JOURNAL OF MATHEMATICAL PSYCHOLOGY 9, 418–455 (1972)

# Trace Resistance and the Decay of Long-Term Memory[1]

WAYNE A. WICKELGREN

*Department of Psychology, University of Oregon, Eugene, Oregon 97403*

A theory of storage in long-term memory is described which characterizes the trace by two properties: strength and resistance. The theory makes four assumptions: (a) The rate of decay of strength equals the force on the trace divided by the resistance. (b) Force is proportional to trace strength and to the similarity of current traces to the previously established trace. (c) Resistance increases as the square root of trace age. (d) The resistance of a trace transfers completely to subsequent increments. The theory accounts for long-term retention functions over delays from 1 min to 2 yr, long retrograde amnesia, unlearning, effects of multiple learning trials and spacing of practice, and possibly recency judgments and retention functions following relearning.

This paper presents a simple mathematical theory of storage in long-term memory and provides some supporting evidence for the theory. The most basic assumption of the theory is that a long-term memory trace has two properties which are critical for its characterization in storage: its *strength* and its *resistance*. Both strength and resistance are nonnegative real variables. Under ordinary circumstances, the probability of correct recognition or recall depends upon the strengths of one or more traces, but does not depend upon the resistance of these traces. However, the susceptibility of a long-term memory trace to decay in storage is assumed to depend upon both its strength and its resistence. In addition, a number of other memory phenomena are explained by the concept of trace resistance, which seem difficult or impossible to explain with only trace strength.

To give the proper perspective regarding the limited goals of the present paper, it is important to note that this is a theory of *storage* in *long-term memory*. Short-term memory, which refers to memory lasting seconds or tens of seconds, is excluded from

[1] This paper was given at the 1971 Mathematical Psychology Meeting and as an invited address at the 1971 APA convention. The research was begun at M.I.T. and continued at the University of Oregon. The work was supported by grants MH 08890 and MH 17958 from the National Institute of Mental Health, Public Health Service to Wayne Wickelgren, by a National Aeronautics and Space Administration grant, NsG 496, to Hans-Lukas Teuber, by the Advanced Research Projects Agency, Department of Defense Contract No. F44620-67-C-0099, to Ray Hyman, and by an NSF grant to the Mathematical Social Science Board to support a summer conference on mathematical models of memory held at M.I.T. in July 1969.

consideration, except for comparative purposes. Processes of acquisition, consolidation, and retrieval of long-term memory are excluded, except that some of the phenomena used to evaluate the theory require certain minimal assumptions concerning retrieval and decision processes in recognition memory. These assumptions will be stated along with the independent evidence that exists to support them. Finally, it should be noted that this theory is concerned with certain quantitative, dynamic properties of memory traces, not with the manner in which memory traces code our knowledge of the world.

## Strength-Resistance Theory

*Definitions*

Let $l$ be the strength of a long-term memory trace.

Let $r$ be the resistance of the trace.

Let $f$ be the force of decay acting on the trace.

Let $t$ be the time since the formation of the trace.

Let $T$ be the retention interval (time from the end of the study period to the retention test).

Let $\pi$ be the similarity of the material currently being studied to the material involved in some previously formed trace.

*Axioms*

A1. $dl/dt = -f/r$.

A2. $f = \pi l$.

A3. $r = \mu t^{\gamma}, \mu > 0, 0 < \gamma < 1$.

Special Case: $\gamma = .5, r = \mu \sqrt{t}$.

A4. The resistance of a trace transfers completely to subsequent increments.

From the first three axioms, it is easy to derive the form of the strength retention function for long-term memory:

$$dl/dt = -\pi l/\mu t^{\gamma},$$

$$\int (1/l)\, dl = \log \lambda - (\pi/\mu) \int t^{-\gamma}\, dt,$$

$$\log l = \log \lambda - [\pi/\mu(1-\gamma)]\, t^{1-\gamma} = \log \lambda - \psi t^{1-\gamma}, \tag{1}$$

$$l = \lambda e^{-\psi t^{1-\gamma}}, \qquad \text{where} \quad \psi = \pi/\mu(1-\gamma), \tag{2}$$

Special Case:

$$\log l = \log \lambda - \psi \sqrt{t}, \tag{3}$$

$$l = \lambda e^{-\psi \sqrt{t}}. \tag{4}$$

Some attempt was made in the present study to determine the optimal value of the exponent of growth of trace resistance ($\gamma$). From these investigations it became clear that the optimal value of $\gamma$ for all studies reported in the present paper was in the vicinity of $\gamma = .5$. Since there were no major differences in the value of $\gamma$ across the various experiments, it seemed parsimonious to conclude that $\gamma$ was constant and equal to .5. However, it should be emphasized that any value of $\gamma$ between .2 and .6 would have worked almost as well as $\gamma = .5$. In any event, throughout the present paper the special case of Axiom 3 will be assumed, namely, $r = \mu \sqrt{t}$.

