Organization and long-term memory search*

DONALD HOMA†

New College, Sarasota, Florida 33578

A recognition RT paradigm was used to assess a number of plausible search strategies in LTM for categorized lists. List length was determined by factorially combining two, three, or five categories with two, three, or five words per category, and test items could be one of four types: (1) positive, (2) a repeated positive, (3) a related negative, or (4) an unrelated negative. It was found that RT increased linearly with category size for both positive and related negatives (about 30-40 msec/item); whereas the increase was much smaller for the unrelated negatives, especially for three and five categories (about 9 msec/item). With an increase in the number of categories, RT increased at the rate of about 40 msec/category for all three test items. A theory of high-speed scanning for categorized material was proposed in which a serial and exhaustive search of the categories is first undertaken. If a positive category match is found, a serial and exhaustive search within the contents of the positive category is initiated; if no category match is found, the search is simply terminated. Some evidence was presented that categories recently probed may provide for a short-circuiting of the category search.

Over 100 years ago, Donders (1868) suggested a method for studying the characteristics underlying simple mental processes, e.g., time required to detect a stimulus, time required to discriminate between two or more stimuli, etc. Donders's method contained two main assumptions: (1) Mental events require "real" time, i.e., their execution is neither instantaneous nor is the event itself too brief for measurement, and (2) complex mental operations are composed of simpler substages which occur serially and whose time components are additive. After some initially promising results, the second assumption was criticized by Kulpe (1896), leading to a diminution of interest in the Dondorian additive model.

Recently, however, the additive model was revitalized by Sternberg (1966, 1969) in his study of retrieval processes from short-term memory (STM). Ss were required to make one response (e.g., pull a lever) whenever a stimulus from a predefined set was presented and to make a different response otherwise (e.g., pull a different lever). The size of the positive set was varied from one to six items on any particular trial. Sternberg interpreted the results as indicating that retrieval of information from STM was serial and exhaustive; the search was serial because reaction time (RT) increased linearly with the size of the positive set and exhaustive because of the equivalence of the slopes for the positive and negative stimuli. From the slope constant, it was hypothesized that each item of the positive set was successively compared to the test stimulus at a rate of about 40 msec per item.

Although alternative models for retrieval from STM have been suggested (e.g., Falmagne & Theios, 1969; Theios, Smith, Haviland, & Traupmann, 1971), relatively few models have been proposed for retrieval in long-term memory (LTM), especially where a recognition paradigm is used.

One approach has been to use relatively long lists but where the task constraints are otherwise similar to that used in memory scanning experiments for STM. In an experiment by Juola, Fischler, Wood, and Atkinson (1971), Ss memorized 10, 18, or 26 common and unrelated English words prior to recognition testing. For both positive and negative items, the increase in RT with list length was relatively small (slopes of about 6 msec/item). A model of recognition from LTM was proposed which included two basic factors: (1) the familiarity of the item being searched, based along signal detection considerations, and (2) a memory comparison process which was initiated only following the failure of a familiarity check to exceed either of two response strength criteria. However, no elaboration of the memory search itself was provided. Although an exhaustive scan similar to that proposed by Sternberg for STM was considered unlikely on the grounds that such a search would have to occur at an extremely rapid rate (6 msec/item or about 165 items per second). Even the small increase in RT with list length is suspect, however, since the number of repetitions of the test stimulus was confounded with memory set size, i.e., as memory set size increased, the number of test stimulus repetitions decreased; this alone may have accounted for the effects of list length (Hintzman, 1969).

One difficulty with the Juola et al study is that the lists of words comprising the memory sets may have been subjectively organized (SO) (Tulving, 1962) prior to testing, and yet, little is known about the effects of organization on recognition RT. Since instructions to recall and instructions to organize can result in recall performances indistinguishable from each other

*Based on a dissertation submitted to the graduate school at the University of Wisconsin at Madison, in partial fulfillment of the requirements for the PhD degree. The author wishes to thank William Epstein, Chairman of the dissertation committee, and Edward Smith for their help.

†Requests for reprints should be sent to Donald Homa, Division of Natural Sciences, New College, Sarasota, Florida 33578.

Fig. 1. Schematic diagram of categorized memory set, with indication of potential variables in memory search.

(Mandler, 1967), it seems likely that the memory sets were subjectively organized prior to testing. Unless it can be conclusively demonstrated otherwise, it is possible that the effects of SO may have masked any attempt to uncover a search strategy for LTM. For example, an exhaustive search was considered untenable primarily on the grounds that the search would be extremely rapid, i.e., 6 msec/item. It can be easily demonstrated, however, that an exhaustive search that utilized the organization of the memory set can fall well within the rates proposed by Sternberg (1966) for STM, e.g., if the 10-, 18-, and 26-item memory sets were subjectively categorized into approximately two, three, and five categories of about five words each and if the effects of organization would be to direct the search only through the positive category, then the rate of search would be increased fivefold relative to that reported (about 30 msec/item rather than the reported 6 msec/item). There is, in fact, cumulating evidence that formal organization can provide for a short-circuiting of the search, whereby organized units are processed (Clifton & Gutschere, 1971; Naus, Glucksberg, & Ornstein, 1972; Williams, 1971). An experiment by Briggs and Swanson (1969) suggests the possibility that an entire ensemble can be initially retrieved prior to some decision process. In the Juola et al study, it is conceivable that an ensemble (subjective category) could be retrieved, followed by an exhaustive or self-terminating search of its content.

