JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **2**, 65-74 (1963)

Recall as a Search Process[1]

DOUWE B. YNTEMA AND FRANCES P. TRASK

Lincoln Laboratory,[2] Massachusetts Institute of Technology, Lexington, Massachusetts

The purpose of this paper is to suggest that it can be helpful to think of recall as a search through memory and that a useful way to analyze a recall task is to consider how a data-processor might be programmed to find the correct response.

A POINT OF VIEW ON RECALL

The view of human memory that will be adopted here is in many ways similar to that proposed by Feigenbaum (1961). There is, however, a difference in emphasis. If behavior is controlled by a hierarchy of programs, as Miller, Galanter, and Pribram (1960) suggest, then some of the processes considered here are probably on a higher level than those with which Feigenbaum was primarily concerned.

A person will be regarded as consisting of two parts, a memory and a processor. For the present purposes the memory may be assumed to contain items of information (chunks, if one prefers) each of which bears a number of tags that describe it and show how it is related to other items in memory. For example, one might want to assume that after paired-associate learning the stimulus would have a tag showing where, or how, to look for the proper response; and if "elephant" were the response, one might assume that the item "elephant" would have tags showing that it belonged to the category of animals, that it began with the letter *e*, and so forth.

This is not a formulation that appears to impose any particular theory of learning. There seems to be no reason why these tags cannot be said to represent habits, or membership in larger *Gestalten*, or even the branches of decision trees. At any event, this memory works at the rather low level with which theories of learning are usually concerned, the level of habits and conditional reflexes. The processor is the unit responsible for obeying the instructions that *E* gives to *S*—and the instructions that *S* gives to himself. We suggest that a useful way to analyze a recall task is to form hypotheses about the items that are likely to appear in memory and the tags that they are likely to bear, and then consider what rules the processor might follow to search through memory, inspecting and testing the items and their tags, in order to find the answer to the question that *E* has posed for *S*.

What is meant by searching and testing will perhaps be clarified by an example. Suppose that *S* is read a list of five words at the rate of about two words a second. After a half-second pause he is read the same list again, but in a scrambled order and with one word left out. He is to say which word was missing.

It may perhaps be obvious that a person can perform this task with some success; but a small experiment was done to make quite sure. Ten members of the laboratory staff

[1] Presented by the first author to Psychol. Sect. of AAAS, Philadelphia, December, 1962, with the title "Recall as a form of data-processing."

[2] Operated with support from the U. S. Army, Navy, and Air Force.

were each given a practice list and then five experimental lists of the kind just described. The 50 experimental lists were all different and were constructed by drawing monosyllabic words from the Harvard PB Lists (Egan, 1948), avoiding rhymes and other obvious similarities. The *E* did not himself know as he was reading a list for the first time which word would be omitted the second time. The mean fraction of the trials on which *S* responded with the missing item was 0.80, and the standard error of the mean was 0.05.

How might a person go about this task? One way would be to hold only the first list in memory, and as each word in the second list was read, tag the similar item in memory as having been repeated. The retrieval procedure would then be simple: run through the words in storage in some convenient order, testing each word in turn, until finally a word was found that was not tagged as having been repeated. A second strategy would be to hold both lists in storage and use a two-level search: run through the first list, testing each word by searching the second list to see whether it appeared there. There are other possible strategies, but it is difficult to think of any way in which it might be feasible for a person to perform the task without making the equivalent of a search through the first list, a search in which items are somehow tested to determine whether they appeared on the second.

Note that *S* does not respond with the missing item because it has been reinforced most often, or because it is in any conventional sense the strongest response. Thus it is reasonable, indeed almost necessary, to assume that some sort of data-processing mechanism can intervene between memory and overt response. We are suggesting that it is convenient to visualize the action of this mechanism as a search and that it can be enlightening to ask by what rules the search might best proceed.

Two illustrations of this approach will be given. The first is a case in which it leads to a re-examination of a familiar problem. The second is a case in which it raises an unconventional question about recency.

