second. Thus, the process is exhaustive and is substantially faster than estimates of the rate of covert speech (Landauer, 1962; Clifton and Tash, 1973). Judging from what subjects report, the search is not accessible to introspection.
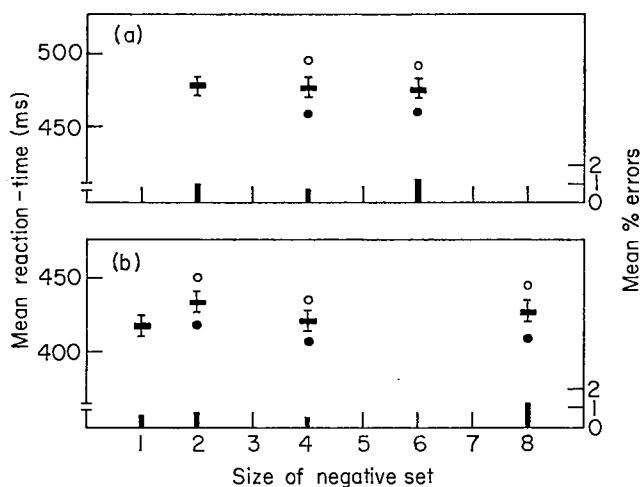
FIG. 3. Results of two experiments in which size of the negative set was varied. Mean latencies of correct responses (heavy bars) with estimates of $\pm$s.e., as functions of size of negative set, for positive sets of size $s=1$ [(a) six subjects; about 720 observations per point; $F=1\cdot02$, $df=2,8$] and size $s=2$, [(b) eight subjects; about 640 observations per point; $F=1.01$, $df=3,18$]. Also shown, for conditions where the two sets differ in size, are separate means for responses to stimuli in the smaller, positive set (filled circles) and for responses to stimuli in the larger, negative set (open circles).

Furthermore, the process seems largely unaffected by how well the set has been learned. The varied-set procedure is a short-term memory procedure: the test item is presented only 2 or 3 s after a single presentation of the set. On the other hand, on the average test trial in the fixed-set procedure, whose results are shown in Fig. 2(b), a subject had been working with the same positive set for 10 min; subjects could recall the sets they worked with several days later. In this case the sets must have been stored in long-term memory. Nonetheless, the similarity of the data led to the conjecture that the same memory was being searched in the two procedures: even if a set is well learned, when it is needed in the item-recognition task its members are "activated", or transferred into an "active memory", perhaps equivalent to the short-term store, where they are more rapidly accessible. Other experiments have confirmed this conjecture (e.g. Sternberg, 1969a, Experiment 5).

The height of the zero intercepts indicates that a large fraction of the RT reflects the duration of processes other than scanning. Even if we subtract estimates of input and output times (based, e.g. on simple visual RT data) we are left with more than 200 ms to account for. In other early item-recognition experiments, other experimental factors were studied. The effects of those factors, and certain additive relations among them (Sternberg, 1969b), pointed to the existence of at least three additional processes, arranged in stages, as shown in Fig. 4, whose dura-
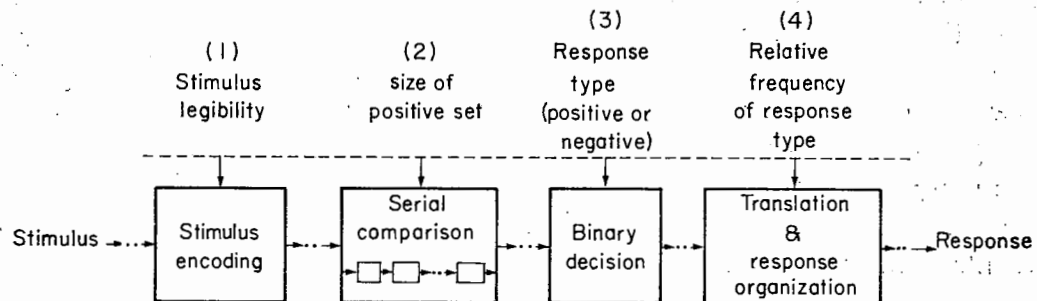
FIG. 4. Four processing stages in item recognition. Above the broken line are shown four experimental factors believed to influence the stages. Vertical arrows show each factor influencing only one stage. The additivity of the effects of factor pairs 1 and 2 (after a session of practice), 1 and 3, 2 and 3, 2 and 4, and 3 and 4 was the main reason for inferring the existence of the stages and the selectivity of factor effects (Sternberg, 1969b). Other factors, not shown, may also influence these same stages, and dots between arrows indicate the possibility of additional stages, as yet unknown.

tions contribute to the zero intercepts. Also shown are four experimental factors whose influence on the stages appeared to be selective after a session's practice. The legibility of the test stimulus influences the duration of an encoding stage in which an internal representation of the stimulus is formed. This representation is then used in the serial-comparison stage, whose duration increases linearly with positive-set size; in each of a series of substages the representation is compared to a memory representation of one member of the set. In the third stage a binary decision is made that depends on whether a match has occurred during the serial-comparison stage that precedes it; the mean duration of the third stage is greater for negative than for positive decisions. The selection and output of a response based on the decision is accomplished in a fourth stage, whose

duration is influenced by the relative frequency with which a response of that type is required. (Other factors, not shown in the figure, may also influence these same stages, of course.)

4. Generalizations and extensions of paradigm and phenomenon

Now let us move on from early data and theory to the question of the generality and robustness of the phenomenon. Reaction-time functions that are approximately linear, and with roughly equal slopes for positive and negative responses, have now been observed in various laboratories with a large variety of stimulus ensembles, both auditory and visual. (Later we shall come to some important exceptions, however.) The stimuli that have been used include visual and auditory digits and letters, two- and three-digit numerals, shapes, pictures of faces, drawings of common objects, words of various lengths, colours, and phonemes (e.g. Burrows and Okada, 1973; Chase and Calfee, 1969; Clifton and Tash, 1973; Foss and Dowell, 1971; Hoving, Morin and Konick, 1970; Sternberg, 1969a; Swanson, Johnsen and Briggs, 1972; Wingfield, 1973). The slopes of the functions are not the same for different ensembles, however; they appear to differ systematically from one ensemble to another, and in an orderly way, as described in section 7.5. The RT functions have been observed to remain linear and parallel in studies with positive sets containing up to 10 letters (Wingfield and Branca, 1970), and up to 12 common words (Naus, 1974).

Changes in the relative frequencies with which positive and negative responses are required alter the zero intercepts of the RT functions but leave their slopes and linearity relatively invariant (Sternberg, 1963; 1969b, Experiment 4).* And, so long as errors do not exceed about 10%, the slope and shape of the functions change very little when subjects are instructed to increase their speed at the cost of accuracy (Swanson and Briggs, 1969; Lyons and Briggs, 1971; Pachella, 1972; but see also Coots and Johnston, 1972, and Lively, 1972).

The phenomenon has been observed in people of various ages, ranging from children to elderly adults, and in normals, alcoholics, schizophrenics, and brain-damaged mental retardates. For some of these groups the slopes and/or intercepts of the RT functions are elevated relative to those of young adults; for example, aging and mental retardation both appear to produce increased slopes (Anders, Fozard and Lillyquist, 1972; Eriksen, Hamlin and Daye, 1973; Harris and Fleer, 1974). More remarkable, perhaps, is the absence of slope differences among certain groups of people, despite the existence of reliable individual differences within groups. Thus, children as young as 8 produce RT functions with higher intercepts, but the same slope as young adults (Hoving *et al.*, 1970; Harris and

---

* In the most complete experiment on this problem 12 subjects were studied in each of three relative-frequency groups, with positive responses required on 25, 50, and 75% of the trials (Sternberg, 1969b, Experiment 4). Mean slope estimates and their s.e.s were 30·5 ± 4, 35·9 ± 4, and 31·1 ± 4 ms/item, respectively, and the amounts by which the estimated slope for negative responses exceeded the slope for positives were 6·5 ± 3, −0·2 ± 3, and −1·1 ± 3 ms/item, respectively. On the other hand, the amounts by which the negative zero-intercept exceeded the positive zero-intercept showed an effect of relative frequency that was both systematic and large; for the same three groups these quantities were −22 ± 15, 50 ± 15, and 95 ± 15 ms, respectively.
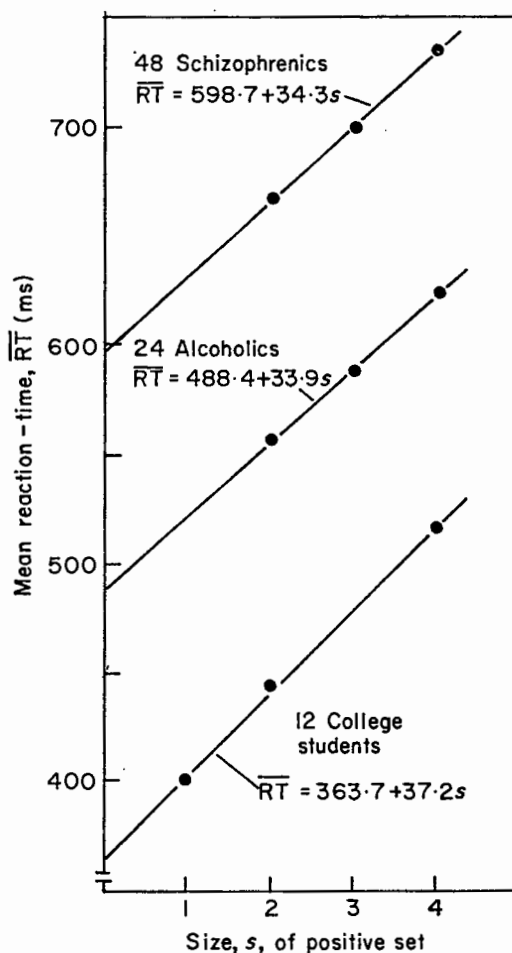
FIG. 5. Results from item-recognition experiments with three groups of subjects. Overall mean RTs as functions of size of positive set, and lines fitted by least squares. Data for schizophrenics (average hospitalization, 15 months) and alcoholics (average hospitalization, 8 months) from an unpublished study by S. F. Checkosky. For both groups the s.e. of the mean slope was about 2·2 ms, and slopes of separately fitted functions for positive and negative responses differed by less than 2·5 ms. Data show performance after an average of 1728 trials of practice (8 test days preceded by 4 practice days, with 216 trials per day). Data for college students from a similar study shown for comparison (Sternberg, 1967a, session 2); data show performance after an average of 630 trials of practice. Standard error of the mean slope for these data was about 3·4 ms.