An attempt to systematize the effects of organization on recognition RT that involves the same essential task

constraints as employed by Sternberg is suggested in the following section. The approach followed here is that the extent to which organization facilitates search in LTM can be optimally determined by pitting specific alternative models against each other that are empirically distinguishable. At one extreme, it is possible to test a search hypothesis that is independent of organization variables; at the other extreme, a mixing of a self-terminating or exhaustive scan is coupled with a comparison process involving the semantic characteristics of the searched material.

THEORETICAL SUBCOMPONENTS OF THE COMPARISON PROCESS: DEVELOPMENT OF SEARCH STRATEGIES FOR LTM

Consider the situation where the positive set has been organized into N categories of size s' . The schematic diagram in Fig. 1 defines all relevant parameters. In the schematic diagram, it has been assumed that a list of nine items has been organized into three categories of size four, two, and three items. The parameters B and K provide the basis for generating a number of different search strategies: (a) B is the time consumed for memory comparison for each item in the positive set, and (b) K is the time required to access the next category for memory search. The category switching parameter, K , may also contain comparison time for the category that has been retrieved (Schaeffer & Wallace, 1970). Two of the potential benefits of search through organized as opposed to random lists might be that (1) decisions about entire categories can be made without examination of their contents and (2) organization may facilitate switching time for categories, i.e., organization tells one what to search next.

It is possible to describe, for a number of likely search strategies, a family of mathematical curves relating performance as a function of (a) list length, (b) semantic relation of the test stimulus to the positive set, and (c) nature of the response required (yes-no). When the memory set is categorized, these predictions can be refined further to include (d) number of categories comprising the memory set and (e) size of the categories. For a categorized memory set, one of three distinct probes can be presented: (1) a positive item (P), i.e., an item which matches a member of the memory set; (2) a negative item which is related to one of the categories in the memory set (P-N), e.g., if one of the categories of the positive set was "four-footed animals" and contained the items "dog," "cat," "lion," and "pig," a P-N item would be a word like "goat," which matches the category but not its membership; and (3) a negative item which is unrelated to the positive set, e.g., if the three positive categories were "four-footed animals," "types of crime," and "weather phenomena," an unrelated negative (N-N) would be a word like "doctor," which matches neither the categories nor its members. To ascertain the nature of search through a categorized list,

consider the following predictions for three different search strategies:

Search Strategy 2

Exhaustive search of the categories: If the test stimulus matches none of the categories, the search is terminated and a negative response is made. If the test stimulus matches one of the positive categories, an exhaustive and serial search is initiated within that positive category only. This strategy is similar to the serial and exhaustive search in STM proposed by Sternberg. The major difference is that allowances for a serial and exhaustive search of the *categories* are made. If no category match is found, the search is immediately terminated. If a category match is found, an exhaustive search is initiated in the contents of the positive category. The outcome of the latter search may be either positive or negative, e.g., a P or P-N stimulus would satisfy this condition.

Search Strategy 3

Exhaustive search of the entire list: The organization of the positive set is not utilized and each item is inspected serially and exhaustively.

Search Strategy 8

Category bypass search, exhaustive search of the positive category only: In other words, search is directed only in the positive category. However, only the items are searched and the categories are bypassed completely.

Thus, Search Strategy 3 defines a retrieval process that makes little or no use of the organization of the material; Search Strategy 8 proposes that the primary benefit of organization is essentially to inform S where to "look next," i.e., the organization defines a subset of the material to be searched [similar to Shiffrin and Atkinson's (1969) self-addressable memory]; and Search Strategy 2 invokes a parallel between the search in STM defined by Sternberg and a search process in LTM, i.e., search may involve an interplay between STM and LTM such that "relevant" material is first retrieved from LTM, and the contents and resulting search occurs in STM.

In order to generate the RT functions for each of the three strategies, consider Search Strategy 2. The presentation of a positive item (P) will result in (1) serial and exhaustive scan of the categories, NK, where N = the number of categories and K = category processing time, and (2) a serial and exhaustive search of the contents of the positive category only, $s'B$, where s' = the size of the positive category and B = item processing time within the category. If the time for encoding the test stimulus and all response time is assumed to average a seconds, then the total RT for this stimulus (assuming additivity) will be $RT = a + NK + s'B$. If a related negative (P-N) test

item is presented, an identical RT results, i.e., a serial and exhaustive scan of the categories, NK, which is positive, followed by a serial and exhaustive scan of the contents of the positive category, $s'B$, plus associated encoding and response time, i.e., $RT = a + NK + s'B$. For an unrelated negative (N-N), the serial and exhaustive scan of the categories, NK, results in a negative match and the search is terminated, i.e., $RT = a + NK$. To summarize the predictions for this search strategy:

$$\begin{array}{ll} \text{Strategy 2: } RT = a + NK + s'B & \text{for P stimuli} \\ RT = a + NK + s'B & \text{for P-N stimuli} \\ RT = a + NK & \text{for N-N stimuli} \end{array}$$

For Strategy 3, no benefits of organization are used and a serial and exhaustive scan of the entire list is made, i.e., if N = the number of categories and category size, s' , is equal for all categories, then the size of the total list is simply Ns' . Since the organization is not utilized, each of the three test types should consume the same approximate amount of time during the search process, i.e., $RT = a + Ns'B$ for all stimuli.