Experiment I: Recall of Items Presented in Dichotic Pairs

The first example has to do with the immediate recall of digits that are presented two at a time, one to one ear and one to the other. In the case to be considered here three such pairs of digits are presented in rapid succession—i.e., at the rate of two pairs a second. Broadbent (1954) discovered that in free recall there is a tendency to report first all of the digits presented to one ear and then all of the digits presented to the other. He also found that when an *S* is required to report the first pair of digits first, then the second pair, and then the third, recall is less successful than when *S* is permitted to give the digits heard in one ear and then those heard in the other. The first finding has recently been confirmed by Bryden (1962); the second, with slight modification, by Moray (1960). Evidently it is easier to report the digits ear by ear than to report them pair by pair.

Broadbent's concept of sensory "channels" has been used to explain these phenomena (Broadbent, 1958, Chapter 9; Broadbent and Gregory, 1961). If the ears are to receive different items at the same time they must function as different channels. It is impossible to switch attention from one channel to another four times a second. The best strategy that *S* can adopt is therefore to fix his attention on one ear and perceive the digits presented to that ear at the time they are presented. He holds the stimuli presented to the other ear in a very-short-term memory called the S system and goes back to perceive them later. The digits are, so to speak, lined up in ordinary short-term memory (the P system) in the order in which they are perceived, and so they are most easily recalled in that order. If they are to be recalled in any other order—for example, pair by pair—then they must be rearranged, which is difficult, just as it is difficult to rearrange an ordinary list of digits and say them backward.

When the task is treated as a search, the assumption that *S* cannot pay attention to two things at once begins to seem unnecessary. Assume for a moment that both members of a pair are perceived and stored in memory at the time of presentation. What sort of plan should the processor be programmed to follow if it is to retrieve all six items and retrieve each item only once?

There are apparently some search plans that the processor can execute more readily than others. It is difficult to repeat an ordinary list of digits backward: the tags representing forward associations are apparently more convenient or more reliable than the tags representing backward associations. Suppose then that what Broadbent's discovery reveals is that it is difficult for the processor to go from one item to another that has appeared simultaneously, just as it is difficult in the backward-span experiment to go from one item to the item which preceded it. It is easy for the search to go forward in time, difficult to go sideways or backward.

If it is best to go forward in time, the processor should be programmed to consider the pairs in temporal order; retrieving one item from each pair, and then return to find the other member of each pair, taking the pairs in temporal order again. In that case there must be some provision for avoiding on the second pass the items that were retrieved on the first. The obvious procedure is to retrieve the digits tagged as heard on one side of the head and then return for those tagged as heard on the other. The alternative would be to choose one member of each pair at random on the first pass, and on the second pass appeal to some sort of record of items that have already been recalled. Such a record is probably not very dependable. If it is to be used at all, the search had best proceed by sides of the head; in that way the record of items recalled will at least be supplemented by the tags showing the sides.

In sum, we suggest that the items are most easily retrieved ear by ear because they have no other characteristic that so neatly divides them into two groups within which the processor may proceed in temporal order.

One way to test this hypothesis is to let the items have a second set of tags that divide them into two parallel sequences. Let there be two types of items, the ten digits and ten words that are not digits. Each pair consists of a word and a digit, and the side on which the word is presented varies haphazardly from pair to pair. On the search hypothesis it should be as easy to recall the words and then the digits as to recall the items heard on the left and then those heard on the right. In fact it should be a little easier: each item will be unambiguously tagged as a word or a digit, but there may at times be a little uncertainty about the side on which it was heard.

Under the attention hypothesis, recalling the words and then the digits should be difficult. The *S* cannot pay attention to both ears at once; and, since the words shift haphazardly from ear to ear, he cannot select in advance the ear to which the next word will be presented. Thus he cannot adopt the strategy of perceiving the words at the time they are presented and going back later to get the digits from the S system. As before, the best that he can do is perceive the items presented to one ear when they are presented and then go to the S system for the items presented to the other ear. Thus the words and digits will usually be lined up in the P system in a scrambled order. Recalling first one type of item and then the other will require that the items be rearranged about as thoroughly as if they were to be recalled pair by pair.

