

Memory Without Organization: Properties of a Model with Random Storage and Undirected Retrieval

THOMAS K. LANDAUER

Bell Laboratories, Murray Hill, New Jersey 07974

A very simple spatial model of memory storage and retrieval is described, analyzed, and discussed. The postulated memory is without organization in the sense that neither the place of storage nor the order of search during retrieval is influenced by the nature of the information being stored or retrieved. The memory consists of a three-dimensional space containing a large number of homogeneously distributed loci at which data may be stored. Data received near each other in time are stored at nearby locations. Access is by an undirected expanding-sphere search. The model exhibits a wide variety of quantitatively and qualitatively humanlike behavior with respect to both standard learning and forgetting paradigms and with respect to frequency effects and other phenomena in word processing.

In less than a second, the average adult can understand any one of over 100,000 different spoken or printed words, and can retrieve with similar speed the names, locations, nature, and other facts about hundreds of thousands of objects, sounds, smells, tastes and feels.

It is tempting to assume that because its apparent functions are enormously fast and powerful, the basic structure and mechanisms of human memory must be enormously complicated. In particular, it has sometimes been proposed (e.g., Norman & Rumelhart, 1970; Quillian, 1967) that data are stored in intricate, purposeful arrangements, that on receipt of new data the system can reorganize its contents to make them more accessible, and that in retrieval the system searches in a systematic manner.

In general, the object of the work reported here was to explore the kinds and ranges of phenomena that might occur in a memory that had no such elaborate internal structure. Important regularities often arise in nature from the random concatenation of large numbers of basically

Portions of this paper were first presented at the annual meeting of the Psychonomic Society, St. Louis, MO, November, 1972. I thank J. K. Landauer, C. L. Mallows, D. E. Meyer, N. Nilsson, and S. Sternberg for very helpful advice. K. I. Ainslie and W. J. Blair for assistance and G. L. Felfoldy, J. Juola, T. Nelson and B. L. Wattenbarger for comments on the manuscript. Reprints may be requested from Thomas K. Landauer, Bell Laboratories, Murray Hill, NJ 07974.

simple and unsystematic events. The question I wish to pose is whether many of the important features of human memory might not reflect the outcome of essentially random underlying processes. One way to approach this question is to try to formulate a model that relies only on random basic processes and to see what human-like properties it may manifest.

Here, I consider a particular model of an intrinsically nonorganized memory store with an undirected retrieval mechanism. I first describe the model and then demonstrate its operation in a number of applications. The applications are chosen both from traditional verbal learning paradigms and from phenomena involving the retrieval of lexical and semantic information stored in long-term memory.

Before describing the model and its applications, however, I must outline the limits of the present discussion. First of all, this is a model of memory only, not of cognition or performance in general. Obviously the total human information processing system contains other important components. In all likelihood, it has a separate mechanism for preprocessing or analyzing perceptual input before or in interaction with memory consultation. In all likelihood it has a separate mechanism for combining information from memory or from perception and memory in order to make inferences or generate novel ideas and actions. It probably has components equivalent to linguistic or perceptual parsers, logic processing units and buffer registers. It must have a *homunculus*, something equivalent to a computer's executive routine, to coordinate all of these. But we will not be directly concerned with these parts of the machinery. Rather, we will be concerned with what known features of behavior could result from the way the basic memory storage and retrieval component is assumed to work.

Second, since the model is potentially relevant in some way to all phenomena that involve memory, some limitation in its applications was necessary. The sample of cases to be reported was chosen to try to satisfy two main goals; (a) to test the model against well-established quantitative facts, and (b) to explore the diversity of phenomena to which it could be related.

A third, related, limitation needs to be clearly understood. The discussion leaves open the possibility that some facts that could be attributed to memory are properly to be explained by functions performed by other parts of the system. For example, an explanation for the benefit of categorization in free recall might be sought in a strategy that the subject adopts in using his memory, rather than in properties of the memory system itself. This distinction will be made more explicit later.

It is necessary to put these demurring remarks in perspective. The fact that the model is incomplete, and like all mortal theories, probably wrong, is not what is interesting about it. The model is not offered as a

full and true account of nature, but primarily as a heuristic exercise in the evaluation of intuitions concerning basic mechanisms. Therefore, what is interesting and important about it is that with very minimal structure, no self-organizing abilities, and a search process that is independent of the target, the model performs a wide variety of lifelike memorial functions.

DESCRIPTION OF THE MODEL, I: STRUCTURE AND STORAGE

The memory store is assumed to consist of a large and—as a working hypothesis—three-dimensional space, containing a very large number of small storage loci. These storage loci correspond roughly to the storage registers of a computer memory, and are homogeneously distributed throughout the space. The amount of information that can be stored in a single locus is assumed to be variable, but to have some as yet unspecified maximum. What the stored information represents is also assumed to be variable, just as it is in a computer. Fortunately, the question of how much of what is in a single locus, while interesting, does not need to be answered in advance in order to derive many important properties of the model. Where necessary for the applications to be described, certain aspects of storage locus content will be specified. (For an initial grasp, the reader can imagine that a single locus can store a whole word, both terms in a paired-associate pair, an atomic proposition, or the information needed to name a simple visual or auditory pattern.) Data are entered into this memory successively in time as the system receives—and perhaps abstracts, encodes, interprets or “parses”—experience. It is assumed that at any given point in time there is just one locus into which new data are written. Data entry occurs only at discrete points in time, but does not necessarily occur at every opportunity. Presumably, data entry is under the control or influence of other parts of the system representing such factors as attention, input preprocessing and the like. The locus available for new data entry at a given moment may be imagined to be at the tip of a pointer. The pointer moves slowly, erratically, and ceaselessly through the space in a three-dimensional random walk. In each successive epoch, the pointer may or may not move one step in any direction. During each epoch, data may or may not be stored (and/or a search for previously stored data may be carried out as described in the next section.) The location of a particular datum relative to another approaches independence as the time between two acts of data entry becomes sufficiently long. However, if the time between two data entries is short, then the two storage loci will be near one another. This is the sole nonrandomness in the entry mechanism. The import of this assumption will become clear as we go along. The structure, or rather the lack of structure, of the memory and a typical path for the recording pointer are schematically illustrated

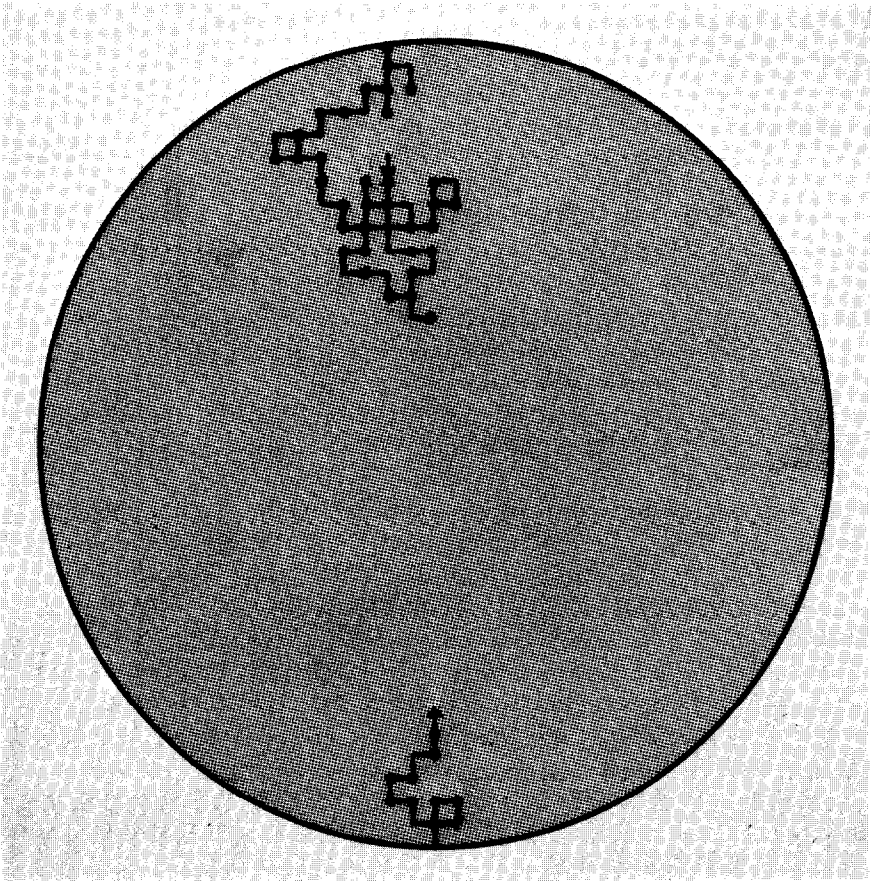


FIG. 1. Schematic diagram of memory structure and movement of the recording pointer. The storage space is assumed to be three-dimensional with a large number of small storage loci homogeneously distributed within it. The point at which data can be entered or a search initiated at any moment describes a random walk. When an edge is encountered the pointer reenters the space from a symmetrically opposite point. Dots in the figure represent successive pointer locations.

in Fig. 1. (The figure, of course, is in only two dimensions, instead of the three postulated in the model.)

One accessory rule is required for the random walk of the pointer. When the pointer encounters an edge of the storage space, it re-enters the space again from a symmetrically opposite point. This is also illustrated in Fig. 1. This rule assures that momentary locations of the pointer are homogeneously distributed over the space.