For Strategy 8, only the contents of the positive category are searched. For both P and P-N items, search should only be a function of the size of the positive category, $s'B$, plus encoding and response time, a, i.e., $RT = a + s'B$. For the N-N items, the prediction is less clear; possibly a "best bet" category is searched in a manner identical to the P and P-N items, i.e., $s'B$, or a "best bet" category is accessed but not searched, i.e., K. To summarize the three predictions:

$$\begin{array}{ll} \text{Strategy 8: } RT = a + s'B & \text{for P stimuli} \\ RT = a + s'B & \text{for P-N stimuli} \\ ? \quad RT = a + K & \text{for N-N stimuli} \end{array}$$

If the parameters a, B, and K are arbitrarily assigned numerical values, $a = 300$ msec, $B = 30$ msec/item, and $K = 50$ msec/category, then the relationships describing RT as a function of the number of categories, N, and category size, s' , as well as for the three types of test stimuli, P, P-N, and N-N, can be illustrated for each of the three strategies (see Fig. 2).

Each of the three search strategies can be empirically distinguished in a variety of ways, including (a) the relationship of the search strategies as a function of the parameters s' and N (e.g., Strategy 8 is independent of the number of categories in memory) and (b) the relationship of each of the strategies for the test stimuli P, P-N, and N-N. Thus, for Strategy 2, the slopes for P and P-N items are the same but different from the slope for N-N items; additionally, the *difference* in RT for P, P-N, and N-N items is the same as N is increased. Neither of these predictions are possible for the other two strategies.

Table 1 contains a listing of search strategies that is partially exhaustive in that they vary primarily in the

Fig. 2. Three potential search strategies as a function of test type, number, and size of categories.

extent to which organization facilitates search. The strategies not discussed at this point provide for search processes that are exhaustive or self-terminating for either the entire contents stored in memory or subsets of the total contents. It can be easily demonstrated that none of the strategies yield identical predictions, i.e., all listed strategies are empirically distinguishable.

Method

Materials

From the Battig and Montague (1969) category norms, an initial 30 categories of 10 words each were selected and semirandomly assigned to two groups of five words each. The only restrictions were that (a) no word contained either an X or Z (neither letter could be displayed by the apparatus) and (b) no word exceeded six letters in length. Each of the two groups of

five words was approximately equated for relative frequency of occurrence as an associate to the category, word frequency (Thorndike & Lorge, 1944), and word length. One of the two sets of words for each category was then randomly designated as a positive set (P1) throughout the experiment and the other as categorically related but negative items (P-N). Five additional items, bearing no obvious semantic or taxonomic similarity to the two sets, were also selected (N-N items) for each category. Each of the N-N items matched one of the five positive items in letter length and approximate word frequency. After all 90 word groups (30 each of P1, P-N, and N-N) had been selected, a final check was made to insure that the N-N words of one category were also unrelated to the members of the remaining 29 categories. In sum, then, the 30 categories of P1, P-N, and N-N items produced 150 P1 items, 150 P-N items, and 150 N-N items, for a total of 450 words. For reasons to be noted shortly, the P1 items were presented for recognition a second time in the testing sequence, i.e., a positive item, P1, when presented a second time in the same testing sequence, was designated as a P2 item.

Table 1
Listing of Potential Search Strategies

Search Strategy	RT Function	Stimuli
1 Exhaustive search of the categories, exhaustive search within <i>all</i> categories	RT = $a + NK + Ns'B$ RT = $a + NK + Ns'B$ RT = $a + NK$	P P-N N-N
2 Exhaustive search of the categories, exhaustive search within <i>positive</i> category only	RT = $a + NK + s'B$ RT = $a + NK + s'B$ RT = $a + NK$	P P-N N-N
3 Exhaustive search of the entire list	RT = $a + Ns'B$ RT = $a + Ns'B$ RT = $a + Ns'B$	P P-N N-N
4 Exhaustive search of the categories, self-terminating search within <i>positive</i> category only	RT = $a + NK + B(s' + 1)/2$ RT = $a + NK + s'B$ RT = $a + NK$	P P-N N-N
5 Self-terminating search of the categories, self-terminating search within <i>positive</i> category only	RT = $a + K(N + 1)/2 + B(s' + 1)/2$ RT = $a + K(N + 1)/2 + s'B$ RT = $a + NK$	P P-N N-N
6 Self-terminating search of the categories, exhaustive search within <i>positive</i> category only	RT = $a + K(N + 1)/2 + s'B$ RT = $a + K(N + 1)/2 + s'B$ RT = $a + NK$	P P-N N-N
7 Self-terminating search of the entire list	RT = $a + B(Ns' + 1)/2$ RT = $a + BNs'$ RT = $a + BNs'$	P P-N N-N
8 Category bypass I: Exhaustive search of the <i>positive</i> category only	RT = $a + s'B$ RT = $a + s'B$ (?) RT = $a + K$	P P-N N-N
9 Category bypass II: Self-terminating search of the <i>positive</i> category only	RT = $a + B(s' + 1)/2$ RT = $a + Bs'$ (?) RT = $a + K$	P P-N N-N

Design

The number and size of the categories comprising the positive set were two, three, or five, combined in a factorial manner to produce nine combinations. Each S was exposed to positive sets of a constant number of categories but to all three levels of category size, e.g., for Ss who worked with memory sets containing two categories, the size of the memory set varied from 4 to 10 words. Category size and test type (P1, P2, P-N, and N-N) were within-S variables, and number of categories comprising the memory set was the only between-S variable.

During an experimental session, S was exposed to all 30 categories. When the positive set contained two categories, there were seven separate lists with two words per category, five lists with three words per category, and three lists with five words per category. When S worked with positive sets of three categories, the number of different lists containing two, three, and five items per category were five, three, and two, respectively. When the positive set contained five categories, there were three, two, and one lists of two, three, and five items per category, respectively.