Another strategy should perhaps be considered before deciding what the attention hypothesis must predict. The *S* might perceive the item presented to the ear to which his attention was already directed, and if that item was not a word, quickly switch attention to the other ear. In this way it would in theory be possible to select the words by switching attention only twice a second on the average. There are two objections to this strategy. First, a digit that is perceived when a word is sought must somehow be excluded from the P system; otherwise the load on the P system will be increased. Second, the average rate of two changes a second is only an average. On a quarter of the trials *S* will be obliged to switch attention on two successive pairs, and on an eighth of the trials he will be obliged to switch on every pair. Under the attention hypothesis it still seems necessary to conclude that recalling the words and then the digits will be at least a little more difficult than recalling the items ear by ear.

There is another, more tentative prediction that can be made from the search hypothesis. It is not a differential prediction that discriminates between the hypotheses, but it is a requirement that one would prefer to see the data satisfy if the search hypothesis is to be accepted.

If the words and digits shift haphazardly from ear to ear the six items fall into various patterns, of which three examples are shown in Table 1. For instance, the list labeled "one crossing" shows the sequence of three words beginning in the right ear, crossing over to the left ear in the second pair, and remaining there in the third. A list like the example labeled "no crossings" seems to be a fairly striking event: *S* is likely to notice that all of the words are on one side. If he does there is a special strategy that he can adopt when given the task of recalling the words and then the digits, or the task of recalling the items on the left and then those on the right. In either case the processor may be instructed to search first for items that are tagged as words and as heard on the left and then search for items that are tagged both as digits and as heard on the right. A search conducted in this manner—i.e., on two redundant principles—should be particularly smooth and efficient; so lists with no crossings should be recalled particularly well.

TABLE 1
EXAMPLES OF DICHOTIC LISTS USED IN EXPERIMENT I

No crossings		One crossing		Two crossings	
Left	Right	Left	Right	Left	Right
bet	7	3	part	0	good
sage	1	worse	8	roam	2
jack	4	house	6	5	coil

Method

The *S*s were 24 students at Tufts University. They were paid for their work and were naive in experiments on dichotic listening.

The stimuli were recorded on a two-track magnetic tape and were presented over earphones. A trial began with a short warning tone sounded in both ears. After a pause of 1 sec. the list of three pairs of stimuli was presented at the rate of two pairs per sec. A pair consisted of a word presented to one ear and a digit presented simultaneously to the other. The digits were three different digits, and the words were three of the following: *bet, coil, good, house, jack, muff, part, roam, sage,* and *worse.* After the last pair *E* stopped the tape-recorder and *S* attempted to repeat the six items aloud. He was instructed to say exactly six items, guessing when he could not remember.

There were three conditions. In the Pairs condition half of the *S*s were instructed to report the first pair of items, then the second pair, and then the third, within each pair giving first the item on the left and then the item on the right; the other *S*s received the same instructions with left and right reversed. In the Sides condition half of the *S*s were to give the items on the left in temporal order and then the items on the right in temporal order; with the other half left and right were reversed. In the Types condition half were to give the digits in temporal order and then the words in temporal order; with the other half digits and words were reversed.

Each S made 16 trials under each condition. The 16 trials were made in a block and were preceded by 3 or (for the first block) 5 practice trials made under the same condition. Order of conditions was balanced across Ss. The 16 lists in a block included 4 of each of the 4 possible kinds—i.e., no crossings, two crossings, a crossing after the first pair, and a crossing after the second pair.

At the beginning of the session S studied and attempted to memorize the ten words, listened to a recording of the words and digits, and then, as training in dichotic listening, made ten trials in which single word-digit pairs were presented.