DESCRIPTION OF MODEL, II: ACCESS

When the homunculus (i.e., the rest of the information processing system) attempts to use stored information, an undirected search is begun. A signal containing some potentially identifying tag for data being sought is sent out from the pointer. If and when the signal encounters a memory location with that tag as part or all of its contents, all of the information in that location is returned to the pointer and thence to the homunculus. The process by which the contents of the signal are matched with those of storage registers is of course important; it involves critical aspects of encoding and pattern recognition. But specifying the details of this process is beyond the scope of the present model. Instead, effective mechanisms by which a partial representation can be matched to a full record or another partial representation is simply assumed. The search signal spreads uniformly in all directions, as an expanding sphere, at a constant rate, up to a limited radius generally smaller than the radius of the entire memory. The search pattern inherent in this process is partly serial, and partly parallel. Whether it is self-terminating or exhaustive depends on whether the homunculus chooses to use information as soon as it is returned or wait until the search has reached either its natural limit or some "deadline" limit set by the homunculus. The order in which memory loci are searched is random with respect to the content of the search signal, depending only on the current random locus of the pointer; first the memory location of the pointer is examined, then all those loci immediately adjacent to the pointer are searched simultaneously, then all those just one step further away are searched simultaneously, and so forth. The region searched at any given instant is thus a spherical shell, expanding from the pointer and growing to some maximum size. The total region that can be reached from one pointer location will be called the *search space*. See Fig. 2. Obviously, the larger the search space relative to the entire memory store the more likely the system will be to examine any random storage location during a single search. For example, if the space that could be searched were as large as the total memory store, then all previously stored data would be accessible. It is assumed, though not used much here, that varying qualities of information contained in the search signal and storage registers might lead to differing effective radii of the search space, giving rise to varying probabilities of retrieval for different circumstances, kinds of material or modes of access. (This presumably would result from a progressive weakening of the search signal as it spreads. Information-rich signals could excite appropriate storage registers at greater distances.) The maximum radius of search might also be under strategic (homunculus) control, for example in response to

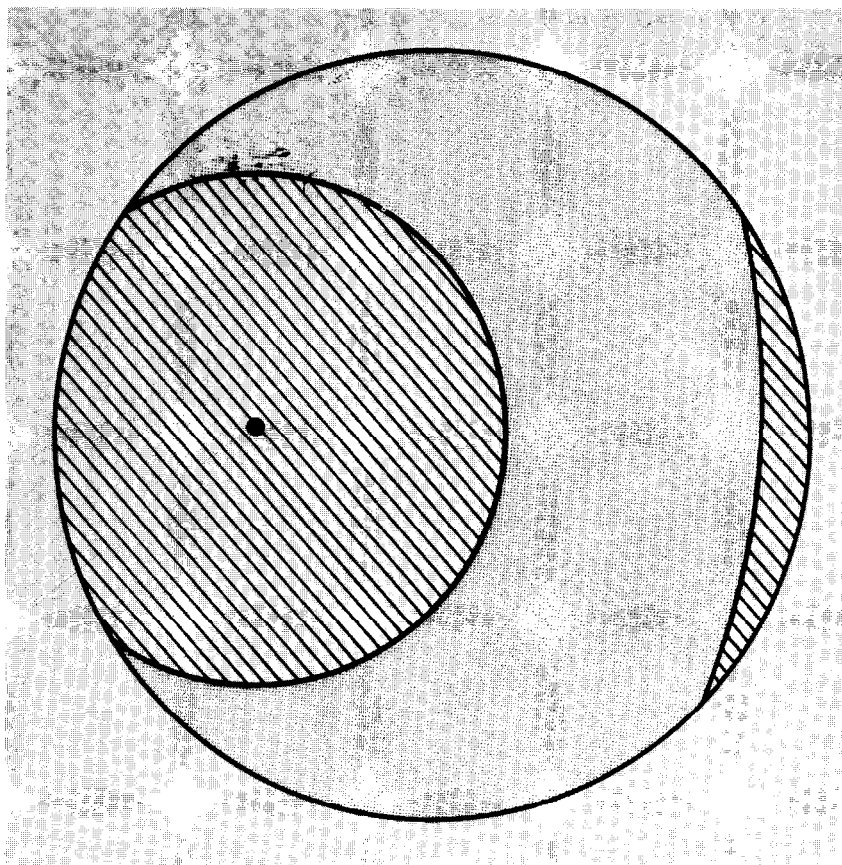


FIG. 2. The possible search space for retrieval at a given moment is defined by a region with radius r . Search proceeds outward from the current locus of the pointer as the surface of an expanding spherical shell. The pointer's location is not influenced by either the occurrence or outcome of a search.

speed-accuracy demands of a particular task. For present purposes, however, it is merely assumed that the maximum search space is constant and smaller than the entire storage space. This means that a datum that is recorded only in a single locus will have a probability less than one of being found during search from a single pointer location. The exact radius of the search space becomes a parameter in the formal mathematical representation of the theory, and in the simulations which will be reported below.

Another rule is needed to take care of occasions when the search signal encounters a boundary surface of the storage space. The rule used is analogous to the one adopted for the random walk of the storage

pointer. The same volume of storage space is always searched, the additional volume beyond the boundary being taken from the symmetrically opposite region of the space. This is also shown schematically in Fig 2. (Although we have been regarding the storage space as a sphere, the boundary rules actually convert it to a closed space. It is the set of formal rules, of course, that determines the model's behavior, not the geometric image used for exposition.)

The path followed by the pointer is assumed to be independent of the retrieval process. That is, the pointer travels inexorably in its random walk as a function only of time. Searches are executed with the pointer location as their center, but the path of the pointer's movement is not influenced either by the occurrence of a search or by its outcome. For example, the pointer *does not* move to the storage location where a fact is found. The time required for memory search is assumed to be a function only of how far the locus of stored information is from the pointer. The spreading search is presumed to be a time-consuming process, like a moving wave around a point of disturbance in a fluid. The time to retrieve information from a given storage location is a simple linear function of how far away from that storage location the pointer was when the search began.

While the foregoing is a reasonably complete verbal description of the system, at least insofar as is required to derive how it will work for the various examples we wish to consider, all of its properties as a dynamic memory storage and access device will hardly be apparent to the reader from such a spare outline. Rather than give more detail immediately, however, what I shall do is describe the operation of the memory in particular tasks, and show how its functions arise from the assumed mechanisms as we proceed.

However, before doing this, let us review briefly the general nature of the posited mechanisms of the memory system. First, as to internal structure, it has hardly any. The system does not organize the location of stored information with respect to the nature of that information. Any kind of data can go anywhere. There are no separate regions for semantic or episodic, lexical or encyclopedic, familiarity or associational data. Such differences must be represented, if at all, in the varying contents of individual registers or in the number and spatial distribution of data in registers resulting from different temporal patterns of input. The density of storage of data is potentially equal everywhere. Only history will determine the actual location of anything within the memory. Any organizing strategies of the subject must be applied prior to entering data in memory or after its retrieval. The memory itself is postulated to be entirely passive. Once entered, its contents cannot be either rearranged or systematically erased. Considerable order may exist within the memory after it has stored a great deal of data, but that order will be only a

reflection, and a distorted one at that, of the order of experience. The order will be more like that of the detritus in a garbage can than the entries in a dictionary. A garbage can is neither self-organizing nor structured, yet one finds coffee grounds and orange peels near each other, things used long ago on the bottom, and so forth. It is this sense of order being imposed only by historical order of entry that is implied by calling this a random storage model.

Let us turn, now, to some specific properties of the model. The properties that have been studied in some detail, and which will be described here, fall into two main categories. To begin with, we consider applications to straightforward human learning and forgetting phenomena. These phenomena, in which important quantitative relations have been established beyond reasonable doubt, provide a necessary first testing ground for any theory of memory.

In this category, we will see how the model provides quantitative accounts for response probability and latency in simple learning, in forgetting after single and multiple learning trials, and in spacing of practice effects with both long and short retention intervals. A number of qualitative properties that are like ones often observed in people will also be mentioned. Among these are oscillation of retrievability, tip-of-the-tongue and slip phenomena, and the "permanence" of memory. Transfer, retroactive and proactive inhibition, and interference generally will be discussed in passing to show how they fit into the model. However, these topics will not be developed fully as they are sufficiently complex to warrant separate treatment. For many of the phenomena discussed in this section "miniature theories" have previously been proposed that give good-to-excellent descriptions of reality. But the present model provides a reformulation that brings an unusual number of these phenomena under the same roof. Similarities to and differences from closely related theories will be discussed.

Next we will consider paradigms involving the retrieval of well-learned facts. Here the focus will be on the processing of words; the domains that have come to be known as "semantic" or "lexical" memory. Among this general class of phenomena, those involving the effects of usage-frequency of words on recognition time and accuracy will be the chief concern. These phenomena contain the kind of regular, quantitative relations that can challenge a model's ability to make detailed predictions. We will also, however, consider some qualitative findings in this area, including association and polysemy effects on lexical decision time.

Following the derivation and discussion of specific properties of the model, there will be some discussion of the general nature of the model and the phenomena to which it is potentially applicable. First, aspects of

its plausibility as a model of brain function will be briefly discussed; then the ways in which the model and the theoretical strategy it represents could be expanded to encompass other phenomena. Finally, inadequacies and incompleteness of the model in its current form will be commented upon.

LEARNING

First consider how the system would learn. If the same datum is encountered on more than one occasion, a record of that datum is ordinarily stored in more than one place. Thus the theory is a multiple copy theory (cf. Atkinson & Shiffrin, 1968; Bernbach, 1969; Bower, 1967; Hintzman & Block, 1971). If successive encounters with the same datum occur very far apart in time, then their corresponding multiple storage loci will be at essentially random places relative to one another. If, in addition, it is assumed that the number of possible storage locations is very large, then each new encounter will leave a new record in a new place, since the probability of its being in the same place as an old record will be vanishingly small.

The probability of correct recall on a given trial will depend on the probability of finding at least one of the copies within the search space surrounding the pointer. This probability will obviously grow with the number of different distinct loci at which the same information is stored. Indeed, it is fairly easy to show that this growth of probability follows the well-known constant-proportionality function of simple learning. (Readers unfamiliar with standard experimental learning paradigms and results are referred to Atkinson, Bower & Crothers, 1965; Hilgard & Bower, 1966; or Kintsch, 1970.) Rather than give a mathematical derivation, I will introduce a simulation procedure that was found convenient for some of the later studies, and use it to demonstrate simple learning as well (however, the formal proof is outlined below). The simulation differs from the model described verbally above only with respect to a few unimportant geometric features and simplifying assumptions. The simulation assumes a large three-dimensional cubical array in place of the spherical store. The pointer walks through this space by the addition, with equal probability, of -1 , 0 , $+1$ in each of three dimensions in each epoch. It thus moves, with equal likelihood, to one of 27 cells. The *distance* of these cells from the last locus is from zero to three units in the city-block metric implicit in this procedure. Data may or may not be stored during a given epoch. Also during each epoch there may or may not be a retrieval attempt. In a given retrieval attempt, search is made only from the single current location of the pointer—if the desired information cannot be found within the search space, an error is scored. The search space is defined by a “city-block radius”, r , the sum of the

absolute number of city-block units in all three directions. [For example, from the point (0,0,0) the search space with radius 2 consists of 25 loci; the point (0,0,0) itself plus six points lying one step away, i.e., (1,0,0), (0,1,0), (0,0,1), (-1,0,0), (0,-1,0), and (0,0,-1), plus 18 points lying two steps away, i.e., (1,1,0), (1,0,1), (0,1,1), (-1,-1,0), (-1,0,-1), (0,-1,-1), (1,-1,0), (-1,1,0), (1,0,-1), (-1,0,1), (0,1,-1), (0,-1,1), (2,0,0), (0,2,0), (0,0,2), (-2,0,0), (0,-2,0), and (0,0,-2).] The simulated reaction times described below consist of the shortest city-block distance between the pointer and the closest storage location at which desired information can be found.¹ (In more complete analyses of the model, the pointer might be allowed to move and try again some number of times before giving up a search. This elaboration gives the model additional interesting features. However, it requires the introduction of a new parameter representing the relation between the time occupied by a movement of the pointer and by the rate of spread outward of the search signal. It seemed more interesting for initial exploration of the model to deny it this freedom.) A final simplifying assumption adopted for simulations was that the number of storage locations be infinitely large. One result (and the implementation) of this assumption is that information is essentially never overwritten. Some of the interesting consequences of relaxing these simplifying assumptions, of allowing search to go one over more than one pointer step, and of allowing overwriting, will be discussed informally below.