Fleer, 1974). And, as shown in Fig. 5, except for their zero intercepts, schizophrenics and alcoholics look surprisingly similar to each other and to normals.*† Of all the groups studied, only aphasics produce data that are qualitatively (rather than merely parametrically) unusual, as discussed in section 7.3 (Swinney and Taylor, 1971).

Finally, let us consider the effect of extended practice in the item-recognition task. This effect seems to depend very much on details of procedure. Several studies have shown that when subjects practice with the same fixed sets over many days, the RT functions become flatter and negatively accelerated. This is particularly true if members of the ensemble are consistently associated with particular

---

* I thank S. F. Checkosky of Bell Laboratories, Holmdel, N.J., for furnishing me with details of this extensive unpublished study.

† The phenomenon may even extend to infrahuman primates (Eddy, 1973), although their high error rates in the task make interpretation of the RT data somewhat difficult.

responses, so that a stimulus that is in any positive set for a subject can never be in any negative set, and *vice versa* (Ross, 1970; Kristofferson, 1972*b*). On the other hand, when sets are changed either from trial to trial (Nickerson, 1966, Experiment 4), or from session to session (Kristofferson, 1972*a*), and stimuli are not consistently assigned to particular responses, extended practice seems to have virtually no effect on the phenomenon. The most complete experiment that demonstrates this invariance with practice was recently reported by Kristofferson (1972*a*).* She used seven subjects who worked for 144 trials per day for 30 days with an ensemble of digits, and with positive sets that changed from day to day. Results are shown in Fig. 6, averaged over positive and negative responses.
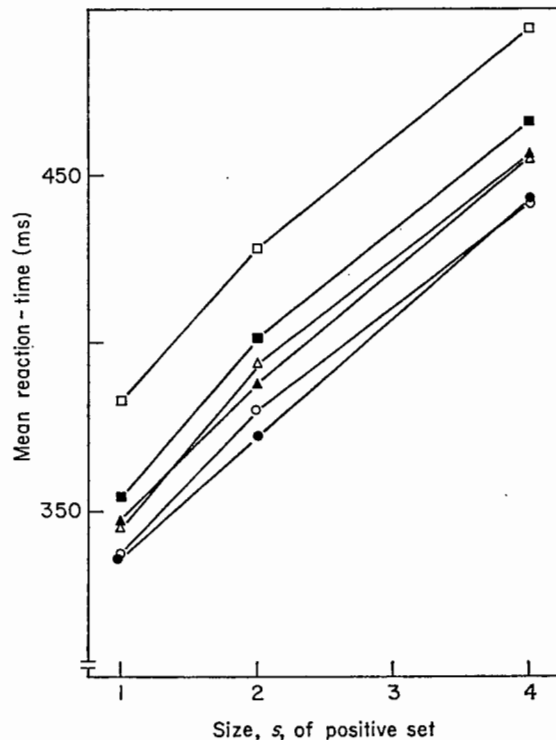
FIG. 6. Effects of extended practice on item-recognition performance: data from a study by M. W. Kristofferson (1972*a*). Mean RTs for seven subjects averaged over positive and negative responses for six consecutive groups of 5 days. Data points from each 5-day period are joined by line segments. Days 1–5, □; fitted line, 349·4 + 36·8 *s*. Days 6–10, ■; fitted line, 321·1 + 37·0 *s*. Days 11–15, △; fitted line, 312·6 + 36.6 *s*. Days 16–20, ▲; 312·2 + 36·2 *s*. Days 21–25, ○; fitted line, 305·4 + 34·6 *s*. Days 26–30, ●; fitted line, 299·2 + 36·0 *s*.

Practice appears to affect only the zero intercept. The average slope stayed very close to 36 ms per digit, deviation from linearity did not change systematically with practice, and the difference between mean positive and negative slope estimates, which are not shown here, was never more than 4 ms. Finally, for each 5-day period for each subject, the percentage of the variance of set-size means accounted for by linear regression was never less than 97.

* If a laboratory task is "artificial", in the sense of calling on skills not often practiced outside the laboratory, we expect relatively large effects of laboratory practice. Seen in this light, the remarkable invariance of the phenomenon with practice, under the conditions investigated by Kristofferson, suggests that the item-recognition task calls on mental operations that are also used in "everyday life".

5. Findings that challenge the model or limit its scope, and what to do about them

The robustness of the set-size effect indicates that it is a phenomenon worth explaining. But there are a number of reasons for questioning its explanation in terms of a high-speed exhaustive scanning process. Some of these reasons can be disputed, since they depend on judgments of plausibility. For example, the exhaustiveness of the inferred comparison process is difficult for many psychologists to accept because it seems inefficient.

Plausibility judgments, however, are highly subjective. I recently described some of this work to a group of Bell Laboratories engineers concerned with the design of special-purpose computers for telephone switching. They were obviously surprised when I commented on the implausibility of exhaustiveness; and when I asked about this later, they said that there were several instances of similar scanning processes being wired into the computer hardware, for the very reason that in those instances they were *faster*, on the average. It is easy to imagine systems in which the average search time is *less* for an exhaustive process than a self-terminating one (Sternberg, 1969a, section 11). See section 7.3 for new findings supporting the idea that human memory retrieval depends on such a system.

There seem to be several other reasons why the scanning model puts a strain on psychologists' intuition, and why we feel the need to seek alternative models. These include the high speed of the process (relative to rates of covert speech or to rates of overt sequences of discrete actions), the fact that subjects are not aware of it (even when introspections include a search, it is reported to be slow and self-terminating), the paucity of connections with theories of memory that have been designed primarily to explain forgetting, and the widespread belief (Kintsch, 1970; McCormack, 1972) that recognition depends on a stimulus having direct access to its representation in memory, rather than on any search process.

More serious for the scanning model than considerations of plausibility and intuition are several features of performance occasionally observed in item-recognition tasks. Some of these merely suggest limitations on the scope of the model. But others suggest that it must either be elaborated or discarded altogether. Discussed below are three of the more interesting challenges to the model, three strategies for elaborating the model to accommodate them, and one kind of finding that should, I feel, be used to limit the model's scope rather than to develop it further.

Let us consider first the serial-position function in the varied-set procedure. For each set size, this function relates the mean latency of positive responses to the position in the memorized list of the item that is matched by the test stimulus. Now, scanning processes lead very easily to predictions of serial-position effects if they are self-terminating—that is, if a positive response is initiated as soon as a match occurs. For example, in such a process, if scanning proceeded systematically from the beginning of the list, the serial-position function would increase monotonically. Different scanning orders and different mixtures of such orders could generate a large variety of serial-position curves, so long as the search is self-

terminating. But in an exhaustive search, this obvious source of serial-position effects is lacking. Since all items are searched, even on positive trials, the number of items searched before the match occurs should have no effect. Without elaboration, the model leads us to expect flat serial-position functions, and in many experiments this is what has been observed. But other item-recognition experiments have produced substantial serial-position effects. They vary in form, some showing recency effects (e.g. Clifton and Birenbaum, 1970), others primacy effects (e.g. Klatzky, Juola and Atkinson, 1971), and others both (e.g. Burrows and Okada, 1971). It is particularly embarrassing for the model if a procedure that gives rise to serial-position effects also produces set-size functions that are linear and parallel, since the latter suggests strongly that the situation is one to which the model ought to apply. A number of experiments showing position effects failed to demonstrate the basic phenomenon, because only one or two set sizes were used, the RT functions were nonlinear or nonparallel, error rates were unusually high, or RTs were outside of the normal range. But unhappily for the model, data from a few experiments whose gross features *are* consistent with the model have also shown serial-position effects. These effects appear to be most marked when the list to be memorized is presented rapidly and the interval between list and test stimulus is brief.

A second finding from the varied-set procedure that presents difficulties for the model arises when the same item appears more than once in a memorized list instead of all items being distinct. More data need to be collected with lists containing duplications. But on the basis of experiments by Baddeley and Ecob (1973) and unpublished data of my own, we can say, at least tentatively, that if the trials on which the duplicated item is itself the test stimulus are excluded, then the set-size function is not unusual; moreover, the effective set size is the same as if all the items in the list were distinct. The difficulty for the model arises when the duplicated item itself is tested, because on such trials the RT is unusually short. Again, the scanning model, as it stands, cannot account for this.

A third troublesome finding is related to the probabilities with which the different members of the positive set are presented as test stimuli. In most experiments these probabilities are equal. But a few experimenters have unbalanced the probabilities in the positive set, either by varying relative frequency in a fixed-set procedure (e.g. Miller and Pachella, 1973; Theios, Smith, Haviland, Traupmann and Moy, 1973; Biederman and Stacy, 1974), or by cueing the subject between presentation of the set and the test stimulus in a varied-set procedure (Klatzky and Smith, 1972). Responses to high-probability positives in a set are faster than to low-probability positives in the same set. If search were self-terminating, and probability influenced the order of search, such effects of probability could be readily explained. But in an exhaustive process this natural source of the probability effects is missing.