Thus, according to this theory, the long-term strength retention function is what might be called an "exponential-power function" with two identifiable parameters in the special case: $\lambda$ representing the degree of learning in long-term memory at the start of the storage period ($t = 0$), and $\psi$ representing the decay rate.

In terms of absolute loss of trace strength, this retention function is a specific quantitative formulation of the notion that the rate of forgetting is constantly decreasing as the retention interval increases [Jost's second law, in Hovland (1951, p. 649)]. The absolute loss of trace strength per unit of time is decreasing, according to this formulation: (a) because strength is decreasing and the force of decay is proportional to trace strength and (b) because resistance is increasing with time since learning.

The decay rate ($\psi$) is jointly determined by two (currently not identifiable) parameters: $\pi$ (the similarity of the interpolated material to the original material) and $\mu$ (the rate of increase of trace resistance for the original material). The values of $\pi$ and $\mu$ may be different for different learning materials, different amounts of study time, different conditions of learning, different types of interpolated materials and tasks, and other variations of the conditions during the retention interval. However, according to the theory, $\pi$ and $\mu$ are independent of the time since the start of the storage period ($t$). Thus, according to the theory, the decay rate ($\psi$) should be invariant with time under constant conditions during the retention interval. This is the central prediction being tested in the present paper when the fit of the proposed retention function in Eq. 3 and 4 is assessed. This invariance of decay rate with delay time can also be assessed by fitting separate retention functions to short delays and long delays under constant conditions to determine whether there are any systematic differences in the estimated decay rate.

The present theory makes no quantitative predictions regarding the dependence of either $\pi$ or $\mu$ on conditions or materials during the learning and retention periods. Assuming the general validity to the present theory, it is conceivable that $\mu$ is a constant over all conditions, but at present I have no evidence for or against that hypothesis.

By contrast, it will be shown to be necessary to assume that $\pi$ varies with conditions that one presumes to affect the degree of similarity between original and interpolated learning. Both structural and semantic similarity probably contribute to the average degree of similarity between original and interpolated material. No specific assumptions characterizing the nature of this similarity are incorporated in the present theory.

However, a number of averaging assumptions are being implicitly made in the present theory, and the validity of these averaging assumptions may be questionable. First, as has been already mentioned, if there are different types of similarity (phonetic, visual, semantic, etc.), these different types are assumed to be representable by their average degree of similarity. Second, even though a subject may be engaged in the same task with the same type of materials under the same conditions throughout the retention interval, the nature of the subject's thought processes cannot be assumed to be constant at all times during the retention interval. At most, the subject might be in a steady state, continually engaging in a sequence of activities which are *not* systematically varying with delay time on a more macroscopic scale. For example, if a subject is learning a series of items during the retention interval, he may go through approximately the same sequence of mental activities in relation to each item to be learned. This sequence of mental activity is assumed by the theory to be characterizable by a single average $\pi$ value in terms of its effect on previous learning. Third, different items presumably have somewhat different $\pi$ values. It is assumed by the theory that these different $\pi$ values for different items can be represented by a single average $\pi$ value.

According to the present theory, one expects the value of $\pi$ and therefore the value of $\psi$ (the decay rate) to be affected by a number of experimentally manipulable conditions.

First, $\pi$ should obviously be increased, if one increases the degree of structural or semantic similarity between original and interpolated learning materials. In the absence of a theory of structural or semantic similarity, one can only rely on subjects' or experimenters' ratings to determine the degree of similarity, but this is probably not a serious handicap, if only a qualitative effect is being investigated.

Second, one would expect $\pi$ to increase somewhat with the rate at which subjects had to learn interpolated materials. There is no particular reason to expect $\pi$ to be proportional to the rate at which new items or pairs have to be learned during retention interval, since, to some extent, greater time spent learning one item may have the same storage interference effect on the retention of a previous item as the same amount of time spent learning two items. However, if the subject is presented with material at a sufficiently slow rate that, for a considerable fraction of the time, he is not learning anything, it is reasonable to suppose that the value of $\pi$ under such conditions will be lower than under conditions of constant learning. The $\pi$ parameter might well be referred to as an "average similarity-over-time" parameter to reflect this averaging of $\pi$ values over different times during the "steady state" retention interval.