In simulations, the memory registers initially contained only zeroes, and on each learning trial the image of an integer "1" was placed in the register at the current pointer location. During retrieval, the search mechanism tested to see if any copy of "1" was stored within r city-block units of the pointer, scored itself successful if at least one was and unsuccessful otherwise. Similar functions would be obtained, of course, for storing other kinds of information, like two associated words, or a simple proposition, and searching for a match between search signal and some component of the stored image.

For predicting probability correct, there are essentially two independent parameters in the model: the size of the search space relative to the total memory, and the rate at which the pointer moves. (For reaction time applications there are two additional time scale parameters, one for the rate of spread of the search signal and one for input-output time in-

¹ The program did not actually search its memory space. This is a slow procedure for a serial-processing computer, despite the fact that it could be a fast one for the partly parallel recognition mechanism assumed in the model. Instead, the program kept track of where each datum would have been stored and computed what would have happened in an expanding sphere search from a given pointer location. The actual stimulation program, excluding input-output, random number generating, and arithmetic subroutines, contains less than 100 Fortran statements. Listings may be obtained from the author.

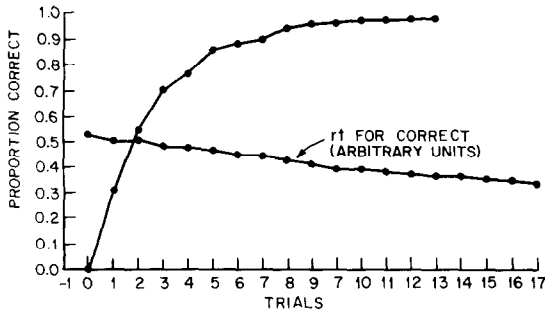


FIG. 3. Simulated learning curve for widely spaced trials. [Total memory equals $12 \times 12 \times 12$; search radius equals seven steps ($\alpha = .33$); random choice of search and entry locus on each trial; $N = 1000$.]

dependent of memory access.) In simulations these parameters were fixed by specifying a total size for the memory, a radius for the search space, and the number of random walk epochs (steps) between successive combined entry and/or retrieval attempts (trials). Values of these parameters and N , the number of simulated subjects, are given in relevant figure captions. (For reaction-time data, the time scaling parameters were fit to the data.)

Figure 3 shows a simulated learning curve with infinitely spaced anticipation trials, with each trial composed of first a search then a new entry of the same datum. Both increases in probability and decreases in reaction time are shown. It is noteworthy that reaction times continue to decrease long after perfect performance is approached. The memory produces perfect recall as soon as there exists at least one copy within r units of every point in the memory space. This will happen long before all of the spaces are occupied by the same information, but the reaction time will still continue to decrease as the same fact is stored in more and more places.

A prime determinant of the rate of learning is the proportion of memory that can be *searched*. This quantity would be expected to vary with differences in input modality or quality, or other factors that might affect the strength of the search signal, or the amount of retrieval time available. Learning rates would also be modulated by attentive factors that would affect whether a given potential input was entered or not and if so how it was coded.

FORGETTING

In this model, forgetting—the loss of the ability to retrieve a given fact—could occur by three separate processes. The first, a relatively short-term process, results from the random walk of the pointer, the sec-

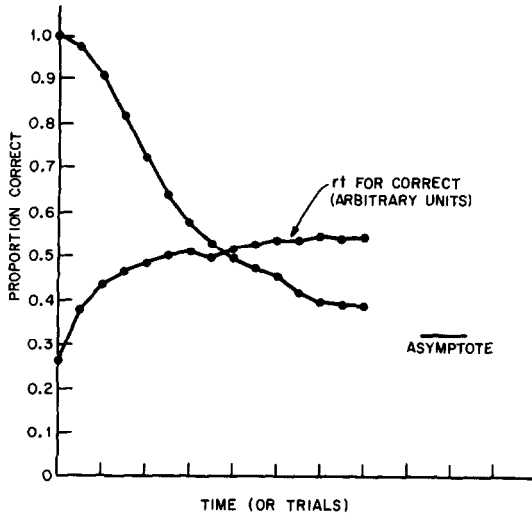


FIG. 4. Simulated forgetting curve for a fact encountered once. [Total memory equals $12 \times 12 \times 12$; search radius equals seven steps ($\alpha = .33$); two random walk steps between trials; $N = 1000$.]

ond from interference, and the third from overwriting. Take these in turn. Consider a datum that has been entered just once in a single memory locus. Immediately after storage, the pointer is certain to be within the search radius of the new datum, and thus the probability of retrieving it is one. However, as time goes by, the pointer will tend to wander away, and the probability that the datum will be within the search radius will diminish. Figure 4 shows the decline in probability of recall of a once-stored item as a function of time (actually the number of steps in the random walk of the pointer). The curve is much like typical ones for human short-term forgetting (e.g., Peterson & Peterson, 1959). Certain features of it are worth noting. The decline in probability is not always smooth. Figure 4, giving averaged data for one thousand simulated Ss, shows a wobble, which while not present in simulations with all sets of parameters, is not atypical. With certain choices of parameters, this wobble in the forgetting curve can be sufficiently pronounced to resemble reminiscence. The wobble results from the wrap-around feature of the memory space; the probability of the pointer getting back in range by going "around the world" rises with time while the distance it has wandered "away" goes up. Unless these two effects are exactly compensatory, forgetting will not be smooth, but rather will contain a damped oscillation, and may even be nonmonotonic. Suggestions of such a wobble have appeared in human forgetting data (e.g., Peterson, 1966).

As time becomes very great, the pointer's location becomes uncor-

related with the point of storage. When this is the case, the probability of recall will be simply the ratio of the volume contained in the search space to the volume of the entire storage space. Thus, the asymptotic probability of finding a once-stored item will be greater than zero. (In particular, it equals the learning rate parameter which characterizes the increase per trial of the curve shown in Fig. 3 for learning with infinite spacing of trials. This will be discussed below.)

Now consider the form of forgetting curves after multiple exposures and entry. First of all, the asymptote toward which such forgetting will head as a result of the walk-away process will clearly be different after two entries than after one, because the probability of finding either one of two places where the same information is stored will be greater than that of finding only one. Figure 5 shows the decline of memory for a given item after it is presented 1, 2, 4 or 8 times in successive trials—as generated by the simulation program. For comparison, a similar figure giving data from human subjects is also presented. No formal parameter fitting was done, but parameter values were chosen that would illustrate the model's human-like qualitative features.

The second forgetting process involves competition with other stored information. If new entries are added such that different information is likely to be returned from searches with the same search signal tag, then unambiguous recovery of the old information will become more difficult. Thus, for example, if first AB is stored, and then later AC, search to

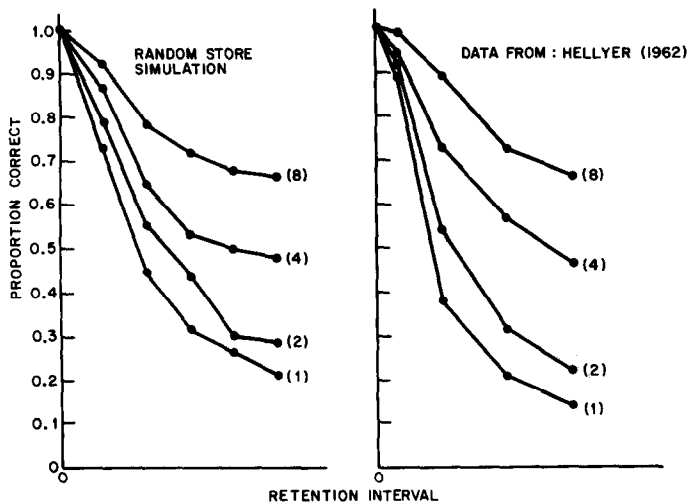


FIG. 5. Simulated forgetting curves for a fact encountered on one to eight successive occasions, along with data from a similar experiment with humans. [Total memory equals $7 \times 7 \times 7$; search radius equals three steps ($\alpha = .18$); two random walk steps between trials; $N = 1000$.]

match A could return either AB, AC or both. Depending on whether the homunculus decides to terminate search at the first match or to search as far as possible, and on how it decides to respond—or not respond—when it receives equivocal output from its memory, one could expect the occurrence of intrusion, confusion or omission errors. In this process, old memories are not lost or erased, just hidden among new ones.

It should be apparent that the mechanism just described concerns the way in which retroactive inhibition affects the system. The system will also be subject to negative transfer and proactive inhibition effects for the same reasons. In general, the presence of AB records in memory will increase the number of AC entries required for a given probability of retrieval of C given A. Trials on AC would eventually lead to a sufficient concentration of AC entries, relative to AB entries, near the pointer to give any desired degree of learning. However, as time elapses after AC learning, the pointer will tend to move away from the area of concentration of new records. This will not only lead to “walk-away” loss of AC—as in the first process—but will also tend to bring the pointer into range of more competing AB records causing “recovery” of AB memory. Thus, prior learning of AB will increase both the rate and total amount of forgetting after the learning of AC to a criterion.