Let us now turn to some of the alternative theorizing strategies we might apply in dealing with these challenges to the model. Most investigators are strongly biased to place the burden of explanation on the second stage shown in Fig. 4—the serial-comparison process. Perhaps because the set-size effect is itself explained in terms of changes in the number of mismatching comparisons, so other

effects, they apparently feel, should be explained in the same terms; since the exhaustive-scanning model does not permit this, it must be discarded and replaced. This is one strategy of theorizing. In section 6 I discuss three of the models that have been proposed as replacements.

A second strategy is to preserve the second stage, but elaborate the model elsewhere, explaining these effects in terms of changes in the duration of other processing stages, such as the encoding stage or the binary-decision stage. For example, it is not inconceivable that the time needed to form an internal representation of the test stimulus in the encoding stage might depend on how frequently or recently that stimulus had been presented, or on the extent to which it was expected, thereby producing effects of trial sequence and stimulus probability. And it is possible when an item that appears twice in the positive set is tested, and two matches occur, that this increases the strength of the internal signal indicating a match, so that the binary decision is facilitated. Of course, unless conjectures like these are made precise and subjected to quantitative experimental tests, the model becomes too flexible. We will see later that the additivity of the effects of stimulus probability and set size can be regarded as one such test.

A third strategy is to invoke the idea of a probabilistic mixture of processes, as shown in Fig. 7. On the basis of an early analysis of the stimulus, a branch occurs,
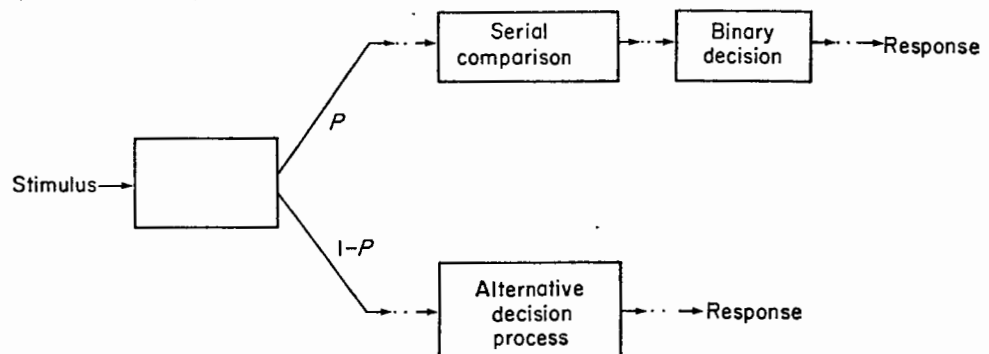
FIG. 7. A probabilistic mixture model. An early analysis of the stimulus on each trial determines whether the serial-comparison process or some alternative process will be used as the basis of the binary decision. The two processes are invoked with probabilities P and $1 - P$ respectively; P may vary across conditions within an experiment.

so that the serial-comparison process is executed on a proportion, P, of the trials, and an alternative process on the remaining trials. For example, in a fixed-set procedure, the test stimulus could first be compared to a representation of the previous test stimulus. If they matched, the previous response would be repeated, and if not, the positive set would be scanned. Again, the quantitative implications of such mixtures have to be determined and tested. For example, depending on details, mixtures need not affect the linearity of the set size function.

Some versions of the item-recognition paradigm produce findings that are sufficiently different from the early ones so that in my opinion they should be used to suggest limitations in the scope or generality of the model, rather than ways in which it ought to be elaborated, at least for the present. Some procedures, for example, consistently produce nonlinear set-size functions, particularly after

practice.   In one class of procedures where this occurs, members of positive sets are distinguished from members of negative sets by semantic or physical features (e.g. Marcel, 1970; Foss and Dowell, 1971), by large differences in familiarity or in frequency of presentation (e.g. Swanson and Briggs, 1971; Clifton, 1973), or by an assignment of responses to stimuli that is consistent over many trials or even sessions (e.g. Kristofferson, 1972b; Ross, 1970; Simpson, 1972; Swanson, 1974). One possibility is that such procedures provide an alternative basis for the positive-negative decision that subjects may find to be more efficient than the scanning process, at least for large sets.

6. Alternative models of the comparison process

6.1   *Self-terminating search*

We turn now to three of the mechanisms that have been proposed as alternatives to the exhaustive comparison stage.

Theios (1973; Theios *et al.*, 1973) has proposed a model whose aim is to account for data from a fixed-set procedure in which stimulus probabilities are unbalanced, and probability effects as well as parallel and linear set-size functions are observed. The proposed process is one of serial self-terminating search, but the list that is searched contains the entire ensemble of test stimuli rather than being limited to members of the positive set.   Each item in the searched list is assumed to be stored in association with a code that indicates the correct response for that item. Positive and negative items are mixed together in the list.   The stored list functions as a "push-down stack", whose members are rearranged from trial to trial in response to stimulus events.   As a result, the more recent and more frequent stimuli tend to be located near the beginning of the list.   The test stimulus is compared to one member of the ensemble after another until a match occurs, at which point the indicated response is initiated.   The model does reasonably well in accounting for both stimulus-probability and set-size effects in experiments in which the positive and negative sets are of equal size.   (Under these circumstances, for example, if the two responses are required with equal frequency, a simple version of the model implies that for any set size the average amount of search preceding positive responses is as great as that preceding negatives, hence the parallel set-size functions.)

One limitation of this model is that it cannot account in a natural way for findings from the more typical procedure, where positive- and negative-set sizes are unequal.   To fit the model here, two special assumptions have been added to it: First, positive stimuli are much more likely than negatives to move to the beginning of the stored list after being presented.   Second, when the positive set has s members, searching the list beyond position $s+1$ contributes no additional time to the RT.   Even with these assumptions, however, the model appears incapable of producing equal effects of positive-set size on RTs of positive and negative responses at the same time as it produces no effects of negative-set size.

A second limitation is that the model, tailored as it is to the fixed-set procedure, cannot readily account for results from the varied-set procedure, without adding

special assumptions about how associations involving items in the negative set on each trial are stored and ordered.   (In the varied-set procedure the negative set on a trial is defined only indirectly and by exclusion; hence these negative associations would have only the 2 or 3 s that elapse between presentation of positive set and test stimulus to be generated and intermingled among the positive associations in the stored list.)   The equivalence of results from fixed-set and varied-set procedures seems to me to be a very strong argument for seeking one model that can explain the performance in both procedures.

A third problem for any explanation in terms of self-terminating search is related to the behavior of the minimum of the RT distribution, as set size is increased.   For a self-terminating process the minimum must be invariant with set size.   The reason is that regardless of set size, there is some probability that the first comparison will produce the match.   (Unlike some of the other properties of the RT distributions produced by self-terminating search, this property does not depend on assuming the durations of successive comparisons either to have equal means or to be stochastically independent.)   Now, without making unacceptably strong distributional assumptions one cannot guarantee that any particular sample estimate of the population minimum is unbiased, or that its bias is independent of set size.   However, the fact that the sample minimum has been found to increase systematically with set size whenever it has been examined suggests strongly that the true minimum RT does not have the required invariance property (Lively, 1972; Lively and Sanford, 1972).   [Furthermore, other statistics can be devised that capture this aspect of self-terminating search, but do not suffer from the estimation difficulties of the minimum; the observed behavior of these statistics also provides evidence against self-terminating search, either through the entire stimulus ensemble or through the positive set only (Sternberg, 1973).]

Finally, experimental support is accumulating for the idea that stimulus probability produces its effects by influencing one or more stages *other* than the serial-comparison stage, and the encoding stage has been proposed as one possibility. There are two sources of evidence: first, in the three experiments where the question could be asked, the effects of absolute stimulus probability and set size were found to be additive, within the limits of experimental precision (Klatzky and Smith, 1972; Theios *et al.*, 1973; Biederman and Stacy, 1974).   The additivity is evidence that the two factors influence different processing stages.   Second, it has been shown that the effects of stimulus legibility and stimulus probability are not additive, which indicates that they influence a common stage (Miller and Pachella, 1973).   Given the assumption that stimulus legibility has its effect by influencing an encoding stage, this finding locates at least part of the probability effect in that stage.

In some models the self-terminating search is restricted to members of the positive set, at least on trials requiring positive responses.   In such cases, and when the stimuli in the positive set are equally probable, one can derive a theoretical relation between the increase in mean RT as a function of set size, and the increase in RT variance.   (The derivation does not require one to assume that components of the RT are stochastically independent.)   Although the variance is observed to increase markedly with set size, it does not increase fast enough relative to the mean

to satisfy this relation. Such failures provide evidence against this variety of self-terminating search. But the variance increase, which is approximately linear with set size and equal in rate for positive and negative responses, is what one would expect from an exhaustive search process with stochastically independent comparison times (Sternberg, 1964).*

6.2 Trace-strength discrimination

A second class of candidates for replacing the exhaustive-scanning model has appealed to several investigators as a way to extend more traditional concepts about memory to this domain; they can be described as pure trace-strength models (Baddeley and Ecob, 1973; Corballis, Kirby and Miller, 1972; Nickerson, 1972). In these models there is no search; instead there is direct access to an internal representation of the test stimulus. Members of the positive set are assumed to have acquired greater strength than nonmembers as a result of their presentation or rehearsal; the binary decision can thus be based on the strength of the accessed test-stimulus representation, as in a "signal detection" analysis. Such models can be made to generate variations in RT by assuming a functional relation between the trace strength (and possibly also an adjustable strength criterion) and the duration of the strength-discrimination process. The models differ in the particular functions assumed and in the rules by which trace strengths are assigned to items.