The tape-recording of the stimuli was prepared automatically by the Lincoln Laboratory TX-2 computer. The voltage produced by a male speaker reciting the ten words and the ten digits into a microphone had been measured every 100 μsec., and the measurements had been stored in the computer's memory. Two of the items could then be reproduced at the same time by programming the computer to set the output voltages of a pair of digital-to-analog converters to the appropriate values every 100 μsec. The outputs of the two converters were recorded on the two tracks of the magnetic tape that was played to the Ss. The recording was made automatically because the facilities happened to be available, not because such elaborate control of the stimuli is considered important in the present experiment.

Results

An item was counted as correctly recalled only if it was reported in the correct position. Figure 1 shows the results.

Note first that recall is less accurate when

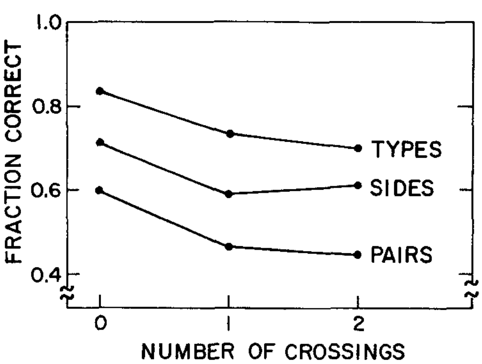

FIG. 1. Immediate recall of six items presented rapidly in dichotic pairs: mean fraction of the items correct and in the correct positions. Legends on the curves indicate required order of recall.

S is instructed to report by simultaneous pairs than when he is instructed to report the items heard on one side and then those heard on the other. The difference in mean fraction correct is significant at the 1% level according to an analysis of variance in which the order of report and the kind of list (i.e., number and location of crossings) were regarded as fixed factors, while Ss and lists within a kind were regarded as sampled from large populations. Thus the present method is adequate to reproduce the phenomenon with which we are concerned.

The crucial comparison is between recall by types of material and recall by sides of the head. Under the attention hypothesis, recall by sides of the head should be more accurate. Figure 1 indicates that it is not, and an analysis of the sort just described rejects at the 1% level the proposition that the mean fraction right is larger in the Sides condition. The attention hypothesis must therefore be rejected. If it is true that attention cannot be divided between the ears, then that inability does not determine the order in which the items are best recalled. On the other hand, the difference between the two conditions is consistent with the search hypothesis: the fact that recall is superior in the Types condition may be interpreted as showing that the items are more reliably tagged as words or digits than as heard on the left or right.

Lists with no crossings were expected to be easy to recall because the search can be organized on two redundant principles. Both in the Types condition and in the Sides condition (and, for that matter, in the Pairs condition, about which no prediction was made) the fraction correct in lists with no crossings differs significantly at the 1% level from the mean fraction correct in other lists. In these tests Ss were considered to be random and lists fixed, or vice versa; it was not possible to treat both as random factors at the same time.

Discussion

The results of other variants of Broadbent's (1954) original experiment seem to be consistent with the hypothesis proposed here. If the digits in a pair differ in voice-quality but not in subjective location, there is a tendency in free recall to report first the items spoken in one voice and then those spoken in the other (Broadbent, 1956). This result was interpreted as showing that attention cannot be switched rapidly from one voice to another; it may also be interpreted as showing that the items were tagged to indicate voice-quality. Moray (1960) presented one digit to one ear and one to the other at the rate of two pairs a sec., but the stimuli presented to one ear were delayed by 1/4 sec. so that the digits alternated from side to side. In this case it would be possible to retrieve all of the items in temporal order in one pass; and indeed recall in temporal order was if anything superior to recall ear by ear. After the present paper was first written we learned that Gray and Wedderburn (1960) had considered free recall of lists that were like our lists with two crossings except that the words formed a familiar phrase. Their data suggest that the tendency to recall the items by types was at least as strong as the tendency to recall them by sides. Quite recently Bryden (1962) has reported an experiment in which digits were presented in simultaneous pairs over two loudspeakers that were placed close together so that it would be difficult to discriminate the side on which a digit was presented. Items from different speakers were often intermingled in free recall, as though the tag indicating the side were not reliable, but there was little or no reduction in the tendency to take the pairs in temporal order, reporting one digit from each pair, and then to go through the pairs a second time, again reporting one digit (perhaps the same one) from each pair. The idea that dichotic presentation divides the items into two groups was emphasized both by Bryden (1962) and by Gray and Wedderburn (1960).