There are many other implications concerning transfer and inhibition phenomena inherent in the model. A full treatment of these matters will require consideration of a large number of details. Among these are the effects of list tagging (e.g., “List I” entered as part of each item), other means of differentiating age of memories (e.g., by relative retrieval times, or stored context information), the conditions under which similarity of stimuli will result in confusions in the match process, and the precise outcome of simultaneous or near simultaneous retrieval of different information by the same match signal. Adequate discussion of interference phenomena, and attention to the large body of relevant data available, is beyond the scope of this paper and will be postponed. It should be noted here, however, that the random store model is not without significant potential to suggest possible solutions for many problems in this area.

In a complete form of the model, forgetting could result from still a third separate process. This is overwriting. If the pointer returns to a storage location that has previously been used and enters something new there, it is reasonable to assume that the new information will replace or degrade the old. This would provide a separate—essentially pure decay—mechanism for long-term forgetting. This mechanism has been ruled out in the computer simulations reported here by allowing any number of facts to be stored with the same spatial address, in effect making the number of storage locations infinite. In a fuller account,

while the short-term forgetting curve would approach the learning-parameter asymptote, forgetting would not cease at that level, but would continue at a very slow rate forever. Such a forgetting function—with a rapid loss at first, leveling out to a slow but steady loss thereafter, is in good accord with what is known about human forgetting.

Certain other interesting qualitative features of human forgetting are also inherent in the model. Given either a single record of a datum, or multiple records of the same datum, the system will show features of memory often taken as evidence that facts once learned are never forgotten. First consider a datum stored only once. While forgetting will proceed rapidly to the walk-away asymptote, and slowly thereafter, there will remain a finite probability of finding that datum and remembering for the life of the organism. While some once-stored items will be irrevocably lost by being overwritten, some will be there forever. Thus, anecdotal reports of subjects having recalled a particular bit of minutia from the past many, many years later, would not be surprising.

Next consider the fate of a fact that has been well learned over many trials. This information would be stored in many different locations. Even after a long time such that many are overwritten, some copies would be likely to remain. If the system began to learn the same fact again, it would show savings for any fact still stored in at least one location.

Things forgotten by either the walk-away or interference processes would not be forgotten permanently. Memory for a given item might oscillate fairly strongly. What could be remembered during one short period of time might be “forgotten” during the next, and “remembered” again at still a later time, as the pointer moves from and to places from which a given locus is found without competition. This would give rise to the temporal fluctuations in recall that occur both in everyday life and in the laboratory (Buschke, 1974; Tulving, 1967).

The multiple storage feature of the model allows different pieces of information related to a common subject to be stored in different places. Recall would therefore often be partial. One result is that the system would exhibit “tip-of-the-tongue” and “feeling of knowing” phenomena (Brown & McNeill, 1966; Hart, 1965; Freedman & Landauer, 1966). For example, after 100 tongue-twisting pages of Dostoevsky's *Crime and Punishment*, one might begin to read and store only “R ” instead of the protagonist's full name. Because separate encounters with that name are stored in different places, rather than being filed together as they might be in a more organized memory, a search for “Dostoevsky's man with an axe” might turn up the information that “there are storage registers containing related information”, “the name is like “R ,” and nothing else. Later, with the pointer in a different place, the same search tag might find “Roskolnikov.”

Alternatively, having found only "R . . .," the subject might try to fill out the whole name by attempting a match to "R . . ." This might return the desired name, nothing, or a similarly spelled or pronounced name like "Rosenkranz." Indeed, through such partial match mechanisms, the system would provide many opportunities for false recognitions, *not quite* random slips of the tongue, malapropisms and the like, based on phonetic, visual, historical or semantic similarities.

Similar considerations would explain differences between recognition and recall. Recognition requires only finding a locus containing the stimulus; any return would indicate presence of the stimulus in memory. However, recall would require finding a locus with the stimulus plus the other information needed to make a correct response. Free recall would require finding loci with both list or context tags and to-be remembered items. (Note that in this formulation even recognition requires search and retrieval, it is just easier search and retrieval.)

SPACING PHENOMENA

The model shows several effects of spacing of practice. To see why performance should benefit from spacing, consider Fig. 6. If the same item is entered in two places, one right after another, the pointer will

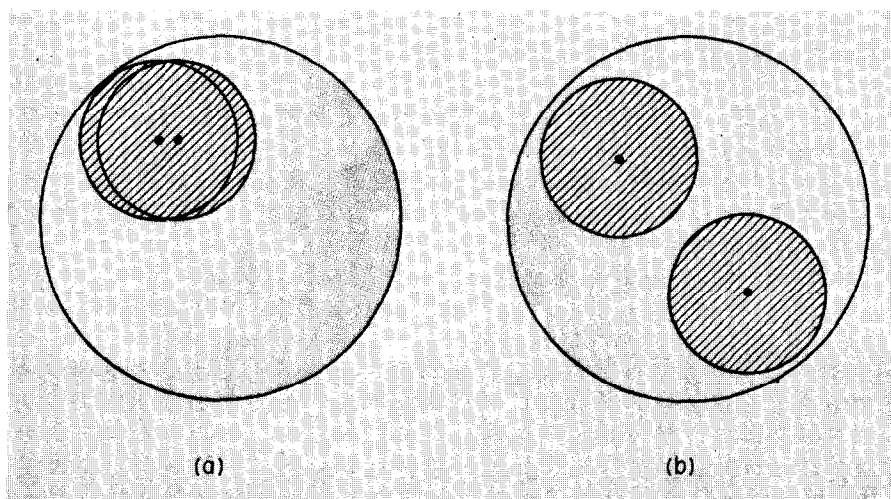


FIG. 6. Illustration of the mechanism of the spaced-practice effect. The dots represent storage loci for two records containing the same information, in (a) resulting from closely massed trials and in (b) from widely spaced trials. Shaded regions show the total area from which one or the other locus can be reached by a search. If the pointer location is uniformly distributed over the storage space, the ratio of shaded to unshaded regions gives the probability of successful retrieval. This corresponds to the situation after a long retention interval following the second trial.

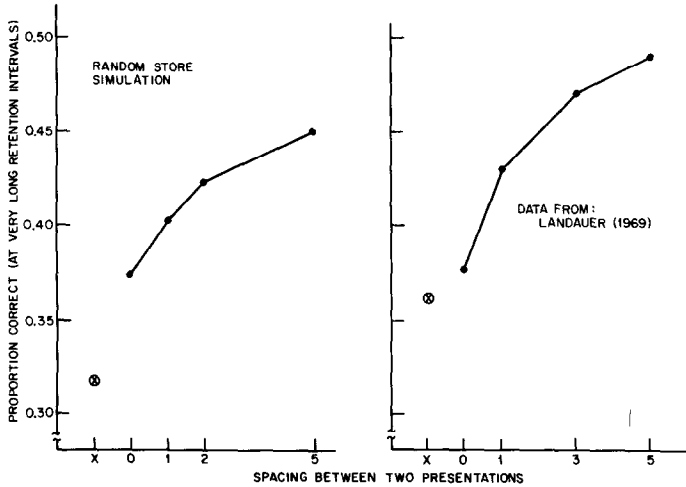


FIG. 7. Simulation of probability of retrieval at long retention intervals as a function of the spacing of two learning trials. The value at x is for a single trial. [Total memory equals $12 \times 12 \times 12$; search radius equals seven steps ($\alpha = .33$); one random walk step per spacing interval, test pointer locus randomly chosen; $N = 2000$.]

have moved very little in between. Therefore, the two storage loci occupied by the same information will be quite close to each other. This means that the total region from which one or the other of them could be reached by a search will be very little different from the total region from which a single storage location could be reached. See Fig. 6a. Thus two trials in immediate succession will not change the asymptotic probability of retrieval very much. On the other hand, if a large amount of time is allowed to elapse between the first and second entry of the same datum, then the two replicas, on the average, will be stored at a considerable distance from each other. In this case, the region where the pointer can be in order for a successful search to occur is as much as twice as large as for a single entry. See Fig. 6b.

Figure 7 shows the results of simulating the effect of two trials with varied spacing. The dependent measure is the probability of finding one or the other or both of the two storage loci after a long retention interval, that is, after the location of the pointer has become random within the storage space. Spacing provides considerable benefit. Again without detailed parameter fitting, results from an experiment with human subjects are shown in the second panel of Fig. 7 for comparison. (The overly large effect of 0 spacing in the simulation results from too large a minimum random walk distance relative to search radius.)

These results, as stated, are for retrieval after long retention intervals. The situation is not necessarily the same at very short retention in-

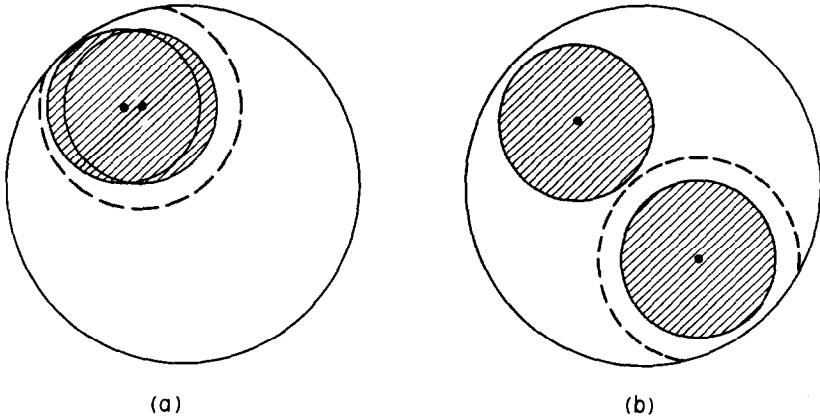


FIG. 8. Illustration of the effects of spacing for short retention intervals. As in Fig. 6, shaded regions represent the areas from which one or the other of two storage loci can be reached by a search. The area within the dotted lines defines the possible loci of the pointer some arbitrary short time after the second of two massed (a) or spaced (b) trials. Unlike the situation after long retention intervals, the probability of retrieval may be greater for massed trials.