During a testing sequence, all positive items comprising a particular list were tested. Thus, when the number of categories contained in the positive set was three and category size was five, all 15 P1 items were tested, as well as 15 tests on each of the corresponding items (P2, P-N, and N-N), for a total of 60 consecutive tests. Since there were two lists of three categories with a category size of five, a total of 30 observations for this particular treatment combination was supplied by each S. The number of observations per combination of category size and number for each test type supplied by each S varied from 25 to 30.

During the testing session, the word lists were presented in a semirandom manner, such that lists occurring early and/or late in the testing session were not predominately composed of categories of a particular category size.

The 30 categories were randomly assigned across category size to insure that each category occurred as often as possible with each category size and in combination with other categories. Whenever a category size of two or three items was required, words were randomly chosen from the full set of five items contained in each of the 30 categories.

For a particular word list, the order of test items in the testing sequence was randomized for each S. For half of the Ss, the positive key was paired with S's dominant hand; for the remaining Ss, the assignment was reversed. A total of 30 University of Wisconsin undergraduates participated in the experiment, 10 Ss each randomly assigned to memory sets containing two, three, or five categories.

Procedure

At an initial meeting, S was informed of the nature of the task and then sent home with the lists containing the positive sets for the following day, e.g., for an S who worked with positive sets containing three categories, a total of 10 word lists were provided, each typed on a separate card. S was instructed that each of the lists had to be memorized prior to testing on that list. On the next day, one of the word lists was selected and S was instructed to write down all the words on that list; if S was unable to recall all the words, he was allowed to study the list again for a few seconds, and written recall was repeated. With only a few exceptions, Ss were able to recall the lists perfectly on their initial attempt. Following successful recall, recognition testing of that set commenced. This procedure of alternating recall and testing was repeated for each of the lists.

S was seated at a two-key response box and display panel and told, in part, that his task was to "... respond as rapidly as possible to words you've been asked to remember." Following successful recall, a final study trial of all the words on that list was presented on the display panel, with words blocked

according to category but otherwise appearing in a different order than they had been ordered on the original study list. Following presentation of the final word for study on the display panel, the panel momentarily went blank (3 sec), followed by the word "TEST." A blank card was inserted following the word "TEST" and served to separate the study and test lists and to alert S to the upcoming testing sequence. All test stimuli remained exposed until a response had been made, which then terminated that display and presented the next test stimulus. A 3-sec lag intervened between each response and presentation of the next stimulus. The initial test item for any sequence was always a filler item, unrelated to the positive set, and disregarded from the later analysis. After the final test item had been responded to, S was allowed to relax briefly, then the next word list was selected for recall and test, etc. The entire session lasted about 1 h.

To insure that S understood the nature of the task and to reduce the initial warm-up effects on the test trials, three practice lists containing digits were used at the beginning of the session. In all, S made approximately 40 responses prior to the first experimental list of words.

Apparatus

A combination keypunch, control unit, and display panel were used to monitor the presentation rate, display and sequencing of the stimuli, and timing of the responses (accurate to ± 1 msec). Stimuli were punched in a coded format on standard IBM cards, which were then interpreted by the control unit and displayed as word stimuli to the S in an adjoining room. The display panel consisted of two horizontal rows capable of containing a maximum of six letter stimuli in the top row or single digits on the bottom row (digit stimuli were used only for practice lists). The S sat in a semidarkened room approximately 1 ft from the face of the display panel. A response box, located just below the viewer, contained two small thumb buttons separated by 2 in. On test trials, the depression of either response button terminated the display, punched the response and associated RT for that trial on the following card, and moved the next IBM card into the keypunch reader. The sequencing of stimuli was automatically repeated until the last response had occurred.

Results

For each S, a mean RT was calculated for the 25-30 observations contributed to each combination of number and size of categories and test types. Deleted from these calculations were any RTs that (a) resulted from an incorrect response (2.40%), (b) exceeded 2.00 sec, and (c) exceeded the mean RT for that cell by three or more standard deviations. The latter two classes accounted for less than 2% of the total responses and were removed, in part, because of occasional misspunches or machine failure to register the initial response of the S.

Each of the test types is shown separately in Fig. 3 as a function of category size and number of categories.

The two main within-S variables of test type and category size were both highly significant [$F(3,81) = 44.86$ and $F(2,54) = 35.24$, both $ps < .01$]. The between-S manipulation of number of categories contained in the positive set was also significant [$F(2,81) = 4.23$, $p < .03$]. The only other source that reached significance was the interaction of Test Type by Category Size [$F(6,162) = 7.14$, $p < .01$].

Fig. 3. Mean RT for each of the test types as a function of category size and number of categories (five, three, and two categories indicated from top to bottom in each cell).

Tests for trend were computed for both the overall effect of category size and the significant interaction of Test Type by Category Size; for both of these tests, only the linear component was significant [$F(1,27) = 51.61$, $p < .01$ and $F(2,81) = 10.45$, $p < .01$, respectively]. In neither case did the quadratic component approach significance (both $F_s < 1.00$). For the Test Type by Category Size interaction, the linear component accounted for over 98% of the variance.

Least-square fits across category size were calculated separately for each test type and number of categories; the resulting RT functions are shown in Table 2. When the number of categories was either three or five, the slopes for P-N and P1 items were of similar magnitude: 33 msec/item and 42 msec/item for P-N, and 30 msec/item and 33 msec/item for P1. In contrast, the slopes for N-N items were considerably less: 6 msec/item and 12 msec/item for three and five categories, respectively. When the positive set contained two categories, slopes for P1 and P-N still exceeded those for N-N items, although the differences between P1 and N-N

were diminished (24 vs 18 msec/item); whereas those between P1 and P-N were increased (24 vs 38 msec/item). Orthogonal contrasts for the linear component of the Test Type by Category Size interaction revealed that P1 and P-N items did not differ significantly in slope [$F(1,81) = 2.54$, $p > .20$]; whereas the N-N items differed in slope from the average of the P1 and P-N items [$F(1,81) = 21.50$, $p < .01$]. In sum, the slopes for P1 and P-N did not differ overall from each other, and both differed from the slopes for N-N items.