In the particular models proposed by Corballis *et al.* (1972) and Nickerson (1972), the most recently presented or rehearsed item in the positive set has the same trace strength, whatever the size of the set. One implication is that the minimum of the RT distribution for positive responses should not increase with the size of the positive set, just as described in section 6.1 for a self-terminating search. Since this implication is violated by existing data, I consider only the particular model proposed by Baddeley and Ecob (1973, Model I), which does not have the above property. Their aim was to account, in the varied-set procedure, for effects of serial position and item duplication within the positive set, as well as linear and parallel RT functions. According to their model, a fixed amount (C) of trace strength is divided among the s members of the positive set. As the set is increased in size, therefore, the difference between the average strengths of the positive and negative items $(C/s$ and zero, respectively) is reduced, making the discrimination more difficult. The duration of the discrimination process is assumed proportional to the reciprocal of the difference between test-stimulus strength and an adjustable strength criterion. To account for the fact that the latencies of negative as well as positive responses increase with set size, the criterion

* If r and s are two set sizes with $r < s$, and β is the slope of a linear function relating mean RT to set size, then it can be shown that the RT variance for a self-terminating search must increase by at least $(s^2 - r^2)\beta^2/3$ as set size is changed from r to s. For example, in an experiment using the fixed-set procedure with 36 subjects, and s-values of 1, 2 and 4 (Sternberg, 1969b, Experiment 4), overall mean RT was well described by $\overline{RT} = 387.7 + 32.5\,s$ ms, and the average RT variance, for both positive and negative responses, was well described by $\text{Var}(RT) = 4072 + 1084\,s$ ms^2. Thus, as s increased from $s = 1$ to $s = 4$, the RT variance increased by about 3252 ms^2, while a self-terminating search with $\beta = 32.5$ ms/item requires an increase of at least 5281 ms^2. Note, however, that the variance does increase precipitously with set size; whereas increasing the set size from 1 to 2 caused a change of only 8% in the overall mean RT, it caused a change of 21% in the overall RT variance.

is assumed to be moved closer to zero as set size grows. In particular, the criterion is assumed to be adjusted to $C/2s$, midway between the average strengths of positives and negatives, causing their RT functions to be parallel and linear.* The repetition effect can be accounted for by assuming that repeating an item in the set gives it a larger proportion of the available strength. And effects of serial-position on RT can be accounted for by variation in strength from item to item as a function of serial position.

At this point in its development there are at least three objections that can be leveled against the trace-strength model. The first and most serious is based on what happens when subjects have to retain a small number of additional items in short-term memory that are not relevant to the positive-negative decision, but are presented at the same time as the positive set and have to be recalled after the decision is made. If the irrelevant symbols must share in the limited available trace strength, the RT should increase with the number of such symbols that have to be retained. But a recent study (Darley, Klatzky and Atkinson, 1972) has shown that even in the varied-set procedure the RT is virtually unaffected by the additional load on short-term memory. This finding presents a serious problem for the strength theory. To avoid this difficulty, a trace-strength theory would seem to require a special limited supply of trace strength that can be shared only by members of the positive set.

The second objection is that a natural interpretation of "trace strength" makes it difficult for the model to account for performance when the positive set is well learned, as in the fixed-set procedure, and when negative stimuli have been presented as often as positives over a series of test trials, so that they should be equally familiar. Despite the large differences in stimulus frequency and degree of learning between these conditions and those of the varied-set procedure, the two procedures produce results that are virtually identical. To explain this, it has to be assumed that frequency of presentation *per se*, which is usually regarded as a critical determinant of strength in long-term memory, has no effect on the trace strength that is used in the decision process.

Finally, since the strength criterion used in the decision is adjustable, it should vary as a function of the relative frequency with which positive and negative responses are required, in order to minimize the mean RT. But it can be shown that such adjustment would change the relative slopes of the latency functions for positive and negative responses, as well as their zero intercepts. And such slope changes are not observed when the relative frequency of the responses is varied (Sternberg, 1963, 1969*b*, Experiment 4).

6.3 *Parallel comparisons*

The final alternative to the exhaustive-scanning stage that is considered here is a model in which the test stimulus is compared in parallel to all the members of the

---

* Insofar as the total trace strength, C, is distributed unequally among the members of the positive set, the linearity of the RT function for positive responses is only approximate: mean discrimination time is proportional to the mean of the reciprocal strengths, which is not identical to the reciprocal of the mean strength.

positive set. In most models of this type that have recently been considered, each of the simultaneous comparisons is assumed to require processing *capacity*; the increase of mean RT with set size is assumed to result from the sharing of a fixed capacity among these increasing demands, thereby reducing the capacity available for each comparison, which determines its duration. By suitably relating the capacity concept to processing time, a linear set-size function can be generated (Atkinson, Holmgren and Juola, 1969; Corcoran, 1971; Townsend, 1971).*

The particular parallel model that has attracted most attention has been considered by Atkinson *et al.* (1969) and by Townsend (1971). All comparisons start simultaneously and have durations that are exponentially distributed. The available capacity is assumed to be equally divided among those comparisons not yet completed, and the mean comparison time is assumed proportional to the reciprocal of the assigned capacity. The mean of the exponential distributions is therefore proportional to the number of ongoing comparisons. As soon as one comparison is completed, the limited capacity is instantaneously reallocated to the others, reducing their means. This continues until the final comparison is completed. If the process is assumed to be exhaustive, the resulting RT functions are linear and parallel. A model like this is proposed, not to account for effects that a serial process cannot explain, but rather as an alternative of possibly equal power.

My first comment about this model also applies to other parallel models that invoke the concept of a limited shared capacity. What is the limited capacity used for, other than the comparison process? Any concurrent task whose introduction has no effect on the RT function cannot be sharing in the same capacity. I mentioned above that introducing a small memory load that is not relevant to the decision has virtually no effect on the RT (Darley *et al.*, 1972). It has also been shown that adding a concurrent size discrimination to the item-recognition task leaves the RT function unchanged (Ellis and Chase, 1971). Therefore neither of these additional processing operations can be sharing the same limited capacity that is used by the scanning process.

One concurrent task that is always present in item recognition, at least in the varied-set procedure, is maintaining the memory of the positive set. In the particular parallel model mentioned above it is (implicitly) assumed that any

---

* Parallel models in which the duration, T, of any particular comparison is *not* influenced by the number of other ongoing comparisons, and in which the comparison durations are stochastically independent as well, are capable of producing an RT function for negative responses that increases with set size, so long as the comparison-time variance, σ_T^2, is assumed to be nonzero. This follows because the time to complete s simultaneously initiated comparisons is as great as the greatest comparison duration, and the mean of the greatest of s durations increases with s. (For positive responses the function would be flat if the process was self-terminating, and increasing if the process was exhaustive.) Without specifying the distribution of T exactly, however, it is not possible to describe the form of the RT function. But such models can be rejected regardless of the T-distribution because the observed effect on mean RT of the number of comparisons is too great relative to an estimate of σ_T^2. In particular, if t_s is the mean duration of the comparison stage for s comparisons, it is known that for any T-distribution, $t_s - t_1 \leqslant (s-1)(2s-1)^{-\frac{1}{2}} \sigma_T$ (Gumbel, 1954; Hartley and David, 1954). Although σ_T cannot, of course, be observed directly, the s.d. of the complete RT when $s = 1$ can be assumed to provide a generous upper bound estimate for σ_T, and can therefore be substituted for σ_T in the inequality. Despite its having been weakened by the substitution, the inequality was nonetheless violated in data from both fixed-set and varied-set procedures (Sternberg, 1963, 1966).

capacity required by a retention process is either fixed, regardless of the size of the memory load, or that its source is different from the capacity used by the scanning process.

Finally, difficulties are presented for this model in accounting for the RT function when it *is* changed by additional processing demands. Pure changes in slope, with no change in zero intercept, are easy to explain by assuming that added demands consume a fixed amount of the available capacity, independent of the size of the positive set. But pure intercept changes require a different rule for capacity sharing, in which the capacity allocated to the additional demand is reduced as the positive set is increased in size. Thus, Wattenbarger and Pachella (1972) found that embedding an arrows–keys binary choice task in an item-recognition experiment increased only the intercept of the RT function. Since performance in the embedded task was the same, regardless of the size of the positive set, any capacity allocated to it would have to be constant. But a constant reduction in the capacity available for the scanning process should have changed the slope of the RT function rather than its intercept. Here, again, the embedded task cannot have been sharing the same capacity. These constraints that are required on the capacity concept seem to me to reduce its appeal considerably.

My second comment concerns the use of the exponential distribution to represent the duration of a processing operation. Despite its popularity and simplicity, and its success in describing aspects of radioactivity and telephone traffic, I think it can be argued that the Markov (no memory) property of the exponential is fundamentally inconsistent with that what we normally mean by the concept of processing over time. Consider the processing of a single item. Given an exponential distribution it follows that, no matter how much time has elapsed, so long as processing is not yet complete the expected remaining time needed to complete it is unchanged. This means that at any point in time the processing that has been accomplished is either all or none and therefore that the processing of an item must be instantaneous. What is arranged in parallel in the particular model proposed by Atkinson *et al.* (1969) and by Townsend (1971), then, is not a set of processing operations, but a set of waiting-times prior to processing. The processing itself occurs at a series of points in time.

7. New findings

7.1 *The translation effect*

For the remainder of this paper I turn my attention away from alternative explanations of the set-size effect, and act as if there is an exhaustive serial-comparison process. In this context I consider some intriguing new findings that bear on several issues concerning the process, and also present new puzzles.

In the assumed scanning process, some internal representation of the test stimulus is compared to internal representations of items in the positive set. What is the nature of the internal codes that can be compared at such high speed? An equivalent question is: to what level is the test stimulus processed before it enters

into the series of comparisons? To what extent do the codes depend on the format or sensory modality in which the stimuli are presented?