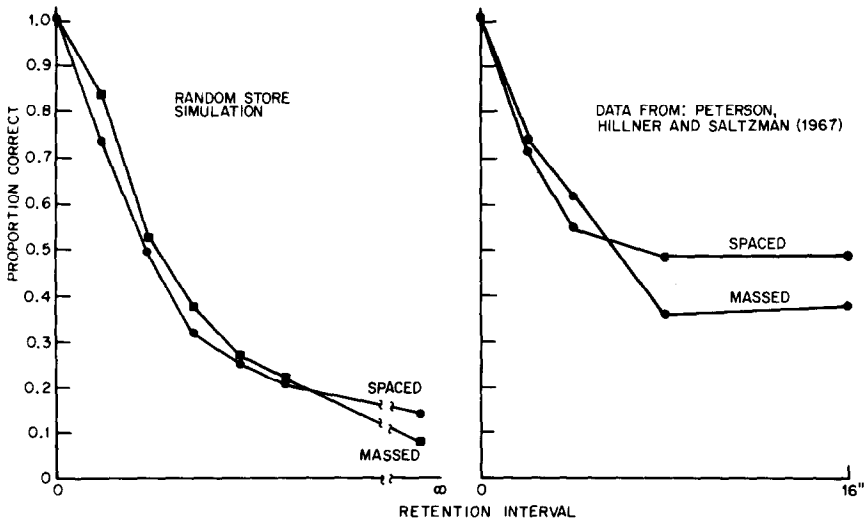


FIG. 9. Simulation of forgetting after massed and spaced trials with parameters chosen to generate a crossover of spacing effects as a function of retention interval. Data from a human learning experiment exhibiting the same qualitative features are shown in the right-hand panel. [For simulation, total memory equals $10 \times 10 \times 10$; search radius equals three ($\alpha = .06$); one random walk step between learning trials for massed condition, randomly chosen entry points for spaced condition; two steps between retention trials; $N = 1000$.]

tervals. Indeed, the model predicts a possible interaction between the effects of spacing and the length of the retention interval. To see why this is so, consider Fig. 8. The broken curves surround areas to which the pointer could move within a short time after the second of two trials. As contrasted to the uniform distribution of possible pointer locations for retrieval assumed in Figs. 6 and 7, here the region from which one or another of the two locations can be reached is larger for two massed trials than for two spaced trials. Put another way, when the retention interval is short, the pointer may still be close to the location resulting from the first of the two trials if they occurred in close succession, but is unlikely to be so if they were well separated in time. As a result, massed trials can lead to superior recall if the retention interval is very short.

Figure 9 shows results of a simulation in which parameter values were chosen so as to show an interaction of retention interval with spacing interval. This figure demonstrates the possibility of a crossover such that at short intervals massed trials are better, while at long retention intervals spaced trials are better. The right-hand panel of Figure 9 shows the same qualitative features from a human learning experiment.

RELATION TO STIMULUS SAMPLING THEORIES

In its explanation of human learning phenomena of the classical varieties described so far, the model is a close relative of stimulus sampling and stimulus fluctuation theories (e.g., Estes, 1950, 1955a,b; Prokasy, 1961). In these theories, learning results from the conditioning of independent aspects of the external and internal environment, so-called stimulus elements, to particular responses. Stimulus elements are assumed to be sampled randomly from a large population, a new sample being drawn in each successive learning or performance trial. Response probability depends on the relative number of appropriately conditioned elements sampled on a given trial. (For more information, see Atkinson, Bower & Crothers, 1965.) It is readily apparent that these mechanisms are very similar to the ones being proposed here. Indeed many of the mean response probability predictions made here can be derived from some form or extension of stimulus sampling theory. However, it remains to be seen whether there are any circumstances in which the two kinds of theory are fully equivalent. In lieu of a definitive analysis, at this time I can only mention some properties of the present model which would not be shown by stimulus sampling theories without the addition of new assumptions or mechanisms. First is the crossover phenomenon in the spacing effect. Published forms of stimulus sampling theory (Estes, 1955a,b) predict superior performance for spaced learning trials at all retention intervals. The prediction is generated by assuming that stimulus elements "diffuse" in and out of the sample set with time. Samples resulting from spaced trials therefore always contain a larger

number of different elements to be conditioned. However, certain additional assumptions about the postulated diffusion process would yield the required interaction with retention intervals. In particular it could be postulated that elements that diffuse out of the sampled set remain nearby, and are therefore more likely to diffuse back in for a short time after a conditioning trial. As a result, the rate of forgetting after the second of two massed trials would be less than after the second of two spaced trials despite an opposite ordering in the forgetting asymptote. (The shape of the simple forgetting curve is also different in the two models. There is a true-to-life flat initial portion in the present model, a purely exponential decline in the Estes model.)

A second class of differences from stimulus sampling theory (Estes, 1950) involve the various reaction time predictions. Stimulus sampling theory as originally proposed contained no mechanism for generating such predictions other than to assume some arbitrary relation between response probability and latency. Subsequently, however, LaBerge (1962) proposed a version in which stimulus elements are sampled one by one with replacement until a criterial number of elements previously conditioned to one response has been "recruited." While there are some qualitative similarities, this theory is not mathematically equivalent to the present model. For example, it specifies a different function relating mean reaction time to number of prior learning trials (a mathematical form of the model for this relation is given later).

Third, there are differences in properties that arise from differences in the sources of randomness in the two theories. For example, short-term forgetting and various practice distribution phenomena arise in stimulus sampling theory from random diffusion and intermixing of conditioned and unconditioned elements. Thus the system would gradually lose information reflecting the order and relative time of occurrence of experience. By contrast, the present model leaves data where entered forever. One potentially observable consequence of this difference concerns correlations in recall. Suppose two items are encountered close to one another in time. According to stimulus fluctuation theory (Estes, 1955a,b), their recall should be positively correlated if tested a short time later, but should become uncorrelated with the passage of time. In the present model, some correlation would persist indefinitely.

Putting such detailed considerations aside, the present theory has a very different conceptual structure from that of the stimulus sampling theories and would be worth considering on this account alone. As one example of this assertion, consider the contrast in interpretations of simple learning offered by the two models. One of the best-established relations in psychology is the "law of constant proportional gain"; that is, in simple learning the amount learned in successive equal periods of

practice (e.g., trials) is, on the average, a constant proportion of what remains to be learned. If the measure is response probability, this relation is mathematically expressed as

$$P_n = P_{n-1} + \alpha (1 - P_{n-1}), \quad (1)$$

where P_n is the probability of response on trial n . Stimulus sampling theory accounts for this remarkable natural law by assuming that successive learning trials expose the organism to random variations in stimulation with the requisite set relations. Most other learning theories either make essentially the same assumption, that the law arises from the probability summation of independent inputs (e.g., Anderson & Bower, 1973), or simply assume that the "strength" of a trace grows appropriately (e.g., Hull, 1943). The present model accounts for the law of constant proportional gain as a fact of retrieval rather than learning. Successive learning trials produce independent data entries, any one of which will serve to support a response if it can be found. The overall probability of response is therefore just the probability of finding any one or more of the n records resulting from n learning trials. It is easy to show that this yields the relation expressed in (1). The learning parameter in this model represents the proportion of memory that can be searched at a given point in time.

LONG-TERM MEMORY RETRIEVAL

We now shift attention from features of the model that describe classical learning and forgetting phenomena to those that are more directly concerned with problems of retrieval of long-term memories. The particular applications considered here involve retrieval of lexical or semantic information about words. These applications are, in a sense, more interesting than the previous ones, in that they provide novel explanations and/or quantitative descriptions for many phenomena for which there are no well-established theories.

First we will consider the relation between frequency and retrieval. In this model each time the same information is encountered, a new record of it is entered in a new location. When the system attempts to access that information, the average distance to the closest place where it can be found will be a function of how many different places it is stored. Quantitatively, the relation between the mean of the distance, R , to the nearest storage location from a given pointer locus and the number of stored replicas, n , is approximated by the equation

$$\text{Mean } (R) \cong kn^{-1/3} \quad (2)$$

when n is large. A derivation of this expression is given in the appendix. The time required for access is assumed to be a linear function of the

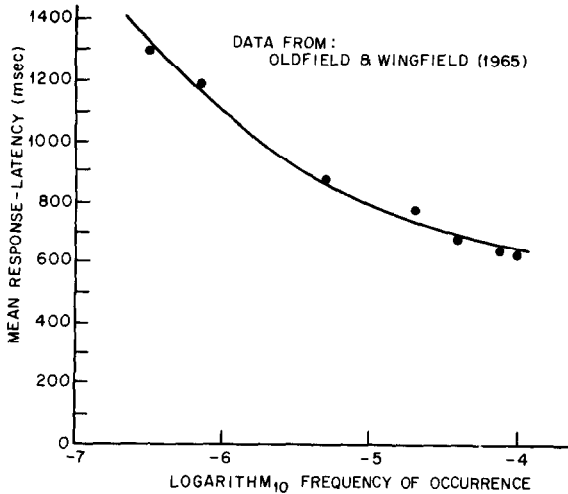


FIG. 10. The time required to say the name of an object represented by a line drawing. The smooth curve is a least-squares fit of the equation $RT = a + kn^{-1/3}$, an approximation to the model for a large storage space. Frequencies are taken from the Lorge magazine count (Thorndike & Lorge, 1944).

distance from the pointer to the nearest of n locations at which searched-for information can be found.²

The virtue of having an explicit expression rather than a simulation for this case is that it is interesting to explore the results of a very large number of encounters with the same information, and the simulation procedure becomes prohibitively costly. Three sets of data involving large frequencies were examined. The first set of data are those of Oldfield and Wingfield (1965) for times required to name line drawings of common objects as a function of the frequency in English of the names of the objects. We need to assume that the frequency of object name words in English is proportionally reflected in the number of loci at which the information needed to go from a picture to a name is stored in the average subject's memory. This seems to be a reasonable assumption.

² Retrieval time differences are treated as if their only source is in the distance, R , that must be traversed by the spreading search signal. This would be the case given three assumptions. First, the time required for matching the search signal and target is independent of R , second, the time for return of stored information is also a linear function of R , and third, the pointer does not move appreciably during a search, i.e., search time is short relative to pointer movement time. For simplicity the receiving point for information returned from memory loci has been assumed to be the pointer, and the transmission of returning information to be like that of the search signal, only in the opposite direction. If return were to some other place, the shortest input-output path would not necessarily be through the nearest memory replica to the pointer, and the expression for the expected search-path distance would be different.