Experiment II: Discrimination of Recency

The foregoing experiment was one illustration of the consequences of the present point of view. The second example has to do with an experiment on keeping track of several things at once (Yntema and Mueser, 1962, Exp. II). The *S* received a long series of items that fell into six categories—animals, countries, numbers, etc. From time to time he was asked what was the last item he had received in one of the categories.

How might the processor be programmed to find the correct response? At first sight the answer seems to be straightforward: the processor should search for items belonging to the given category and choose the most recent. But how is it to decide which is the most recent? What rules or tests might it use to determine that one item occurred after another? Various hypotheses can be devised, but simplicity is attractive. It is tempting to assume that items in memory carry time-tags that the processor can examine to determine their relative recencies.

Usually one thinks of time and intervening activity as influencing the outcome of an experiment by destroying the contents of memory. Here one finds himself asking whether time, or something correlated with it, may not influence *S*'s response by forming a part of the contents of memory. This is a question that could have arisen from a data-processing analysis of a number of memory tasks. This could at least in theory be a pertinent question in transfer experiments where items on the first list must be discriminated from items on the second.

An initial attempt has been made to see whether items in memory may reasonably be said to carry time-tags and, if so, to get some idea of how precisely those tags are discriminated.

Method

The *S* was given a deck of 220 cards, which he turned over one by one, putting each card face-down before looking at the next. About half were inspection cards; on each there were two words, typed one above the other, which *S* was to read silently. The other cards were question cards. There were two words typed side by side with a large red question mark below them. The *S* was to judge which he had seen more recently and record his decision by checking the left or right column of an answer sheet. The task was self-paced and usually took about 20 min.

A word could appear at most twice—i.e., on an inspection card and then on a question card—and *S* was told of this fact. Sometimes one of the items on a question card was a word that had not appeared in the deck at all. The *S* was instructed that such an item was to be treated as a word that he had not seen for a very long time; so these questions were in effect a test of his ability to recognize the words that he had seen. The first seven cards in the deck were inspection cards; question cards and inspection cards alternated thereafter. There were 107 question cards, of which only the last 75 were used in scoring the results. Six levels of recency were considered; the 75 questions included five instances of each of the 15 possible comparisons of one level with another.

Two sets of words were used. One was a set of 259 concrete nouns that seemed to evoke good visual images—words like *pitchfork, marshmallow,* and *smokestack.* The other was a set of 259 abstract words chosen with the hope that they would be more difficult to recognize—words like *consequently, several,* and *informative.* Four decks were made up using the concrete words and four using the abstract. All eight decks were alike in the sense that any could be derived from any other by simple substitution—i.e., by substituting word *a* on the card or cards where word *a'* occurred, word *b* where *b'* occurred, and so on. However, the sequence of correct choices, left or right, was not the same in the abstract decks as in the concrete decks.

The 32 *S*s were divided into four groups. Each group was assigned one of the concrete decks and one of the abstract. Four of the *S*s in a group were given the concrete deck first, and four the abstract first. All *S*s began with a practice deck in which there were only 31 questions and in which the items were boys' and girls' given names. All three runs were done in a sitting.

The *S*s were 32 students at Tufts University. They were paid for their work.

Results

The data on the concrete nouns are shown in Fig. 2. Let one of the words on a question