In a number of laboratories, experiments that bear on this issue have revealed what can be called a *translation effect*. These experiments are concerned with what happens when the test stimulus is not identical to a possible member of the positive set, but is related to it by association. In some instances the associations were known by the subject before he arrived in the laboratory, such as the relation between printed letters and their spoken names. In other instances the associations were learned in the laboratory.

In one of their experiments, for example, Cruse and Clifton (1973) taught subjects a set of eight letter–digit associations. In the item-recognition experiment that followed, which used sets that changed from trial to trial, a set could contain either all letters or all digits, and the test stimulus could be either a letter or a digit, with equal probability. Consider a trial on which the set contains digits. If the test stimulus is also a digit, we have the standard task. But if the test stimulus is a letter, the subject must decide whether the digit associated with that letter is contained in the set. If the subject did this by translating the test stimulus into the "language" of the set and then searched for the translation, one might expect an increase in the zero intercept, reflecting the translation time, but no change in the slope. Exactly the reverse occurred: when the "languages" of set and test-stimulus differed, so they were related by association, the slopes of the RT functions for both positive and negative responses increased substantially—from 37 to 94 ms/item, on the average—but there was no increase in the zero intercepts. The same effect occurs, although less dramatically, when the two languages are printed letters and their spoken names, even when the subject knows the sensory modality of the test stimulus in advance (Chase and Calfee, 1969). And similar effects have been observed in several other cases, with slope increases ranging from 20% to over 100% (e.g. Swanson *et al.*, 1972; see also Burrows and Okada, 1973; Juola and Atkinson, 1971; Klatzky *et al.*, 1971; Nickerson, 1972; Sternberg, 1967a; Wattenbarger, 1970).

Several explanations are available, and it is too early to decide among them. The most obvious is that for some reason yet to be determined, if their "languages" differ, each member of the set is translated into the code of the test stimulus, or into some common code, after the test stimulus is presented. If each translation operation added the same mean increment to the RT, the result would be a change in slope, but no change in intercept, as observed. (If the test stimulus instead of the set were translated, the effect should be seen in the intercept, but not the slope.)

Two recent findings by Clifton and his associates, however, have led them to question this explanation: first, increases in the rate at which the positive set is presented that do not change the RT-function under normal conditions increase its intercept when translation is required (Clifton, Cruse and Gutschera, 1973). Second, variations in the learned associations that change the associative latencies for overt translation do not always change the slope of the RT function when covert translation is required in the recognition task (Clifton, Gutschera, Brewer and Cruse, 1973). An alternative explanation for the effect that they consider is that

both direct and translated representations of the positive set are generated before the test-stimulus is presented, but that the translated representation is stored in a less "active" form that is scanned more slowly than the direct representation.

Whatever the explanation of the translation effect, it seems evident from its existence that even after several seconds the memory code of the positive set may strongly reflect the form in which the stimuli were presented. This idea appears to conflict sharply with the ideas about recoding (of, e.g. visual stimuli into their spoken names) that psychologists have been led to by the study of errors in short-term memory.

7.2   Partial selectivity of search

The second set of findings I would like to consider bear on the selectivity of the scanning process. In most item-recognition experiments the memorized list that defines the positive set is relatively homogeneous, and there is no basis on which search could be selective. Suppose instead that the list is partitioned into distinct sublists, each containing a different category of items. It would presumably benefit the subject if he could confine his search to the "relevant" sublist—that is, the sublist containing the category to which the test stimulus belongs. For scanning to be selective in this way, two conditions must be met. First, the category of the test stimulus must be ascertained. Second, the list must be organized in memory in such a way that selective access to the relevant sublist is possible. In considering findings related to this question it is worth noting that if the scanning process is sufficiently fast and categorizing sufficiently slow, then selective search might be slower, on the average, than unselective search, so that selective search might not occur even if selective access was possible.

It has been shown that if the subject is informed of a relevant subset sufficiently in advance of the test stimulus, he can scan selectively, limiting his search to just that subset, even while maintaining the "irrelevant" items in short-term memory. That is, the subject is not obliged to search the entire set of recently presented items, even though they may be required for later recall (Darley et al., 1972). But perhaps more interesting is whether selective scanning occurs when there is no advance warning: is scanning of a recently memorized list that contains two or more well-defined sublists limited to the sublist whose category matches that of the test stimulus? A series of experiments by Naus et al. (1972; Naus, 1974) addressed this question. They used ensembles of common words organized into several different semantic categories. In one experiment (Naus et al., 1972), for example, the words were either girls' names or animal names. The positive set contained either words from just one category or equal-size sublists of words from the two different categories. The category of the test stimulus was always represented by at least one of the words in the positive set. For both kinds of sets the RT functions for positive and negative responses were linear and parallel. However, when the list contained two semantic categories rather than one, the slope was reduced by 25%, and there was no substantial intercept difference. If scanning were completely selective the slope would have been reduced by 50%, because only half of the words would have been scanned. On the other hand, if scanning were completely unselective, and if we assume that relevant and irrelevant words are scanned

at the same rate, then the slopes would have been equal. Neither of these was the case.

The 25% reduction suggests two simple interpretations. One is that on the average, exactly half of the irrelevant words, as well as all of the relevant words, are scanned at the standard rate. This was the interpretation made by Naus *et al.* (1972). It means that search is partially but not wholly selective. The alternative interpretation is that search is not selective at all, but that irrelevant words are scanned at exactly twice the rate of relevant ones. I use the word "exactly" because of results when the experiment was generalized to lists containing three and four equal-length sublists from different categories (Naus, 1974). The more categories, the smaller the fraction of the list containing words from the relevant category, and the greater the slope reduction. In each case, the observed slope of the RT function was within 1% of what would be expected from either of these interpretations.

A model of the process that corresponds to the Naus *et al.* interpretation is as follows. Words in the memorized list are stored in a set of bins, with one category per bin. Either because the bins are unmarked, or because the category of the test word is not ascertained, selective access does not occur. Instead, the bins are entered and searched successively in random order. The words in each bin are searched exhaustively, and at the standard rate, whether or not they are in the same category as the test word. However, when the bin containing the relevant category is entered, this fact is detected, and the search stops as soon as this bin has been scanned. Given this model, an irrelevant bin has a probability of exactly one-half of being searched, but the relevant bin is always searched. Hence, irrelevant words add half as much time, on the average, as relevant words.

How is the detection of category relevance achieved? Naus *et al.* suggests that the category "label" of the test word, $L(t)$, is compared to the category label, $L(b)$, of the bin words, and that this occurs at the same time as the contents of the bin are being scanned. But the data for short lists would then require that retrieving $L(t)$ and comparing it to $L(b)$ be very brief processes. An alternative is that the categorical identity of test word and bin word [t and b] or of test word and bin label [t and $L(b)$] can be recognized "directly", without having to delay the search by the time required to retrieve the category, $L(t)$, of the test word. In either case it is an open question why the search should be exhaustive within bins but not across bins.

Before we leave this topic, it is perhaps worth mentioning two objections to an alternative interpretation of these experiments, according to which all the irrelevant words are scanned, but at twice the rate. This interpretation has the advantage of being more consistent with results of studies (using different kinds of items and different procedures from Naus *et al.*) in which the test stimulus could be drawn from a category not represented in the positive set. For such stimuli the RT function typically has a slope that is neither as small as zero—which would indicate no search—nor as large as the slope for same-category stimuli—which would indicate search at the normal rate (Homa, 1973; Lively and Sanford, 1972; Milles and Morin, 1970; but see also Okada and Burrows, 1973). However, one objection is that the exactness of the 2:1 ratio between mismatching times for words from same and different categories that is required for the Naus *et al.* data calls for an

explanation in terms of process structure rather than parameter values. The other objection is based on the existence of some data indicating that when the words from various categories are scrambled in the list, rather than being arranged into sublists, and subjects do not rearrange them in memory, then the effect disappears (Naus, 1974).

7.3   *Relation between search structure and search rate*

If we had an acceptable explanation for the exhaustiveness of the scanning process in item recognition, this would seem to reduce the pressure to find alternative models. I turn next to some recent data that bear on this issue and seem to support the explanation I put forward some time ago (Sternberg, 1966, 1969a). The difficulty with exhaustiveness is that it strikes one as inefficient. Why should scanning continue after a match has occurred? But suppose we have a system in which a separate match-testing operation is needed to determine whether a match has occurred. If match-testing takes time and cannot occur concurrently with scanning, then, for a self-terminating search, the scanning process would have to be interrupted after each comparison. Now, the response depends only on whether there is a match, not on which item produces it. Therefore it might be faster, on the average, to store information that a match had occurred in a location that was examined only after all comparisons had been completed. But this would depend delicately on the relative speeds of scanning versus match-testing. Roughly speaking, the slower the scanning process, the less the efficiency to be gained from exhaustiveness.

Some support for this explanation seemed to be provided by the discovery that when the task was changed so that the response depended on which item produced the match, the scanning process became substantially slower and appeared self-terminating (Sternberg, 1967b). But this finding could be explained in other ways as well.

Table I shows results from five recent studies that provide evidence for self-terminating search in the item-recognition task itself, for either some subjects or

TABLE I

Experiment	Group or condition	Mean slope Negative responses (ms/item)	Mean slope ratio (positive/negative)
Clifton and Birnbaum (1970)	9 subjects	31	0·99
	3 subjects	47	0·47
Corballis and Miller (1973, group R)	6 subjects	45	1·02
	4 subjects	85	0·58
Klatzky and Atkinson (1970)	5 flattest	37	1·00
	5 steepest	113	0·44
Pachella (1972)	Before	50	1·0
	After	105	0·6
Swinney and Taylor (1971)	Normals	44	1·18
	Aphasics	121	0·55

some treatments.   The evidence is based on the relation between the slopes of the RT functions from positive and negative responses: for a self-terminating search the slope ratio expected is 0·5; for an exhaustive search, 1·0.   There is a fine dividing line between this kind of *post hoc* analysis and numerology, but I feel that the results are at least suggestive.   What is interesting is the association between small slope ratios and slow scanning rates; this association fits with the idea that it is the exhaustiveness of the process that makes its high speed possible, that the rate of self-terminating search is slower because of the time consumed by checking for the occurrence of a match after each comparison, and that although self-termination can sometimes speed positive responses, it must slow negatives.   Except possibly for the first two sets of data, this association cannot have resulted from a selection artifact.