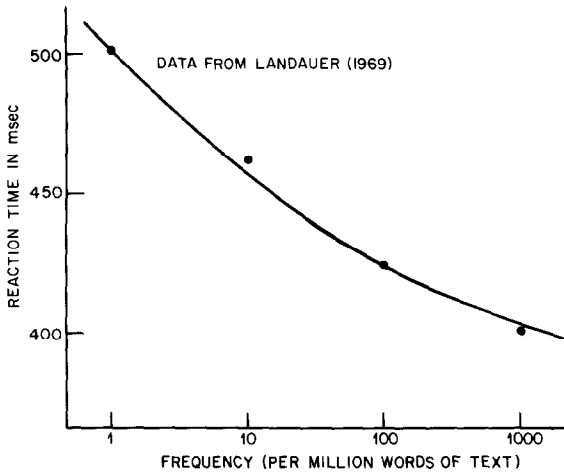


FIG. 11. The same equation described for Fig. 10, fit by least squares to data on the time required to decide that a visually presented word is not a member of a pre-defined semantic category, as a function of word frequency.

tion if one believes that objects with common names occur commonly, and that a constant proportion of the occasions on which objects are seen lead to new entries of object-name associations in memory. The fit of the model to the Oldfield and Wingfield data is shown in Fig. 10.

The second set of data was reaction times from a task in which subjects decided whether a visually presented word was or was not a member of a prespecified category. For this case we need to assume that a constant proportion of times that a printed word is encountered the system stores a record of the—graphic, syllabic, articulatory or phonemic—“name” of the printed word along with a visual representation. Presume further that in performing the categorization task the subject often or always determines the “name” of the word in a first stage and then searches again to discover whether it is or is not a category member (see Meyer, Schvaneveldt & Ruddy, 1974, for evidence of such a process; see Collins & Quillian, 1969; Landauer & Meyer, 1972; Meyer, 1970; Smith, Shoben & Rips, 1974 for some discussion of how category membership information might subsequently be retrieved from long-term memory). It is assumed that frequency has its primary effect during the first stage, that of recognizing the word.³

The model was fitted to some unpublished data from an experiment in which subjects judged whether visually presented words were or were

³ To be precise, we need to assume that later stages are either independent of frequency, or depend on frequency according to a function that is itself a linear function of that describing the relation for stage I.

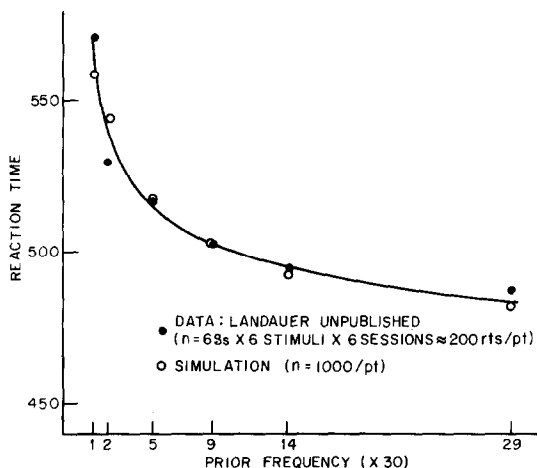


FIG. 12. Data for RT to press a key indicating the correct of three possible responses in a PA task as a function of number of previous equally spaced trials. The solid line is a least-squares fit derived from the same equation used in Figs. 10 and 11. The filled circles are for the human data, the open circles for simulation "5s".

not category members (ANIMALS or LIVING THINGS) (Landauer, 1968). The data presented in Fig. 11 are only for negative instances, which constituted 80% of the trials. What is plotted is the mean reaction time for 20 subjects judging 60 noncategory words of each frequency class shown. The words were matched on length. The smooth fitted curve was derived from the formula in Equation (2) with asymptotic minimum and slope constant estimated from the data. The model provides a good fit to the data.

The third set of data was obtained from a new unpublished experiment in which the frequency of a word-digit association was varied experimentally, holding constant such factors as word length, graphemic and phonemic composition of words—by counterbalancing—and spacing of practice during learning. Subjects learned sets of 96 paired associates in a series of 36 sessions (Landauer, 1973). These paired associates consisted of common four-letter nouns paired with the responses 1, 2, or 3 made on a keyboard. Among the pairs, there were six each that occurred, respectively, 1, 2, 5, 9, 14 or 29 times per session. The data shown in Fig. 12 are for the first occurrence only of each word during sessions 31–36. There were six subjects over whom responses and the particular pairs occupying the different frequency classes were counterbalanced. The least-squares fit of Eq. (2) to these data is shown as the solid line. Also shown are discrete values obtained indirectly from the simulation program. (These provide indication that the mathematical and simulation approximations yielded very similar predictions.)

It is worth noting that the model does not always expect superior performance for common as compared to rare words. In "episodic" (Tulving, 1972) recognition experiments in which the *S* is to decide whether a presented word has occurred previously in the experiment, just the opposite might be predicted (and is observed, e.g., Shepard, 1967). Recorded tokens (particular occurrences) of the presented word would have to be searched for and then a decision made on the basis of returned context-tagged or recency information as to whether any such token appeared in the experimental set. Common words would find a higher proportion of tokens that had not appeared in the experiment, thus leading to a potentially more difficult decision.

Perhaps the best-known effect of word frequency is on the probability of correct recognition of a word under degraded stimulus presentation. To apply the model to this case, we can assume that the incoming information—what is being looked for during the search process—is incomplete and therefore search for a complete match in memory will often fail. Search must then be made for a partial match. If this were so, other things equal, the first word that matched adequately would be more likely to be a common than a rare word. Since it is assumed that each word is represented in memory in proportion to its frequency in English, the model reduces to a special form of the urn or sophisticated guessing model proposed by Pollack, Rubenstein and Decker (1959). In this theory word-images are stored in multiple memory loci in proportion to frequency. These images are examined until one is found that is acceptably similar to the perceived stimulus. The probability that the first matching image corresponds to the stimulus word is then just the number of memory loci containing images of the actual stimulus word relative to the total number of memory loci containing acceptably similar images. The probability of correct identification will differ between common and rare words only insofar as their frequencies differ relative to other words with which they are similar. Landauer and Streeter (1973) found that the frequency relations among orthographically similar four-letter words are consistent with this account of the superior recognition of common words.

The final applications to be considered here concern similarity effects in retrieval of knowledge about words. Meyer and Schvaneveldt (1971) had subjects judge whether two presented letter strings were both words. Judgments were faster if the two words were associates of each other than otherwise. The model explains such results as follows. It is necessary first to assume that associated words are usually encountered near to each other in time in the average subject's experience. Thus DOCTOR and NURSE, two of Meyer and Schvaneveldt's associated words, are assumed, on average, to occur near each other in common discourse or reading. If so, the minimum distance between records of

unassociated words like DOCTOR and BUTTER will, on the average, be greater. Now consider a search that will decide that both of them are words. The time to find both words will be that required to reach the closest record of each, and this will depend on the most distant of the two. The farther apart these two closest records are, the greater will be the average distance to the most distant of them. Thus, on the average, two associated words will be found faster than will two unassociated words.

A similar prediction holds for judgments of three strings. The greatest expected maximum of the distances to the nearest records of all three will occur when none are associated; the smallest when all three are associated; and an intermediate value is to be expected if any two of the three are associated, even if they are separated by a non-associated word. This last expectation has been confirmed in results obtained by Meyer, Schvaneveldt and Ruddy (1972) in an extension of the original paradigm, and replicated by Loftus (1973) using a somewhat different task.

It is worth noting that these predictions follow from the theory either if the separate two or three strings are assumed to be evaluated simultaneously during the same search, or if they are evaluated in successive searches from different pointer locations.

One version of an explanation based on the successive-search assumption is of special interest. Suppose that the subject evaluates each word in turn, using as many successive memory searches as needed to find a stored token of the first word, then the second, etc. Since a search cannot be completed successfully unless the pointer is within range of a record of the word, the random walk of the pointer will have had to take it near such a record for the word to be evaluated, and it will still be nearby just afterward. Because associated words tend to be stored near each other (due to history, not intrinsic organization), the next word will tend to take less time to evaluate if it is an associate. If three words are processed in succession with an unassociated word inserted between two associated words, the facilitative effect will be attenuated but not lost. The remaining facilitation will result from the fact that the particular token of the unassociated middle word that is found will be, on the average, nearer than the average of its type to that of the first word, and therefore also nearer than average to recorded tokens of words associated with the first word. The special interest of this mechanism is that it is closely related to the "location shifting model" considered and rejected by Meyer, Schvaneveldt and Ruddy (1972) and Loftus (1973). The important difference is that they assumed single rather than multiple storage. Under this assumption an unassociated middle word would be stored at random with respect to the third word, and would be expected to remove entirely the effect of association between the first and third.

The observation of a significant association effect allowed rejection of the location shifting model. However, assuming multiple copies of each word, as described above, such a model would be consistent with the results in question.

A similar kind of explanation applies to an observation of Rubenstein, Lewis and Rubenstein (1971). They reported that subjects were faster to judge that a string of letters was a word if it had two or more common meanings rather than just one, with total frequency of usage held constant. For this case, it is only necessary to assume that words with multiple meanings are stored in more disparate places because their experienced occurrences, on the average, are more widely distributed in time. Thus the different meanings of YARD are likely to have been encountered at different times. To determine whether such a letter string is a word, one needs to find only one representation of it, and the farther the multiple representations are apart, on the average, the closer will be the nearest one. If tokens of all meanings are searched for at once, a particularly plausible assumption for homophonic homographs, the search time will depend on the distance to the closest. Here the prediction follows from the same argument that predicts spacing effects in learning.

RELATIONS TO THE BRAIN

Because the theory is presented as a spatial model, it is tempting to interpret it also as a concrete physical model of the brain. There are certain respects in which such interpretations can be made easily and yield interesting and plausible predictions, others which require postulation of mechanisms with no obvious foundations.

The most obviously applicable idea is that memories are stored multiply in randomly distributed, very small "registers." If so, it would be difficult to remove particular memories by removing particular pieces of tissue. It is postulated that any datum can go anywhere. Therefore, one would not be likely to erase the same memory by extirpating the same place in two different animals. Erasing a memory would depend on removing all the places it is stored. Therefore, (a) the probability of destroying a particular memory would depend on the amount of tissue removed, not its location, and (b) the better learned a particular fact (or habit), the less would be the probability that removal of a given amount of tissue would destroy its memory.