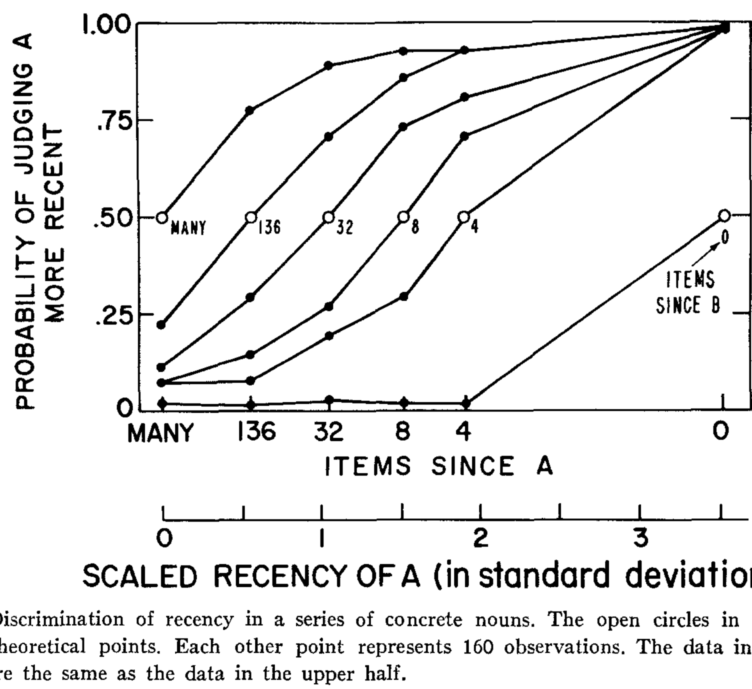

FIG. 2. Discrimination of recency in a series of concrete nouns. The open circles in the middle of the figure are theoretical points. Each other point represents 160 observations. The data in the lower half of the figure are the same as the data in the upper half.

card be arbitrarily designated Item A; the other is Item B. On the horizontal axis is indicated the number of items that S had seen since he had last seen Item A. The number ranges from "many," which means that Item A had not previously appeared in the deck, to zero, which means that Item A had appeared on the immediately preceding card. The items seen since A include the words on question cards as well as the words on inspection cards; and if a word appeared both on an inspection card and on a question card it was counted twice. The parameter of the curves is the number of words seen since Item B, and the ordinate is the observed probability of choosing Item A as the more recent. Thus each curve is a psychophysical function: it shows for a fixed recency of Item B the probability that Item A will be judged more recent. The distance that is required for one of the curves to rise from an ordinate of 0.5 to 0.75 is by definition a just noticeable difference in recency.

To make the figure easier to interpret, the spacing on the horizontal axis has been chosen so that the psychophysical functions would as nearly as possible be normal sigmoids with the same slope—i.e., so that the horizontal axis would be an equal-discriminability scale of recency. The point at zero items since A is an exception: it had to be located rather arbitrarily since these data give almost no information about its position on a scale of discriminability. Scaled values of recency are shown on the auxiliary axis at the bottom of the figure.

In summary, Fig. 2 shows that it is not unreasonable to think of time-tags on items in memory. The discrimination of recency was not sensitive, but it extended far beyond the memory span.

Figure 3 shows the data on the abstract words. Evidently these words were not chosen very shrewdly. The irregularity of the psychophysical functions suggests that this set of words was not as homogeneous as the other set—a suggestion that is confirmed by analyses of variance. And the attempt to choose words that would be difficult to recognize was only moderately successful: comparison of the top curve in this figure with the top curve in Fig. 2 shows that the