In the Clifton and Birenbaum (1970) study, among 12 subjects, nine had slope ratios near 1·0 (range 0·80–1·25) and three had ratios near 0·5 (range 0·41–0·56). The mean scanning rate for the three was 50% slower.   In the Corballis and Miller (1973, group R) study, among 10 subjects, six had slope ratios near 1·0 (range 0·83–1·25) and four had ratios near 0·5 (range 0·50–0·65).   The mean scanning rate for the four was 89% slower.   In the Klatzky and Atkinson (1970) study, I divided subjects into two groups on the basis of the means of positive and negative slopes.   Again, the slow-scanning group displays a mean slope ratio near 0·5, and the others, who show the normal scanning rate, have a mean ratio of 1·0.   In the Pachella (1972) study the treatment placed extreme emphasis on speed.   This led to a reduction in intercept, but at the same time seems to have changed subjects from fast and exhaustive to slow and self-terminating.   Finally, in the Swinney and Taylor (1971) study, the normals appear to be fast and exhaustive, and the aphasics slow and self-terminating.[*]

In addition to supporting an explanation of exhaustiveness, some of these data seem to indicate bunching of slope ratios in the vicinity of 0·5 and of 1·0.   If this could be better documented in other studies, it would represent strong evidence against some of the models (such as those involving trace strength or processing capacity, considered in sections 6.2 and 6.3) that explain the set-size effect without a scanning process.

7.4   *Exhaustive search in long-term memory*

A number of experimenters are currently using RT methods to explore the processes underlying relatively accurate retrieval of information from large lists or categories that are well learned and presumably stored in long-term memory (e.g. Atkinson and Juola, 1974; Indow and Togano, 1970; Landauer and Meyer, 1972; Lovelace and Snodgrass, 1971).   Several reasons prompt me to discuss the model developed on the basis of a recent set of studies by Juola, Atkinson and colleagues, in which the item-recognition paradigm has been extended to well-learned ordered lists of up to 30 random words (Juola, Fischler and Wood, 1971; Atkinson and Juola, 1973, 1974).   First, these studies suggest that familiarity (or trace strength) does play a role in such tasks, while also showing how its role is limited, relative to

the one assumed for it in the models of pure trace-strength discrimination discussed in section 6.2.   Second, they reveal a search process in long-term memory that is qualitatively similar but quantitatively different from the exhaustive-scanning process we have been considering.   And third, they suggest a possibly fruitful elaboration of the exhaustive-scanning model, as applied to the original paradigm.

According to the model developed by Juola and Atkinson, the learning of a list of words has two effects.   One is that the familiarity value associated with a representation of each word in a lexical memory is increased; the other is that in a different part of the long-term memory the words are stored as an array.   Thus, two alternative bases for a recognition decision are established.   When a test stimulus is presented, there is direct access to its representation in the lexical memory, where its level of familiarity can be determined.   The difficulty with making the decision at this point is that even when positive items are from well-learned lists, and negative items have not been seen before in the experiment, familiarity is not a reliable indicator of list membership.   This idea is represented in Fig. 8(a).   Although responses based exclusively on familiarity would be uniformly fast, they would contain many errors.   Therefore, when higher accuracy is called for, the process indicated in Fig. 8(b) is used.   When the familiarity level is either high enough or low enough to be reliable, fast responses are made, based on familiarity discrimination alone.   The speed of these responses does not depend on how extreme a familiarity value the test stimulus has, and they lead to some small number of errors.   But when the familiarity level falls in the range of unreliability that lies between the two adjustable criteria, a search of the stored array is performed, which leads to an accurate response.

The result is a special instance of the probabilistic mixture of processes discussed in section 5 and illustrated in Fig. 7, in which both the mixing probability, P, and the mean RT on nonsearch trials are independent of list length.*   As shown in Fig. 8(c), the obtained RT function is a probabilistic mixture (weighted average) of two functions, one constant and the other increasing linearly; relative to the latter, the mixture is still linear, but has a reduced slope and intercept.

Parameter values in the fitted model that represent effects of list length on the search time for positive versus negative responses tell us that the search is exhaustive, and that although it takes about 70 ms longer to initiate than a search of short-term memory, it proceeds at the higher rate of about 10 ms (rather than 35 ms) per word.†   These parameter estimates are remarkably invariant across experi-

---

* Given that the RT function is based on correct responses only, the proportion, P, is a mean of two proportions, one for positive test stimuli, determined by the relation between the criteria (C_N and C_P) and the familiarity distribution for positive stimuli, and the other for negative test stimuli and determined by the relation between the criteria and the familiarity distribution for negative stimuli.

† These values unfortunately suggest an inconsistency with experiments using short lists. According to the parameter estimates, if the RT function produced by a search in active memory is $T_A = \alpha + 35\,s$ ms, then the function produced by a search in long-term memory is $T_L = \alpha + 70 + 10\,s$ ms.   For a list of length $s = 1$ or 2 that is contained in both memories (as in the fixed-set procedure), T_A is the smaller, so the active-memory search should be used.   But for a list of length $s = 3$ or more, the search in long-term memory is the faster. These considerations would be hard to reconcile with the similarity of results from varied-set and fixed-set procedures (Fig. 2) if this similarity could be demonstrated for words as it has been for digits.   They are also hard to reconcile with the effects on the RT function of fully loading the active memory with irrelevant material (Sternberg, 1969a, Experiment 5).

ments and procedures, an outcome that lends considerable support to the model, as does its overall goodness of fit to results from several studies. On the other hand, several of the special assumptions that play a critical role are open to question. First, the familiarity distribution of the memorized words, which presumably depends on degree of learning, is assumed to be independent of list
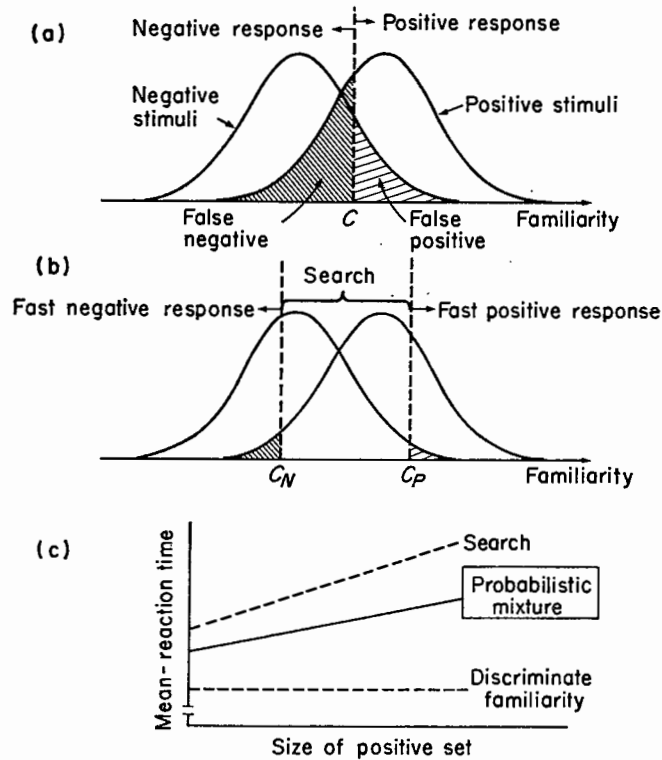

FIG. 8. The Atkinson–Juola (1974) model. (a) Overlapping familiarity distributions for positive and negative stimuli. In the middle range of values, the familiarity measure available by direct access to the representation of the test word in the lexical memory is not a reliable indicator of list membership. (b) Hence decisions are based on familiarity discrimination alone only outside the middle range of familiarity values. When the value is inside this range, a search of the stored array is performed that uses more time, but guarantees perfect accuracy. (c) Mean RT-function when search time (top function) is linear with list length, and mean RT on nonsearch trials (bottom function) and mixing probability are independent of list length. The overall RT function is a weighted average of the two contributing functions.

length. Second, the adjustable criteria are assumed to be independent of list length, even though lists of different lengths are learned by different subjects. Third, all errors are assumed to be generated by the familiarity-discrimination process; search trials are totally error free.

Elaborations of the exhaustive-scanning model for short lists contained in active memory, similar in spirit to the Juola–Atkinson model, would seem to be worth pursuing. Of course, if the ensemble is small, as it has usually been, and positive and negative sets are not distinguished by systematic differences in familiarity, the scanning process might be bypassed only very occasionally. But suppose these few occasions tend to be the trials on which familiarity of the test stimulus is high because it was drawn from certain serial positions, or because it had been duplicated

in the positive set.   Then this kind of elaboration of the model might be able to explain some of the troublesome effects considered in section 5.*

Atkinson and Juola (1974) have successfully applied such an elaborated model to an experiment by C. Darley and P. Arabie that used the varied-set procedure with words as stimuli, but differed from the typical procedure with short lists in that it used a very large ensemble of test stimuli.   Their aim was to introduce large familiarity differences between positive and negative stimuli.   Whereas the positive words on a trial were presented shortly before the test stimulus, a negative test stimulus might never have been seen in the course of the experiment.   Some of the negatives were deliberately made more familiar, however, by being drawn from the previous positive set.   According to the model, responses to such negatives should include more decisions based on scanning, and fewer based on familiarity discrimination.   This implies that their RT function should be steeper than the function for negative words that were novel, and this is what was observed.