An analysis of the slopes for the N-N items was determined to see if they differed from zero. For each S and for each memory set containing two, three, or five categories, slopes for the N-N items were calculated across category size. Only in the case of two categories was the slope significantly different from zero [$t(9) = 3.48$, $p < .01$]; for three and five categories, these differences only approached significance [$t(9) = 1.49$ and $t(9) = 1.68$, $.10 < p < .20$ in both cases]. Estimated standard errors of the slopes for two, three, and five categories were 5.24, 3.89, and 7.30, respectively. For the case of five categories, one S had an unusually large slope value (49 msec/item); otherwise, 8 of the 10 Ss had values less than 20 msec/item and 4 had slopes less than 10 msec/item; for the three category case, the largest slope value was 24 msec/item, and 6 Ss had values less than 10 msec/item. For two categories, however, 6 Ss had slopes greater than 20 msec/item and the remaining 4 Ss had slopes of less than 10 msec/item. Because it seemed that the slopes were positive only for the case of two categories, the mean RT for two categories and the average of three and five categories was determined as a function of category size and test type. These RTs are shown in Fig. 4. For the average of three and five categories, it would appear that both P1 and P-N items not only increased linearly with category size but also with approximately equal slopes; for P-N and P1, the slope constants were 37 and 32 msec/item, respectively. The corresponding slopes for P2 and N-N items were considerably smaller, 13 and 9 msec/item, respectively.

Some indication of the extent to which the slope constants for each test type (derived from category size) increased or decreased, as the number of categories in the positive set was increased, can be determined by applying a least-squares fit of the slopes for each test type across the number of categories. In other words, it

Table 2
RT Functions for Each Test Type as a Function of the Number of Categories

Number of Categories	Test Type			
	P1	P2	P-N	N-N
Two	$23.96s' + 668$	$17.23s' + 650$	$37.76s' + 658$	$18.16s' + 677$
Three	$29.79s' + 720$	$8.74s' + 727$	$32.68s' + 722$	$5.77s' + 758$
Five	$33.48s' + 769$	$17.80s' + 744$	$42.09s' + 782$	$12.28s' + 791$
Mean	$29.08s' + 719$	$14.59s' + 707$	$37.52s' + 721$	$12.07s' + 742$

might be useful to know if the slope constants *within* a category were altered as the number of categories in the positive set increased. These derived estimates indicated that P1 items tended to have the largest increase across number of categories (about 3 msec/category), although this was primarily attributable to the increase from two to three categories. Both P-N and P2 items increased at a smaller rate (1.90 and 0.81 msec/category) and the slope for N-N items was actually negative (-1.29 msec/category). It would appear, then, that the slopes tended to be relatively stable across the number of categories contained in the positive set, although there was a tendency for P1 and P-N items to show some increase.

In a similar manner, it is possible to determine whether the intercepts of the original RT functions tended to increase with an increase in the number of categories contained in the positive set. If a least-squares fit is calculated on these intercepts for each test type, then an indication of the increase in RT due to an increase in the number of categories in the positive set can be derived. For P1, P-N, and N-N, the increase in intercept values as a function of the number of categories varied between 30 and 40 msec/category; the actual estimates were 33, 40, and 35 msec/category, respectively. For P2, the estimate was slightly lower, 28 msec/category. Thus, unlike the relatively stable estimates provided by slope changes across the number of categories, the intercept values increased across the number of categories and were relatively similar for each test type; adding a category to the positive set increased RT for all test types by about 30-40 msec.

Since the intercepts of the original RT function were influenced by the number of categories, whereas the slope constants were not, it is possible to apply a least-squares fit for the slope and intercept parameters simultaneously. When this analysis was computed for each test type separately, the following revised RT function was obtained:

$$\begin{aligned} \text{P1} \quad \text{RT} &= 579 + 29.0 s' + 42.1 N \\ \text{P-N} \quad \text{RT} &= 568 + 37.5 s' + 45.8 N \\ \text{N-N} \quad \text{RT} &= 639 + 12.0 s' + 31.0 N \end{aligned}$$

Although the slope and intercept parameters (for s' and N) are altered slightly from the foregoing estimates of these parameters, they provide estimates which are reasonably close, especially for the slope parameters. The variability in the intercept parameters can probably be attributed, in part, to the fact that this estimate was a between-S fit, based on only 10 Ss in each case, where a constant number of categories comprised the positive set.

Error Rates

The mean error rate for P1, P2, P-N, and N-N items was 4.85, 1.15, 2.96, and 0.58, respectively; for two,

Fig. 4. Mean RT as a function of category size and two and the average of three and five categories for each test type.

three, and five categories, the overall error rate was 1.62, 2.90, and 2.79, respectively. For none of the 36 combinations of test type, number, and size of the categories did the error rates exceed 7.50%.

Discussion

The results of the present experiment are supportive of a serial and exhaustive search both within and among the categories. Nonorganizational models (Strategies 1, 3, and 7), as well as the category bypass models (Strategies 8 and 9), can be rejected on the basis of the obvious utilization of the categories in the search, i.e., RT increased significantly as the number of categories in the memory set was increased.