These are, of course, the central facts, first established and summarized by Lashley (1934, 1950), that have given rise to various theories of distributed memory traces. The present model provides a novel and rather straightforward explanation of such facts.

The most problematical aspect of the theory as a brain model is the randomly walking pointer. While one might be able to concoct an hypothesis for the pointer's physical basis, it would be highly speculative

and would probably serve no present purpose. However, if some mechanism corresponding to the pointer exists, variations in its behavior could have interesting implications. For example, if the pointer's movement were restricted to a small region, it would begin to selectively overwrite recently entered data. The observed effect would be a markedly decreased ability to store new memories that would last. However, loci within the search radius—thus a fair proportion of all previously learned facts, especially those stored many places (i.e., most of LTM)—would still be retrievable, as would the last few data entered. By contrast, if the pointer began to move much more rapidly than normal, short-term forgetting would become very rapid, while the storage and retrieval of long-term memories would be largely unaffected. These expected contrasting pathologies match very closely ones that have been observed in brain damage cases (e.g., Milner, 1966; Shallice & Warrington, 1970). A sudden displacement of the pointer would produce a temporary loss of memory for recent events, much as in retrograde amnesia from trauma or ECS. It is worth noting that, if nothing else, this analysis provides a counterexample showing that differential effects of brain injuries on memories at short and long retention intervals do not require the assumption of separate short- and long-term memories.

A final aspect of the theory with a possible physiological interpretation is the assumption that memory search is accomplished through the agency of a "broadcast" search signal that releases the contents of memory storage loci. This is not a new idea, physiologically speaking, having been discussed in various forms by many people (e.g., John, 1967; Landauer, 1969; Pribram, 1971). It is, however, an idea still in need of proof.

OTHER PHENOMENA

There are, of course, a host of facts known about human learning and memory for which a random storage model has yet to account. Among the more interesting in the present context are phenomena that suggest intrinsic organization and reorganization in memory. These include clustering in recall, the effect of meaningfulness and other prior knowledge on new learning, context and set effects, release from PI, and the influence of mnemonic aids. It is not my purpose here either to catalog such phenomena or to attempt to provide model-based explanations of them all. Indeed many of them are probably outside the domain of the model entirely. However, it seems worthwhile to suggest ways in which the explanation of additional phenomena could be approached, without modifying the basic structure of the model as an account of memory storage and retrieval or diminishing its established predictive value.

Let us take as an example of a phenomenon to be explained the semantic clustering of items in free recall. As observed by Bousfield

(1953) and many followers, when subjects are given randomized lists containing semantically categorizable words, say animal and plant names, their order of recall is nonrandom; words from the same categories tend to occur near each other in the output protocol.

Assume that subjects categorize the words as they store them, tagging them both as "on list x" and also as "animals" and "plants." Then in searching they use—at least some of the time—the category name as well as the list tag as the search signal to find the required doubly tagged words. This explanation is cast in terms of strategies applied by the homunculus in inputting and retrieving data from its memory. The memory itself remains passive and unorganized in the sense meant here.

This trick would allow the system, if clever enough, to use its unorganized memory to store organized sets of facts. By storing "A is a B," "B is above C," and the like, various kinds of associative networks could be formed. Stored networks could be searched in multiple passes, using the outcome of one search as the target for the next, and so forth. Anderson and Bower (1973) have argued that passive networks of elementary propositions have the potential to explain almost everything now known about "organizational" properties of memory. And, indeed many of their detailed proposals concerning sentence parsing and representation, match criteria, list and order information storage, and so forth could easily be grafted onto the present model. No attempt to do this kind of thing is made here, because the focus of interest is at the level of basic storage and retrieval mechanisms, rather than on properties and interrelations among the contents of memory. Nonetheless, it should be apparent that the proposals made here are, by and large, potentially complementary to, rather than in necessary opposition to work on content organization. (Hopefully, it is also apparent that these proposals show that some phenomena whose explanation might have been sought in content organization can instead result from an unorganized storage and retrieval mechanism.)

PROBLEMS

The theory as presently developed has two types of deficiencies. The first is that there are, as already mentioned, certain relevant phenomena to which it has not yet been applied successfully. Some of these may appear superficially to pose serious difficulties. One example is the linear set-size effect for reaction times to identify items in working memory (Sternberg, 1966). I have found no plausible account for this phenomenon that does not require adding a completely separate mechanism. Another, possibly tractable example is the well-known primacy effect that occurs in list-learning of many kinds. Without elaboration the random storage model correctly predicts positive recency effects that disappear with time, because the pointer is likely to begin retrieval near

the last-stored item. But positive primacy effects would apparently require additional explanation. One possibility would be consideration of processes outside the memory storage system as such, for example, rehearsal strategies or attentive biases on input such that early items are entered more times. Another possibility would be special storage and input assumptions, for example, that more than one word is put into each successive register, one being the "current" word, the others chosen from the set of previously entered words. This would also result in more copies of early items. These explanations are close relatives of ones proposed previously in other guises. It is not surprising, of course, that this simple system does not immediately explain everything, or that previous explanations of some phenomena will not be altered or replaced by the postulation of a new storage and retrieval mechanism.

The second deficiency is that the various predictions from the model have not been tied together sufficiently. Learning and forgetting curves, spacing functions, and reaction time predictions all follow from a common set of assumptions and mechanisms. In principle they should be capable of simultaneous prediction in appropriate situations. But the applications reported here are piecemeal, some involving a tailored choice of assumptions and parameter values. To obtain stronger tests of the model it should be applied to experiments in which several predictions can be tested at once. For example, reaction times as a function of number of repetitions, spacing, and retention interval should be simultaneously predictable from the same parameter estimates derived from the same experiment. Similarly, in lexical decisions the facilitative effects of association and of homography are predicted from assumptions about language statistics that, while plausible taken one at a time, ought to be evaluated by simultaneous estimates made on the basis of the same language sample. If then applied to results of comparable experiments, the absolute sizes of reaction time effects should be mutually consistent between the two phenomena. Data on which to make such tests need to be collected.

SUMMARY AND DISCUSSION

I have outlined a minimal heuristic model of human memory storage and retrieval in which there is no self-organizing ability, and in which retrieval processes are undirected—in the sense that they do not vary with the nature of information sought. The model has been applied to a variety of phenomena in human learning, forgetting and retrieval. It seems to possess a large number of lifelike properties. In particular, it gives a good account of many classical human learning phenomena, including acquisition and forgetting curves and spacing of practice effects. It adds to these a natural and successful means of generating reaction time predictions. It also provides, for the first time, a quantitative process

model for predicting recognition probability and reaction time as a function of word frequency. It is also qualitatively consistent with some recent results from semantic memory experiments. And it offers avenues of expansion to still other domains.

Some general characteristics of the model are of interest in relation to certain classificatory distinctions sometimes made in discussing memory models. First, consider the distinction between long-term and short-term memory. Many phenomena typical of both long- and short-term memory and their differences can be explained by the one-device conception used here. Among these are the backward-S shape of the forgetting curve, the occurrence of positive recency effects, and the possibility of differential pathological impairment of retention over short and long intervals. The present model shows that these properties do not require separate long- and short-time memories as has sometimes been supposed.

Second, consider Tulving's (1972) distinction between semantic and episodic memory. The essential difference is that semantic memories (e.g., word meanings) as contrasted to episodic ones (e.g., autobiographical events) do not admit of accurate or unique temporal placement. In the present conception, temporal information could either be stored directly as temporal context or could be inferred to some extent from search times (very recent events are closer). In the present model autobiographical events would usually be stored along with unique temporal contexts; and repeated events, like the information that defines words, would be stored in multiple places along with multiple temporal contexts. Consequently, the homunculus could not retrieve unique time-of-occurrence information about often-encountered semantic knowledge, but could about nonrepeating occurrences. Thus, the multiple record idea provides a single-process explanation of the qualitative differences that led Tulving to suggest separate memories.

Third, consider whether human memory is properly to be considered associative, content addressable, or random access. In this model, information is found by matching aspects of contents, and, while information is entered in temporal order, its retrieval does not depend on retracing that or any particular order. Thus the system appears to have features of associativity, content addressability and random access. Yet both recognition and recall always require search, and the data entry process leaves correlations in the storage loci, and thus in retrieval ease from given pointer loci. These features might be viewed as evidence that the system is not associative, not content addressable, and not randomly accessible. (Note also that the model provides potentially very rapid access despite the dependence on search.) The conclusion appears to be that the model constitutes a very mixed breed of memory device, in that none of these popular distinctions seems especially useful in character-

izing it. In fact, the mere existence of the model as a logical possibility raises questions as to the kinds of properties that would require theories to make qualitative distinctions between memories of different durations, different content types, or different general modes of access.

No claim is made or implied that this model can account for all phenomena of learning and memory. It has not been developed to do so, and there are a multitude of memorial phenomena for which its predictions have not yet been derived, or for which its mechanisms will provide only part of the answer. For example, its relation to phenomena like clustering in free recall have not been worked out, although there are promising directions for exploration in this regard. Moreover, there are some phenomena for which it appears, by itself, to make wrong predictions—notably some serial position effects. These unexplained phenomena may eventually require modifications of the model in various ways, for example, by the addition of input and output variables and strategies.

Nonetheless, the significant fact is that, in the absence of strong assumptions regarding organization, the model does perform in many ways as humans do. This suggests that it is worth considering to what extent and for what purposes stronger assumptions are really necessary. The successes of this model make it appear quite likely that a large range of important phenomena of human memory may be explicable as the result of simple and basically random underlying processes.

REFERENCES

- Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, 79, 97-123.
- Anderson, J. R., & Bower, G. H. *Human associative memory*. New York: Winston, 1973.
- Atkinson, R. C., Bower, G. H., & Crothers, E. J. *An introduction to mathematical learning theory*. New York: Wiley, 1965.
- Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 2. New York: Academic Press, 1968.
- Bernbach, H. A. Replication processes in human memory and learning. In G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation*, Vol. 3. New York: Academic Press, 1969.
- Bousfield, W. A. The occurrence of clustering in the free recall of randomly arranged associates. *Journal of General Psychology*, 1953, 49, 229-240.
- Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation*, Vol. 1. New York: Academic Press, 1967.
- Brown, R. W., & McNeill, D. The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 325-337.
- Buschke, H. Spontaneous remembering after recall failure. *Science*, 1974, 184, 579-581.
- Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 240-247.
- Estes, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, 57, 94-107.