It is interesting that even for the novel negative words, the model tells us that search occurred on more than half of the trials and that if, instead, all decisions about these words had been based on familiarity discrimination, about 20% of the responses would have been errors.   This result can be taken as further evidence of the inadequacy of a pure trace-strength model for explaining accurate performance.

7.5   *Relation between memory search and the memory span*

The similarity between retrieval mechanisms in short-term and long-term memory that is suggested by the work discussed above, along with its elucidation of the role of familiarity, seems to help integrate results and interpretations from the item-recognition paradigm with more traditional concepts and issues of memory. But nothing that I have yet said explicitly links measures of accuracy when there is no time pressure with measures of speed when there is.   Indeed, an important general question that is not often raised in relation to RT studies is to what extent the processes invoked in speeded tasks are the same as those used in tasks without great time pressure.   At the very least, it would be comforting to know that important parameters of performance in studies both of RT and of accuracy depend on the same factors.   For this reason, I find particularly encouraging the result shown in Fig. 9, from a study by Cavanagh (1972), which is the last new finding I shall mention.

Cavanagh found seven classes of items, ranging from digits to nonsense syllables, for which both memory-span and memory-scanning data had been obtained for similar groups of subjects under similar conditions.   Thirteen studies in the literature provided traditional measures of the immediate memory span for ordered sequences of items from these classes.   These values were averaged for

---

* One difficulty in applying the model as it stands to experiments using short lists and typical procedures is that changes either in overall error rate or in the relative speeds of positive and negative responses would presumably be mediated by adjustment of one or both criteria, and would therefore require concomitant changes in slopes of the RT functions.   But, as discussed in section 4, changes in error rates and zero intercepts have been produced by experimental variations that leave slopes relatively constant.

each class, and it is their reciprocals that are plotted on the abscissa.   If one uses a space metaphor for immediate memory capacity, and thinks of memory span as the number of items that fill a fixed space, then the reciprocal is a measure of the space required per item.   (Proportionately less of the fixed available space is presumably used by an item list that is shorter than the memory span.)   For these same classes of items, 32 studies in the literature provided measures of the memory-scanning rate, in terms of slopes of the best-fitting linear functions of RT versus set size.   These values of time required per item are plotted on the ordinate. What Cavanagh discovered, then, is that the time per item in the RT task is proportional to the space per item in the recall task.
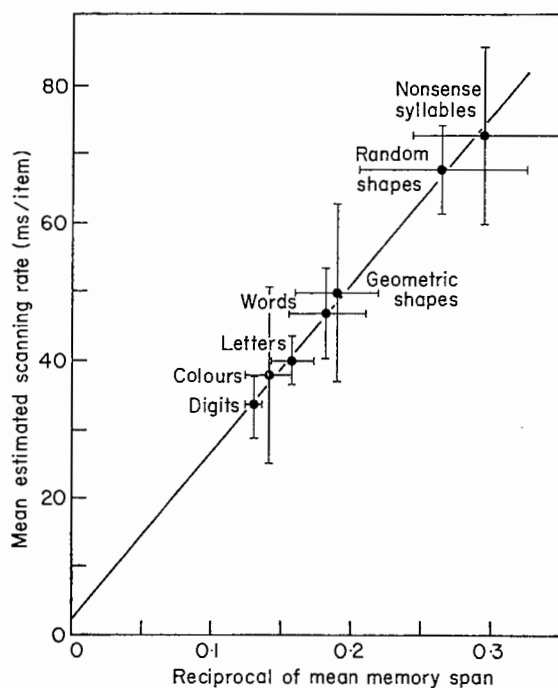

FIG. 9.   The relation discovered by Cavanagh (1972) between mean slope of the RT function in item recognition (time per item; unit is 1 ms) and reciprocal of the mean memory span (space per item; unit is one memory span).   Estimates of ±s.e. are based on pooled variance between studies, within classes of items.   Intercept and slope of fitted line are 2·4 ms/item (not significantly different from zero) and 245·5 ms/span, respectively.   Based on data from 45 experiments.

Suppose that this remarkable relation could be confirmed and made more precise in studies specifically designed for the purpose.   Its simplicity would then suggest that both measures are in some sense unitary or pure.   It would suggest, for example, that the immediate-memory span indeed reflects the capacity of a single, short-term or active memory, rather than depending on contributions from more than one memory system as has recently been argued (Craik, 1971; but see also Shallice, in press).   The relation would support the idea that a straight line provides a meaningful approximation to the RT function; a simple quantitative law would be even more surprising if it involved the slope of a line fitted to what was really a curve.   It would also suggest that so long as they are reliable, small slope differences in item recognition may be important.

Several directions of interpretation suggest themselves (see Cavanagh, 1972).

One possibility is that the average number of features that make up the internal representation of an item varies from one class of items to another. If the memory space required by an item, as well as the duration of mismatching comparisons involving that item, are both proportional to the number of features in its representation, then Cavanagh's finding would be expected. One possible difficulty for this kind of interpretation (and one that makes the finding itself particularly surprising) arises if we consider the effects of adding to the items features that are redundant, but that nonetheless must be recalled in the memory-span task. (An example would be adding a different colour to each member of a digit ensemble.) It seems likely that this would reduce the memory span without necessarily influencing performance in the item-recognition task, where the redundant feature could be ignored. If so, one would expect *natural* variations in redundancy from one class of items to another to similarly disrupt the relation between tasks.

On a dynamic theory of the memory span, involving a recycling "rehearsal" process that refreshes a decaying trace (e.g. Broadbent, 1958), Cavanagh's finding could mean that the rehearsal rate and memory-scanning rate are proportional. But if the process that refreshes the trace is the same rehearsal often identified with covert speech, this interpretation is violated by a recent experiment (Clifton and Tash, 1973), which showed that syllabic word length influences the rate of covert speech, but not the memory-scanning rate.

8. Summary

This paper has been concerned with the recognition of items in relatively short memorized lists, investigated with RT methods. First I described some of the early experiments that led to the idea of a high-speed exhaustive scanning process, and indicated some of the other processing stages that the data seem to require. Then I considered extensions and generalizations of the early findings, obtained in a number of laboratories, that show the phenomenon to be relatively robust, and the estimated scanning rate to be remarkably invariant across subject populations and practice. But effects of duplication of items in the list, their serial positions, and the relative frequency with which they are tested, show the scanning model to be either wrong, or insufficiently detailed in its description of how these effects might arise in the encoding and decision stages. Such effects have led others to propose alternative models of the comparison process. I considered three contrasting examples: an alternative serial process involving self-terminating search, a parallel process, and a direct-access process involving trace-strength discrimination. Each of these alternatives has weaknesses that appear to be at least as serious as the inadequacies of the exhaustive-scanning model. My current preference is for a strategy of theorizing that retains the exhaustive-scanning process and elaborates the model by investigating the other processing stages and considering mixtures of processes.

Finally I described five recent developments that I find intriguing and that seem to shed more light on the scanning process as well as on other issues in memory research. The "translation effect" tells us about the coding of information in memory. Work with lists of words that are organized into categories suggests that

search is partially but not wholly selective. The correlation between slopes and their ratios supports the idea that the high speed of the scanning process may depend on its exhaustiveness. Extension of the paradigm to long lists in long-term memory illuminates the role of familiarity in recognition and suggests elaborations of the scanning model for the case of short-term memory. And the relations between the scanning rate and the immediate memory-span links the recognition-RT approach to more traditional studies of recall accuracy.

References

ANDERS, T. R., FOZARD, J. L. and LILLYQUIST, T. D. (1972). Effects of age upon retrieval from short-term memory. *Developmental Psychology*, 6, 214–17.

ATKINSON, R. C., HOLMGREN, J. E. and JUOLA, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception and Psychophysics*, 6, 321–7.

ATKINSON, R. C. and JUOLA, J. F. (1973). Factors influencing speed and accuracy of word recognition. In KORNBLUM, S. (Ed.), *Attention and Performance* IV. Pp. 583–612. New York: Academic Press.

ATKINSON, R. C. and JUOLA, J. F. (1974). Search and decision processes in recognition memory. In KRANTZ, D. H., ATKINSON, R. C., LUCE, R. D. and SUPPES, P. (Eds), *Contemporary Developments in Mathematical Psychology*. Vol. 1. Pp. 242–93. San Francisco: W. H. Freeman.

BADDELEY, A. D. and ECOB, J. R. (1973). Reaction time and short-term memory: implications of repetition effects for the high-speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology*, 25, 229–40.

BIEDERMAN, I. and STACY, E. W., JR. (1974). Stimulus probability and stimulus set size in memory scanning. *Journal of Experimental Psychology*, 102, 1100–7.

BROADBENT, D. E. (1958). *Perception and Communication*. New York: Pergamon Press.

BURROWS, D. and OKADA, R. (1971). Serial position effects in high-speed memory search. *Perception and Psychophysics*, 10, 305–8.

BURROWS, D. and OKADA, R. (1973). Parallel scanning of semantic and formal information. *Journal of Experimental Psychology*, 97, 254–7.

CAVANAGH, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, 79, 525–30.

CHASE, W. G. and CALFEE, R. C. (1969). Modality and similarity effects in short-term recognition memory. *Journal of Experimental Psychology*, 81, 510–14.

CLIFTON, C., JR. (1973). Must overlearned lists be scanned? *Memory and Cognition*, 1, 121–3.

CLIFTON, C., JR and BIRENBAUM, S. (1970). Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, 86, 69–76.

CLIFTON, C., JR, CRUSE, D. and GUTSCHERA, K. D. (1973). Recoding processes in recognition: some effects of presentation rate. *Memory and Cognition*, 1, 387–94.