Category Search

A self-terminating search among the categories would produce a greater increment in RT for the N-N items relative to P1 and P-N items as the number of categories was increased. In other words, all the categories must be searched for an N-N item; whereas only half, on the average, must be searched in order to locate a positive category match for the P1 and P-N items. An estimate of category search time can be determined by measuring the separation between RT functions for a particular test type as the number of categories is increased. Thus, the separation between RT functions for P1 and P-N should be approximately half that found for the N-N items, i.e., P1/P-N should be about 1.00 and P1/N-N and P-N/N-N should be about .50. The actual least-squares estimates

were .91, 1.35, and 1.48, respectively, derived from the category comparison times of 42, 46, and 31 msec/category for P1, P-N, and N-N items, respectively. According to an exhaustive category scan, these ratios should all be 1.00. While the estimates for category search times were computed from only 10 Ss in each group, these ratios would appear to conform more closely to an exhaustive scan of the categories (a second set of category scan estimates, which did not simultaneously estimate search time *within* the categories, provided even better support for an exhaustive scan; for P1, P-N, and N-N, these estimates were 33, 40, and 35 msec/category, respectively, or ratios of .82, .94, and 1.14 for P1/P-N, P1/N-N, and P-N/N-N, respectively).

Within Category Search

According to a self-terminating search within the category, the slopes for P1 items should be approximately half that for the P-N items for two, three, or five categories comprising the positive set. For an exhaustive scan of the category, the slope for P1 and P-N should be equivalent, i.e., the ratios should be 1.00. Under both theories, the slope for the N-N items should be zero (or close to zero), since the search should have been terminated following a noncategory match. For two, three, and five categories, the slope ratios for P1/P-N were .67, .91, and .82, respectively. The corresponding slope ratios for P1/N-N were 1.33, 5.00, and 3.00; for P-N/N-N, these ratios were 2.00, 5.50, and 3.67, respectively. When the positive set contained two categories, the predictions were ambiguous regarding either a self-terminating or exhaustive search; however, for three and five categories, the P1/P-N slope ratios would appear to favor an exhaustive scan, i.e., when the estimated slopes for P1 and P-N were averaged across three and five categories, the memory comparison times were 32 and 37 msec/item, respectively.

The nonzero positive slope for the N-N items is difficult for either a self-terminating or exhaustive search model, although the positive slope was greatest for the two category memory set; for two categories and the average of three and five categories, the slopes were 18 and 9 msec/item. Part of the positive slope might be attributed to those cases in which the N-N items resulted in a positive category match (with a resulting search within the category), i.e., with increasing category size, the probability that an N-N item might be interpreted by S as similar to one of the categories should increase, in spite of experimental precautions to select nonsimilar items. Regardless, for three and five categories, the slopes for N-N items were about 20%-25% those for either P1 or P-N items.

In summary, then, it would appear that the search through the contents of a category was exhaustive, as it apparently was for search across the categories. While the category estimates are based on groups of only 10

Ss, it is interesting that the category comparison time was approximately the same as comparison time for items within the category, i.e., about 40 msec/category and about 35 msec/item. For items unrelated to the positive set, i.e., N-N items, it would appear that the search is terminated following an unsuccessful category match. Of additional interest was the implication that the magnitude of comparison times for categorized memory sets, involving both categories and their contents, are similar to that estimated by Sternberg for STM.

THEORETICAL MODIFICATIONS OF MEMORY SEARCH

Potential Contributions of STM

If STM is viewed as akin to a rehearsal buffer, then it would appear that direct influences of STM in the present experiment would play a minimal role, especially for memory sets that exceed STM capacity and for negative items. It can be demonstrated that, even if S (a) filled a STM buffer prior to testing and rehearsed its contents during the testing sequence and/or (b) filled his STM buffer prior to testing and each test item "pushed out" one of the buffer items, with the current item taking its place (Waugh & Norman, 1965), then the majority of items would not be found in the buffer, i.e., the search would occur in LTM. None of the Ss mentioned using rehearsal strategies during the testing sequence in the present experiment. If a push-down view of STM is considered, then the only positive items (P1) that would find a match in STM would be those occurring in the first few test trials, i.e., if a buffer of five items is initially filled with items from the positive set prior to testing, then the only positive items (P1) that would find a STM match would be those items that occurred in the first five test trials of the sequence. After five trials, no additional P1 items would find a match in the buffer, because any positive item would have to be either a repeat (P2) or a P1 item not in the buffer. Similarly, it is not clear what role STM could play for either type of negative item, i.e., P-N or N-N items, since the negative set was never defined for S. It should be noted that an undefined negative set differs from that employed in the typical Sternberg paradigm and avoids the criticism of Theios et al (1971) that stimulus probability (a) affects RT for *both* positive and negative items and (b) excludes the necessity of having negative items stored in memory during the memory search.

Nonetheless, a push-down model provides one interesting possibility for occasionally short-circuiting the proposed memory search for material that is primarily in LTM. Two related assumptions are required: (1) Information may be initially accessed in LTM but its contents are transferred to STM for further analysis, and (2) the end result of a search is to prime that location such that a repeated test of that location largely

bypasses a complete search of the relevant material. The proposed interplay between STM and LTM might be initially to access categorical information from LTM (categorical search), with the resulting contents searched in STM. Thus, the category processing constant estimated in the present experiment (about 40 msec/category) might be composed of a number of substages, e.g., search and comparison, where the search locates the categories and the comparison involves a matching of associates between the test item and the category being processed (Schaeffer & Wallace, 1969, 1970). The best match would result in that category being transferred into STM, where its contents would be checked for a match with the test item. The similarity in item comparison time within a category (about 35-40 msec/item) and the comparison times suggested by Sternberg for STM suggest that this possibility cannot be ruled out.