Third, if both original and interpolated learning materials are randomly selected from a given population of items, it is reasonable to expect the $\pi$ parameter to vary depending on the average degree of similarity (distinctiveness) of each item compared to every other item in the population. Along these lines, one would assume that words, which can be distinguished both on the basis of structure and meaning, have a lower average mutual similarity than do nonsense materials, which presumably have a lower probability of being meaningfully distinguished.

Fourth, a greater number of study trials or a greater amount of study time on a given learning trial permits the subject more time to recode the material in a different form from the original structural (phonetic and/or visual) representation. Although the subject might use additional study time simply to engage in "rote" rehearsal of the originally encoded form, everyone's experience indicates that subjects frequently use study time to encode learning material more effectively. In general, such additional encoding will serve to reduce the similarity of the learned material to material expected to be encountered in retention interval. Thus, one expects to find that, in general, greater amounts of study time produce both higher degrees of learning and lower decay rates. However, for very unfamilar material (e.g., Russian or Chinese words), achieving a high degree of learning may involve recoding the unfamiliar material into more familiar concepts by virtue of its similarity to the more familiar concepts. In this case, the effect of recoding could actually be to increase the degree of similarity of the original material to the material encountered during the retention interval. Exactly these types of "inconsistent" effects of study time and degree of learning on decay rate will be noted in the present study. Such results are discussed in the present paper only to contraindicate any attempt to postulate a direct relationship between degree of learning and decay rate. Such results are not in any sense predictions of the present theory.

Axiom 4 is rather independent of the first three axioms, and it may well require reformulation at some future time, even if the rest of the theory is supported by future studies. Axiom 4 states that the increments to trace strength contributed by multiple learning trials all start off with a resistance determined by the time since the first learning trial. It will be shown that this explains a number of qualitative findings regarding the efficacy of spaced practice and relearning, and the assumption also fits the results of one quantitative study. However, whether other quantitative studies will support this surprising assumption remains to be seen. It seems likely that for Axiom 4 to be valid it would be necessary for a certain minimum trace strength to be established on the first learning trial, and perhaps for a certain minimum trace strength to still be available at the time of any later learning trial. However, all such limitations or reformulations of Axiom 4 are beyond the scope of the present paper, since the necessary evidence is not available.

To test this theory, we need to know when the storage phase begins in relation to the period of active study of the material. Wickelgren and Berian (1971) present evidence that long-term memory consolidates primarily over the first 30 sec after the end of the period of active study, but that the onset of the storage (decay) phase may be delayed until 30 sec to 3 min after the study period. Actually, the evidence in the Wickelgren and Berian study indicated that the long-term trace was not subject to any degradative forces until 30 sec to 3 min after the end of the study period, but indicated nothing about when the trace resistance might be assumed to begin increasing. The present paper does not attempt to determine the forces of decay and the onset of the increasing trace resistance in any detail. Rather, we shall assume that trace resistance begins to grow immediately after learning and that the trace becomes subject to decay at or around a minute following learning.

Thus, we will confine ourselves to retention functions for periods greater than a minute (0.9 of a minute in some cases). The decision to limit consideration to retention intervals greater than a minute is also justified by the suggestion in the Wickelgren and Berian study that short-term memory may persist to some small degree for as long as a minute after learning. Because trace resistance is assumed to begin accumulating immediately after learning, the value of $t$ in Eqs. 3 and 4 will be taken to be exactly the same as the retention interval ($T$). In the future, this may have to be decreased by some small quantity. Such a change in the value of $t$ would have only a very small effect on the results reported in the present paper for even the experiments with the shortest retention intervals.

The following six sections of the paper discuss a variety of qualitative and quantitative memory phenomena which support the present strength–resistance theory of storage in long-term memory. The first section discusses the phenomenon of long retrograde amnesia, which provides perhaps the single most dramatic example of the usefulness of the concept of trace resistance. The second section is concerned with documenting the previous available evidence for the existence of similarity-dependent storage interference in long-term memory as embodied in Axiom 2 of the present theory. The third section is concerned with the role that trace resistance might play in explaining why relearning of previously learned material is different from original learning of that material. The fourth section is concerned with presenting the results of a large number of studies concerned with assessing the form of the long-term memory retention function and the invariance of its decay rate as a function of time. The fifth section is concerned with assessing the validity of Axiom 4 which is concerned with the effects of multiple learning trials and the spacing of these learning trials on the form of the retention function. The sixth section is concerned with the possibility that recency judgments may be determined, in some case, by the resistance of a memory trace. This is a very natural assumption considering that resistance is assumed to be monotonically increasing with time, permitting resistance to serve as a "biological clock."