CLIFTON, C., JR, GUTSCHERA, K. D., BREWER, E. and CRUSE, D. (1973). Recoding in a character classification task: some inconsistent effects of recoding difficulty. Unpublished report, Cognitive Processes Laboratory, University of Massachusetts.

CLIFTON, C., JR and TASH, J. (1973). Effect of syllabic word length on memory-search rate. *Journal of Experimental Psychology*, 99, 231–5.

COOTS, J. H. and JOHNSTON, W. A. (1972). The effects of speed and accuracy strategies in an information-reduction task. *Perception and Psychophysics*, 12, 1–4.

CORBALLIS, M. C., KIRBY, J. and MILLER, A. (1972). Access to elements of a memorized list. *Journal of Experimental Psychology*, 94, 185–90.

CORBALLIS, M. C. and MILLER, A. (1973). Scanning and decision processes in recognition memory. *Journal of Experimental Psychology*, 98, 379–86.

CORCORAN, D. W. J. (1971). *Pattern recognition*. Harmondsworth: Penguin Books.

CRAIK, F. I. M. (1971). Primary memory. *British Medical Bulletin*, **27**, 232–6.

CRUSE, D. and CLIFTON, C., JR (1973). Recoding strategies and the retrieval of information from memory. *Cognitive Psychology*, **4**, 157–93.

DARLEY, C. F., KLATZKY, R. L. and ATKINSON, R. C. (1972). Effects of memory load on reaction time. *Journal of Experimental Psychology*, **96**, 232–4.

EDDY, D. R. (1973). Memory processing in *Macaca speciosa*: mental processes revealed by reaction-time experiments. Unpublished Ph.D. thesis, Carnegie-Mellon University.

ELLIS, S. H. and CHASE, W. G. (1971). Parallel processing in item recognition. *Perception and Psychophysics*, **10**, 379–84.

ERIKSEN, C. W., HAMLIN, R. M. and DAYE, C. (1973). Aging adults and rate of memory scan. *Bulletin of the Psychonomic Society*, **1**, 259–60.

FOSS, D. J. and DOWELL, B. E. (1971). High-speed memory retrieval with auditorily presented stimuli. *Perception and Psychophysics*, **9**, 465–8.

GUMBEL, E. J. (1954). The maxima of the mean largest value and of the range. *Annals of Mathematical Statistics*, **25**, 76–84.

HARRIS, G. J. and FLEER, R. E. (1974). High speed memory scanning in mental retardates: Evidence for a central processing deficit. *Journal of Experimental Child Psychology*, **17**, 452–9.

HARTLEY, H. O. and DAVID, H. A. (1954). Universal bounds for mean range and extreme observation. *Annals of Mathematical Statistics*, **25**, 85–99.

HOMA, D. (1973). Organization and long-term memory search. *Memory and Cognition*, **1**, 369–79.

HOVING, K. L., MORIN, R. E. and KONICK, D. S. (1970). Recognition reaction time and size of the memory set: a developmental study. *Psychonomic Science*, **21**, 247–8.

INDOW, T. and TOGANO, K. (1970). On retrieving sequence from long-term memory. *Psychological Review*, **77**, 317–31.

JUOLA, J. F. and ATKINSON, R. C. (1971). Memory scanning for words versus categories. *Journal of Verbal Learning and Verbal Behavior*, **10**, 522–7.

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

KINTSCH, W. (1970). Models for recall and recognition. In NORMAN, D. A. (Ed.), *Models of Human Memory*. Pp. 331–73. New York: Academic Press.

KLATZKY, R. L. and ATKINSON, R. C. (1970). Memory scans based on alternative test-stimulus representations. *Perception and Psychophysics*, **8**, 113–17.

KLATZKY, R. L., JUOLA, J. F. and ATKINSON, R. C. (1971). Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology*, **87**, 281–8.

KLATZKY, R. L. and SMITH, E. E. (1972). Stimulus expectancy and retrieval from short-term memory. *Journal of Experimental Psychology*, **94**, 101–7.

KRISTOFFERSON, M. W. (1972*a*). Effects of practice on character-classification performance. *Canadian Journal of Psychology*, **26**, 540–60.

KRISTOFFERSON, M. W. (1972*b*). When item recognition and visual search functions are similar. *Perception and Psychophysics*, **12**, 379–84.

LANDAUER, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, **15**, 646.

LANDAUER, T. K. and MEYER, D. E. (1972). Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, **11**, 539–49.

LIVELY, B. L. (1972). Speed/accuracy trade off and practice as determinants of stage durations in a memory search task. *Journal of Experimental Psychology*, **96**, 97–103.

LIVELY, B. L. and SANFORD, B. J. (1972). The use of category information in a memory search task. *Journal of Experimental Psychology*, **93**, 379–85.

LOVELACE, E. A. and SNODGRASS, R. D. (1971). Decision times for alphabetic order of letter pairs. *Journal of Experimental Psychology*, **88**, 258–64.

LYONS, J. J. and BRIGGS, G. E. (1971). Speed-accuracy trade-off with different types of stimuli. *Journal of Experimental Psychology*, **91**, 115–19.

MARCEL, A. J. (1970). Some constraints on sequential and parallel processing, and the limits of attention. In SANDERS, A. F. (Ed.), *Attention and Performance* III. *Acta Psychologica*, **33**, 77–92.

McCORMACK, P. D. (1972). Recognition memory: how complex a retrieval system? *Canadian Journal of Psychology*, **26**, 19–41.

MILLER, J. O. and PACHELLA, R. G. (1973). Locus of the stimulus probability effect. *Journal of Experimental Psychology*, **101**, 227–31.

MILLES, K. P. and MORIN, R. E. (1970). The effect of a basis for stimulus classification on recognition reaction time. Paper presented at a meeting of the Midwestern Psychological Association, Cincinnati, May 1970.

NAUS, M. J. (1974). Memory search of categorized lists: a consideration of alternative self-terminating search strategies. *Journal of Experimental Psychology*, **102**, 992–1000.

NAUS, M. J., GLUCKSBERG, S. and ORNSTEIN, P. A. (1972). Taxonomic word categories and memory search. *Cognitive Psychology*, **3**, 643–54.

NICKERSON, R. S. (1966). Response times with a memory-dependent decision task. *Journal of Experimental Psychology*, **72**, 761–9.

NICKERSON, R. S. (1972). Binary-classification reaction time: a review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, **4**, 275–318 (whole number 65).

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

PACHELLA, R. G. (1972). Memory scanning under speed stress. Paper presented at a meeting of the Midwestern Psychological Association, Cleveland, May 1972.

POULTON, E. C. (1973). Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, **80**, 113–21.

ROSS, J. (1970). Extended practice with a single-character classification task. *Perception and Psychophysics*, **8**, 276–8.

SHALLICE, T. (1974). On the contents of primary memory. In RABBITT, P. M. A. (Ed.), *Attention and Performance* V. New York: Academic Press (in press).

SIMPSON, P. J. (1972). High-speed memory scanning: stability and generality. *Journal of Experimental Psychology*, **96**, 239–46.

STERNBERG, S. (1963). Retrieval from recent memory: some reaction-time experiments and a search theory. Paper presented at a meeting of the Psychonomic Society, Bryn Mawr, August 1963.

STERNBERG, S. (1964). Estimating the distribution of additive reaction-time components. Paper presented at a meeting of the Psychometric Society, Niagara Falls, Ontario, October 1964.

STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652–4.

STERNBERG, S. (1967a). Two operations in character recognition: some evidence from reaction-time measurements. *Perception and Psychophysics*, **2**, 45–53.

STERNBERG, S. (1967b). Retrieval of contextual information from memory. *Psychonomic Science*, **8**, 55–6.

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

STERNBERG, S. (1969b). The discovery of processing stages: extensions of Donders' method. In KOSTER, W. G. (Ed.), *Attention and Performance* II. *Acta Psychologica*, **30**, 276–315.

STERNBERG, S. (1973). Evidence against self-terminating memory search from properties of RT distributions. Paper presented at a meeting of the Psychonomic Society, St Louis, November 1973.

SWANSON, J. M. (1974) and BRIGGS, G. E. (1971). The neglected negative set. *Journal of Experimental Psychology*, **103**, 1019–26.

SWANSON, J. M. and BRIGGS, G. E. (1969). Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, **81**, 223–9.

SWANSON, J. M., JOHNSEN, A. M. and BRIGGS, G. E. (1972). Recoding in a memory search task. *Journal of Experimental Psychology*, **93**, 1–9.

SWINNEY, D. A. and TAYLOR, O. L. (1971). Short-term memory recognition search in aphasics. *Journal of Speech and Hearing Research*, **14**, 578–88.

THEIOS, J. (1973). Reaction time measurements in the study of memory processes: theory and data. In BOWER, G. H. (Ed.), *The Psychology of Learning and Motivation*, Vol. 7. Pp. 43–85. New York: Academic Press.

THEIOS, J., SMITH, P. G., HAVILAND, S. E., TRAUPMANN, J. and MOY, M. C. (1973). Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, **97**, 323–36.

TOWNSEND, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception and Psychophysics*, **10**, 161–3.

WATTENBARGER, B. L. 1970. The representation of the stimulus in character classification. Unpublished Ph.D. thesis, University of Michigan.

WATTENBARGER, B. L. and PACHELLA, R. G. (1972). The effect of memory load on reaction time in character classification. *Perception and Psychophysics*, **12**, 100–2.

WINGFIELD, A. (1973). Effects of serial position and set size in auditory recognition memory. *Memory and Cognition*, **1**, 53–5.

WINGFIELD, A. and BRANCA, A. A. (1970). Strategy in high-speed memory search. *Journal of Experimental Psychology*, **83**, 63–7.