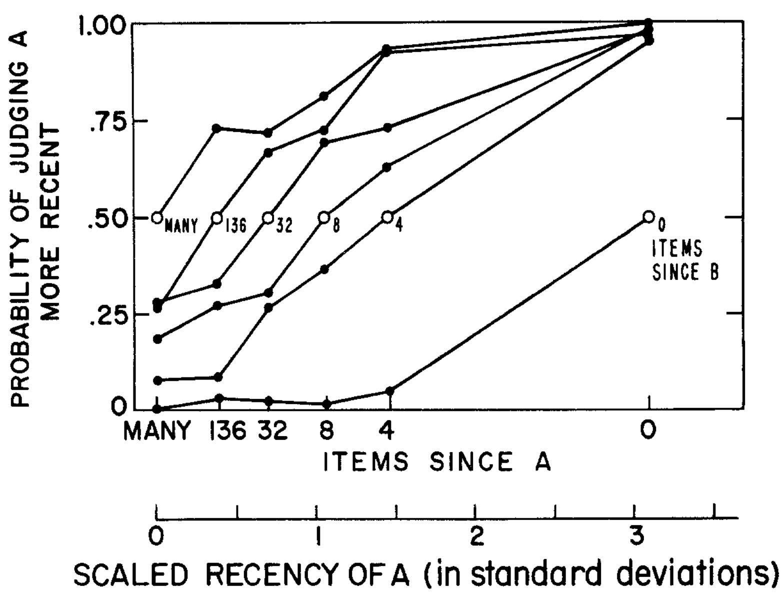

FIG. 3. Discrimination of recency in a series of abstract words. (See legend of Fig. 2.)

total number of errors of recognition increased by a factor of only 1.6. Nevertheless, an equal-discriminability scale was constructed from the data in Fig. 3, and the scaled values of recency are indicated at the bottom of the figure.

It is instructive to compare the two scales. When the unreliable points at maximum recency are discarded, the scale derived from the abstract words is much the shorter of the two. It is not surprising that the point labeled "many" should lie closer to the other points: the abstract words were more difficult to recognize. But there is also a reduction in the scaled distance between an item 136 words back in the series and an item four words back. The change appears to be reliable. No direct test of its significance is available; but when questions about words that had just appeared or had never appeared are ignored, the difference in the fraction of the questions answered correctly in the two kinds of decks is significant at the 1% level. How this result is to be interpreted is not obvious. Must it be taken to show that the time-tags on words that are difficult to recognize are less precise than those on words that are easy to recognize?

Summary

We have suggested that it can as a practical matter be useful to regard recall as a search through memory and consider how a data-processor might be programmed to find the desired response. A demonstration experiment in which recall was examined by what might be called the "method of the missing item" was discussed to point out that a process of searching and testing may reasonably be assumed to intervene between memory and overt response.

Two examples were given to illustrate the suggestion that this point of view can be helpful. The first was an experiment on immediate recall of six items presented two at a time, one to one ear and one to the other, at the rate of one item per ear every half-second. Each pair of items that were presented together consisted of a word and a digit, and the side on which the word was presented varied haphazardly from pair to pair. Recall was more successful when *S* was instructed to report the items of one type and then the items of the other type than when he was instructed to report the items heard on one side and then those heard on the other. The conclusion was that Broadbent's discovery about the preferred order of recall should not be ascribed to an inability to shift attention rapidly from ear to ear; the task is better described as a problem of data-retrieval.

The second example was an experiment on an unconventional question that arises when recall is viewed as a search by a data-processor: may items in memory be assumed to carry time-tags that the processor can examine to determine which occurred more recently? The *S* received a long series of words, alternately seeing new words and judging which of two he had seen more recently. An equal-discriminability scale of recency in a series of concrete nouns was constructed and was compared with a similar scale for a series of abstract words.

References

Broadbent, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, **47**, 191-196.

Broadbent, D. E. Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, 1956, **8**, 145-152.

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Broadbent, D. E., and Gregory, Margaret. On the recall of stimuli presented alternately to two sense-organs. *Quart. J. exp. Psychol.*, 1961, **13**, 103-109.

Bryden, M. P. Order of report in dichotic listening. *Canad. J. Psychol.*, 1962, **16**, 291-299.

Egan, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58**, 955-991.

Feigenbaum, E. A. The simulation of verbal learning behavior. *Proc. Western Joint Computer Conf.*, 1961, **19**, 121-132.

GRAY, J. A., AND WEDDERBURN, A. A. I. Grouping strategies with simultaneous stimuli. *Quart. J. exp. Psychol.*, 1960, **12**, 180-184.

MILLER, G. A., GALANTER, E., AND PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MORAY, N. Broadbent's filter theory: Postulate H and the problem of switching time. *Quart. J. exp. Psychol.*, 1960, **12**, 214-220.

YNTEMA, D. B., AND MUESER, G. E. Keeping track of variables that have few or many states. *J. exp. Psychol.*, 1962, **63**, 391-395.

(Received April 1, 1963)