Support for the second assumption is based on extremely sketchy data, but its implications are fairly provocative. It was found that RTs for P1 items from the *same* category as the previous item in the testing sequence were of shorter duration than P1 items from a *different* category, i.e., RTs to P1 items different from the P1 item on the previous trial appeared to be a function of the categorical identity on the two trials. For the case of three categories, the RT function for *same* category P1 items was $RT = 637 + 31s'$. For *different* category P1 items, the RT function was $RT = 718 + 30s'$. Not only was the difference in RT primarily due to intercept differences, but also the overall RT function for P1 items was similar to the latter estimate, i.e., overall $RT = 720 + 30s'$. These results would be consistent with a one-slot STM buffer where only the category of the previous trial was stored. When this same analysis was applied for the case of five categories, a similar savings in RT was noted that was localized primarily in the intercept: $RT = 696 + 43s'$ for the *same* category P1 items and $RT = 765 + 40s'$ for *different* category P1 items. The overall slope estimate for P1 for the five category case was $RT = 769 + 33.5s'$. The major inconsistency of these RT functions, however, is that the intercept estimates for *same* category P1 items should be equivalent for three and five categories, and this is not the case. Again, only a fraction of the total data lent itself to these analyses (on the average, a P1 item following a P1 from the same category was less than 8%), but it suggests a process similar to what is typically referred to as a repetition effect in RT studies (Bertelson & Renkin, 1966; Keele, 1969). Here, however, RTs are not only reduced for identical repeats (the same item following itself) but also a similar reduction in RT was found for different items but bearing the same categorical membership. One theoretical possibility that would be consistent with this tentative result would be to assume that the category of information from the previous trial relevant for an item match temporarily resides in a STM buffer. Upon

presentation of the next item, a check is first made to determine if the current item is from the same category as the previous item; if not, a search of the categories is initiated.

An experiment by Naus, Glucksberg, and Ornstein (1972) lends itself to this interpretation. In their experiment, Ss worked with either one or two categories, while list length was held constant. It was found that overall RTs were significantly reduced for the two category case. The interesting aspect to their data was that the slopes for the two category case were about 75% of the slope for the one category case, when RT was plotted as a function of list length. Two models were considered, both of which are self-terminating for the categories. In their random entry model, one of the two categories is selected at random and its contents searched. If the appropriate category is selected, the other category is bypassed entirely; otherwise, the search is directed to the second category. In the direct entry model, the appropriate category is always selected and the remaining category bypassed. Since the slope reduction favors the random entry model, it is assumed that categories are searched in an arbitrary manner. In terms of the present experiment, however, the direct entry model may still be appropriate for their results. With just two categories, the probability of an item being followed by an item from the same category should be higher than in the present experiment (also, the use of only two types of test items—positive and related negatives—rather than four, as in the present experiment, would increase this probability). If category repeats for positive items occurred 25% of the time, then a direct entry model could account for their results. Regardless, the model proposed by Naus et al does not seem necessarily inconsistent with the proposed exhaustive category scan for the present experiment.

Role of Item Similarity in Search

Although search through categorized lists is suggested here as being serial and exhaustive for both categories and their contents (except for N-N items), the data do provide for deviations from this strategy. In particular, RTs for N-N items were shown to increase slightly as a function of category size, a prediction inconsistent with either a self-terminating or exhaustive search model. Although the positive slope for N-N items was statistically significant only for the case of two categories, the slopes for three and five categories tended to be positive as well. One modification of the proposed search strategy, which does not fundamentally alter the basic nature of the search, is to assume that each unrelated negative is not, in fact, always treated as such; rather, with a certain probability, an item originally selected as having no obvious semantic similarity to the positive set is, nonetheless, treated as a related negative (P-N). The probability that an item originally selected as an N-N item bears some semantic similarity to *at least*

Table 3
Estimated and Actual RT and Slope Values for the
Modified RT-Function N-N Items

N	s'	RT		Slope	
		Actual	Estimated	Actual	Estimated
2	2	716	716		
	3	728	723	18 Msec	7 Msec
	5	769	737		
3	2	773	753		
	3	770	763	6 Msec	10 Msec
	5	789	784		
5	2	817	825		
	3	827	840	12 Msec	15 Msec
	5	853	871		

one of the positive or related negatives should increase with list length. In the present experiment, for example, an unrelated item might be treated as such 90% of the time; on the remaining trials, however, it might be treated as related to at least one of the categories.

Perhaps the simplest assumption is to let d_N be the probability that an N-N item is treated as a P-N item, where N = the number of categories; for these trials, an exhaustive search is initiated within that category positively matched. If $1 - d_N$ is defined as the probability that an N-N item is treated as such, then the RT function for N-N items can be modified in the following manner:

$$\begin{aligned} RT(N-N) &= (1 - d_N)(a + NK) + d_N(a + NK + Bs') \\ &= a + NK + d_N Bs' \end{aligned} \quad (1)$$

where d = probability of N-N match for a single category, d^N = probability of N consecutive noncategory matches, and $d_N = 1 - d^N$ = probability of at least one related match. By solving Eq. 1 for d_N , it can be shown that

$$d_N = \frac{RT - (a + NK)}{Bs'} \quad (2)$$

Estimates for a and K are available from the original RT function for N-N items, i.e., $a = 639$ msec and $K = 31$ msec/category. An estimate for B is best provided by the estimates derived for both P-N and P1 items (about 35 msec/item). With these estimated values, it becomes necessary only to find a least-squares solution to Eq. 2 for all values of N and s' . If different values for d are substituted in Eq. 2 with the parameter estimates $a = 639$ msec, $K = 31$ msec/category, and $B = 35$ msec/item, then a good estimate for a least-squares value can be determined for d . When this was carried out, the best estimate for d was .89, i.e., the probability that an N-N item was treated as such was .89 and the probability that it was treated as a P-N item was .11. Table 3 shows the estimated RT and slope values for the modified RT function, relative to the actual data.