- Estes, W. K. Statistical theory of distributional phenomena in learning. *Psychological Review*, 1955, **62**, 360-377. (a)
- Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychological Review*, 1955, **62**, 145-154. (b)
- Feller, W. *An introduction to probability theory and its applications*, Vol. 1. (2nd ed.) New York: Wiley, 1950.
- Freedman, J. L., & Landauer, T. K. Retrieval of long-term memory: "Tip-of-the-tongue phenomenon." *Psychonomic Science*, 1966, **4**, 309-310.
- Hart, J. T. Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 1965, **56**, 208-216.
- Hellyer, S. Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, 1962, **64**, 650.
- Hilgard, E. R., & Bower, G. H. *Theories of learning*. (3rd ed.) New York: Appleton-Century-Crofts, 1966.
- Hintzman, D. L., & Block, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297-306.
- Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
- John, E. R. *Mechanisms of memory*. New York: Academic Press, 1967.
- Kintsch, W. *Learning, memory and conceptual processes*. New York: Wiley, 1970.
- LaBerge, D. A recruitment theory of simple behavior. *Psychometrika*, 1962, **27**, 375-396.
- Landauer, T. K. Two hypotheses concerning the biochemical basis of memory. *Psychological Review*, 1964, **71**, 167-179.
- Landauer, T. K. Word categorization as a function of frequency, length and stimulus degradation. Paper presented at a meeting of the Psychonomic Society, St. Louis, MO, November 1968.
- Landauer, T. K. Reinforcement as consolidation. *Psychological Review*, 1969, **16**, 82-96.
- Landauer, T. K., & Meyer, D. E. Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 539-549.
- Landauer, T. K. & Streeter, L. A. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 119-131.
- Landauer, T. K. Reaction time effects of artificial category size. Paper presented at a meeting of the Psychonomics Society, St. Louis, MO, November 1973.
- Lashley, K. S. Learning III. Nervous mechanisms in learning. In C. Murchison (Ed.), *A handbook of general experimental psychology*. Worcester, MA: Clark University, 1934. Pp. 456-696.
- Lashley, K. S. In search of the engram. In Society for Experimental Biology, *Physiological mechanisms in animal behavior*. New York: Academic Press, 1950.
- Lindsay, P. H., & Norman, D. A. *Human information processing: An introduction to psychology*. New York: Academic Press, 1972.
- Loftus, E. F. Activation of semantic memory. *American Journal of Psychology*, 1973, **86**, 331-337.
- Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, **1**, 242-300.
- Meyer, D. E., & Schvaneveldt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 1971, **90**, 227-234.
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Activation of lexical memory. Paper presented at a meeting of the Psychonomic Society, St. Louis, MO, November 1972.
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition*, 1974, **2**, 309-321.
- Milner, B. Amnesia following operation on the temporal lobes. In C. W. M. Whitty & O. L. Zangwell (Eds.), *Amnesia*. London: Butterworth, 1966.

- Norman, D. A., & Rumelhart, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models for human memory*. New York: Academic Press, 1970.
- Oldfield, R. C., & Wingfield, A. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 273-281.
- Peterson, L. R. Reminiscence in short-term retention. *Journal of Experimental Psychology*, 1966, **71**, 115-118.
- Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193-198.
- Peterson, L. R., Hillner, K., & Saltzman, D. Supplementary report: Time between pairings and short-term retention. *Journal of Experimental Psychology*, 1962, **64**, 550-561.
- Pollack, I., Rubenstein, H., & Decker, L. Intelligibility of known and unknown message sets. *Journal of the Acoustical Society of America*, 1959, **31**, 273-279.
- Pribram, K. H. *Languages of the brain*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- Prokasy, W. F., Jr. Nonrandom stimulus sampling in statistical learning theory. *Psychological Review*, 1961, **68**, 219-224.
- Quillian, M. R. Word concepts: A theory and simulation of some semantic capabilities. *Behavioral Science*, 1967, **12**, 410-430.
- Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Homographic entries in the internal lexicon: Effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 57-62.
- Shallice, T., & Warrington, E. K. Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 261-273.
- Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156-163.
- Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review*, 1974, **81**, 214-241.
- Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652-654.
- Thorndike, E. L., & Lorge, I. *A teacher's wordbook of 30,000 words*. New York: Columbia University Press, 1944.
- Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.
- Tulving, E. The effects of presentation and recall of material in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 175-184.

(Accepted March 3, 1975.)

APPENDIX

The derivation of Equation 2 may be sketched as follows. Assume that there are n replicas of a given fact stored randomly within the total memory storage space. Recall that the search space on any retrieval attempt is defined as a region lying within a certain radius of the pointer. The number of replicas within the search space on any given trial will be a binomially distributed random variable M , with expectation αn , where α is the ratio of the search space volume to the volume of the total memory space.

(Note that we have assumed exactly n replicas uniformly distributed over the total storage space. In practice the number of replicas is taken to be a fraction of some index of exposure frequency. In addition, the

distribution of replicas within the total search space ordinarily will not be strictly uniform since it depends on temporal spacing of entries. However, the present simplifying assumptions are convenient and probably cause no appreciable difficulty in the situations to which Equation 2 is applied.)

For expository reasons, the search space has been assumed to be a three-dimensional sphere. For more generality, in the following derivation the dimensionality of the storage and search spaces is made a parameter, a .

The search space is assumed to be symmetrical in shape; that is, it is a line of twice the search radius in one space, a circle or sphere of that radius in two or three space, or a hypersphere of the same radius in a space. Without loss of generality we may arbitrarily assign the search space unitary radius. Because access time is supposed to be related to search distance, we are interested in the moments of the distribution of distances from a random pointer location to the nearest of the M replica storage points lying within the search space. Let R equal that distance. (We are concerned with cases involving large n and moderate α , so the probability that $M = 0$, in which case no replica is found, is small enough to be ignored.)

At any time after the initiation of search the radius so far reached will have a particular value, and a related volume of the space will have been searched. If the radius reached is r , then the volume of space so far searched relative to that of the whole search space is,

$$\frac{\text{volume up to radius } r}{\text{volume of search space}} = r^a.$$

Given $M = m$ replicas lying within the search space, we want first to derive the density function for the random variable, R , the distance to the nearest replica. This density function will allow us to find the mean and other moments of R . We will start by determining the distribution function, $F(r|m)$ of R for given M . We will then differentiate to get its density function $f(r|m)$, use this density function to obtain $E(R|m)$, and finally take the expectation with respect to the binomial distribution of M to find moments of R .

For a particular replica,

$$\Pr\{\text{replica not contained within radius } r\} = 1 - r^a \quad 0 \leq r \leq 1.$$

For all m replicas,

$$\Pr\{\text{replica not contained within radius } r|M = m\} = (1 - r^a)^m.$$

This is the probability $\{R > r\}$, hence the desired distribution function is

$$F(R|m) = \Pr\{R \leq r\} = 1 - \Pr\{R > r\} = 1 - (1 - r^a)^m \quad m > 0.$$

Differentiation yields the density function

$$f(r|m) = mar^{a-1}(1 - r^a)^{m-1}.$$

The j th raw moment conditional on $M = m$ is then given by

$$\begin{aligned} E(R^j|M = m) &= \int_0^1 r^j f(r|m) dr \\ &= am \int_0^1 r^j r^{a-1} (1 - r^a)^{m-1} dr \end{aligned} \quad (3)$$

and taking the expectation over the binomial distribution of M ,

$$\mu_j = E(R^j) = \sum_{m=0}^n \binom{n}{m} \alpha^m (1 - \alpha)^{n-m} \left[am \int_0^1 r^j (1 - r^a)^{m-1} r^{a-1} dr \right],$$

letting $s = r^a$, interchanging summation and integration, and using the binomial expansion sum, we have

$$\mu_j = n\alpha \int_0^1 s^{j/a} (1 - \alpha s)^{n-1} ds.$$

The derivation is exact to here. A more tractable form can be obtained by making certain approximations that will yield close to exact values when n is large, α is moderate and j and a are small. To help in developing these approximations, we first make the change of variable $t = \alpha(n - 1)s$, which gives

$$\mu_j = \frac{n\alpha}{[(n - 1)\alpha]^{j/a+1}} \int_0^{(n-1)r} t^{j/a} \left[1 - \frac{1}{n-1} \right]^{n-1} dt.$$

Now we make three approximations that depend on n being large. First, replace

$$\frac{n\alpha}{[(n - 1)\alpha]^{j/a+1}} \text{ by } \frac{1}{(n\alpha)^{j/a}}.$$

Second, replace $[1 - t/n - 1]^{n-1}$ by e^{-t} , giving

$$\mu_j \cong \frac{1}{(n\alpha)^{j/a}} \int_0^n e^{-t} t^{j/a} dt.$$

Finally, replace the upper limit of integration, n by ∞ , giving

$$\mu_j \cong \frac{1}{(\alpha n)^{j/a}} \Gamma\left(\frac{j}{a} + 1\right). \quad (4)$$

For the mean, Equation 4 provides the approximation

$$\text{Mean}(R) = \mu_1 \cong K_1 n^{-1/a}, \text{ where } K_1 = \alpha^{-1/a} \Gamma\left(\frac{1}{a} + 1\right).$$

Numerical determinations of exact and approximate mean values with $a = 3$, $\alpha = .33$ and selected values of n between 30 and 207 found close agreement. The maximum difference was $<1\%$ and decreased steadily with increasing n . In applications to data, the constant K_1 is replaced by K which indicates the inclusion of multiplicative factors for the relation of n to exposure frequency, and the relation of r to physical search time.

Similarly, for the variance, Equation 4 provides the approximation

$$\text{Var}(R) = \mu_2 = \mu_2^1 - \mu_1^2 = K_2 n^{-2/a},$$

where

$$K_2 = \frac{1}{\alpha^{2/a}} \left\{ \Gamma\left(\frac{2}{a} + 1\right) - \left[\Gamma\left(\frac{1}{a} + 1\right) \right]^2 \right\}.$$

Attempts were made to estimate a from the data shown in the last three figures by iterative nonlinear least-squares fitting procedures, but the results were not especially illuminating. In all cases $a = 3$ appeared to be satisfactory, but the goodness-of-fit did not depend very strongly on the choice of a . It appears that extremely precise data would be necessary to determine the dimensionality of the search space empirically.