Although the original data, as well as the estimate for K , were somewhat noisy, it would appear that the modified RT function can take into account the slight rise associated with increasing category size, especially for three and five categories. The estimate for two categories, however, is not very good. It should be emphasized that the particular correction for similarity suggested here pertains only to the specific conditions of the experiment, e.g., not only the number and size of the categories but also the original set of words chosen for the experiment.

CONCLUSION

The major conclusion of the present experiment is that search through categorized lists is serial and exhaustive for both categories and their contents and that this search occurs at a rate similar to that proposed by Sternberg for much smaller lists. Furthermore, item comparison time and category comparison time require about the same amount of processing time.

It is interesting to speculate on the efficiency of the proposed category search, both in terms of speed and accuracy. Intuitively, an exhaustive search would not seem a likely candidate for relatively long lists, e.g., 25 words. Yet, an exhaustive category search precludes a high error rate—in the present experiment, all the relevant ensembles of information are checked—and, in fact, only 20% of the 25-item word list is actually examined, i.e., only those five items in the appropriate category of the five category memory set are searched. For the unrelated negatives, it is proposed that none of the words are examined, at least when the list is categorized (although all the categories are searched). For more complex hierarchical memory systems containing many levels of subcategorizing ala Mandler (1967), a serial and exhaustive search could maintain high accuracy while adding only a small increment in total time to scan, e.g., for a hierarchy containing five major categories, each of which subsumed five subordinate categories of five words each, a possible search strategy would be to successively exhaust the larger order categories until the appropriate item match is found. For this example, only 5 of the 125 items need be examined, or 4% of the total word list, and only 10 of the 25 "categories" need be searched. In terms of total search time, the relative savings for such a search could be great as well, if the levels of the hierarchy are processed as rapidly as the levels considered in the present experiment, i.e., $K = 40$ msec/category. Thus, $RT = a + NK + Bs'$ for positive items in the present experiment. With five additional subordinate categories, each containing five items, the RT could conceivably be: $RT = a + NK + ML + Bs'$, where M = the number of subordinate categories subsumed by the superordinate category and L = subordinate category processing time. If $L = 40$ msec/subcategory, then the total increase in RT when one goes from a 25- to a 125-item categorized

word list would be only $ML = 5 \times 40 = 200$ msec. For an unrelated negative, N-N, conceivably no increase in RT would be added as the list was increased from 25 to 125 words, because unrelated words would be rejected after the initial category scan. Previous speculation that Ss might initiate their category search where they last probed could provide for additional short-circuiting of the search, viz., a search initiated in a location *embedded* in the hierarchy rather than beginning at the "top." In conclusion, the necessity of searching through long lists need not preclude a systematic serial and exhaustive scan.

REFERENCES

- Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, 80(3, Pt. 2).
- Bertelson, P., & Renkin, A. Reaction times to new versus repeated signals in a serial task as a function of response-signal time interval. *Acta Psychologica*, 1966, 25, 132-136.
- Briggs, G. E., & Swanson, J. M. Retrieval time as a function of memory ensemble size. *Quarterly Journal of Experimental Psychology*, 1969, 21, 185-191.
- Clifton, C., & Gutschera, K. D. Hierarchical search of two-digit numbers in a recognition memory task. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 528-541.
- Donders, F. C. Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 92-120. Transl. by W. G. Koster in W. G. Koster (Ed.), *Attention and performance II*. *Acta Psychologica*, 1969, 30, 412-431.
- Falmagne, J. C., & Theios, J. On attention and memory in reaction time experiments. *Acta Psychologica*, 1969, 30, 316-323.
- Hintzman, D. L. Recognition time: Effects of recency, frequency, and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, 79, 192-194.
- Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, 10, 8-14.
- Keele, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, 80, 243-248.
- Kulpe, O. *Outlines of psychology*. New York: Macmillan, 1895.
- Mandler, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), *Advances in the psychology of learning and motivation: Research and theory*. Vol. 1. New York: Academic Press, 1967.
- Naus, M. J., Glucksberg, S., & Ornstein, P. A. Taxonomic word categories and memory search. *Cognitive Psychology*, 1972, 3, 643-654.
- Schaeffer, B., & Wallace, R. Semantic similarity and the comparison of word meanings. *Journal of Experimental Psychology*, 1969, 82, 343-346.
- Schaeffer, B., & Wallace, R. The comparison of word meanings. *Journal of Experimental Psychology*, 1970, 86, 144-152.
- Shiffrin, R. M., & Atkinson, R. C. Storage and retrieval processes in long-term memory. *Psychological Review*, 1969, 76, 179-193.
- Sternberg, S. High-speed scanning in human memory. *Science*, 1966, 153, 652-654.
- Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, 30, 276-315.
- Theios, J., Smith, P. G., Haviland, S. E., & Traupmann, J. Memory scanning is a serial, self-terminating process. Wisconsin Mathematical Psychology Program, Report 71-1, 1971.
- Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, 1944.
- Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychological Review*, 1962, 69, 344-354.
- Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, 72, 89-104.
- Williams, J. D. Memory ensemble selection in human information processing. *Journal of Experimental Psychology*, 1971, 88, 231-238.

(Received for publication March 5, 1973;
accepted March 18, 1973.)