## Long Retrograde Amnesia

When a human being suffers a severe head injury, he frequently cannot remember events that occurred for some time prior to the injury. Clinical studies of this retrograde amnesia (RA) have indicated two rather different types: *short RA* and *long RA* (Russell, 1959; Whitty, 1962). Short RA refers to a permanent (irreversible) loss of memory for events occurring just a few seconds prior to the injury. The range of times considered to fall in this class is not firmly agreed upon, but it is roughly from about one sec to one min prior to the injury, with the predominant tendency being for permanent short RAs under 10 sec (Russell, 1959; Whitty, 1962). Short RA is generally interpreted as evidence for a consolidation process in the establishment of long-term memory which requires a period of seconds or tens of seconds following acquisition to form a stable long-term memory. The time for consolidation indicated by short RA phenomena is in complete agreement with the consolidation time indicated by the human experimental study by Wickelgren and Berian (1971).

Long RA refers to the less frequent loss of memory for events occurring tens of minutes, hours, days, weeks, months, and even years prior to the injury. Long RA is closely associated with the severity of the head injury as measured by the duration of the period of post-traumatic confusion in mental functioning, including post-traumatic amnesia, PTA (Russell, 1959). Usually, long RA is reversible, so that the patient gradually recovers his temporarily lost memory up to the last few seconds prior to the injury. The last few seconds are lost forever, presumably because of inadequate consolidation. Occasionally, the long RA is never fully dissipated.

The exact nature of the difference between reversible long RA and irreversible long RA is not understood, but in either case the long RA is defined primarily on the basis of trace age, not trace strength or trace importance. It is the more recently established long-term memories which are lost (either temporarily or permanently), not the weakest or the least important. When a patient recovers from long RA, the recovery is also temporally defined on the basis of trace age. The oldest memories are recovered first, that is to say the long RA "shrinks" toward the time of the injury. Similar temporally defined and shrinking RA have been found following electroconvulsive shock in humans (Williams, 1966). Russell (1959) argues persuasively that long RA phenomena indicate that long-term memories must be continually changing as a function of their age over a period of years. The concept of trace resistance handles such phenomena very nicely. However, the important thing to notice is that trace resistance, which is increasing with age, is a *different* property of the memory trace from the strength property, which determines the accuracy of recognition and recall. Strength is continually decreasing with increasing trace age (in the absence of renewal by new learning).

There is at least one major complication in the explanation of long RA by the present theory, which is that there are frequently "islands" of memory within the period of time included in the RA. That is, the subject may remember one or two brief events that occurred during the period for which he can remember nothing else (Russell, 1959). Since the "islands" tend to be near the boundary of events that can be remembered, this probably indicates that there is some variability in the degree to which memory traces of a given degree of resistance are subject to the amnesic forces initiated by a head injury. There are also other explanations of islands that are consistent with the theory.

## Similarity-Dependent Storage Interference

Axiom 2 of the theory assumes that the degradative force on a memory trace is greater the greater the similarity of interpolated learning to original learning. Considerable evidence will be presented later in this paper to indicate that the rate of decay is greater for conditions where the degree of similarity of interpolated learning to original learning is assumed to be greater.

In addition, there are a very large number of studies of retroactive interference using recognition-matching tests that are all consistent in showing similarity-dependent storage interference (unlearning). The findings are that an AB-CD paradigm produces about the same performance on a retention test as the control (rest) condition, while AB-AC or AB-CB paradigms produce greater retroactive interference, and $AB\text{-}AB_r$ produces the greatest retroactive interference of all on recognition-matching test (Delprato and Garskof, 1968, 1969; Garskof, 1968; Garskof and Bryan, 1966; Garskof and Sandak, 1964; Garskof, Sandak and Malinowski, 1965; Goggin, 1968, 1969; Goulet and Bone, 1968; Houston and Johnson, 1967; Keppel and Zavortink, 1969; McGovern, 1964; Postman, 1965; Postman and Stark, 1969). Unfortunately, the most influential of these studies, namely, Postman and Stark (1969), has led many to conclude that similarity-dependent storage interference does not occur (in recognition-matching tests).[2] Actually, Postman and Stark got differences in exactly the right directions predicted by similarity-dependent storage interference, but performance was so high in the AB-AC, AB-CB, and AB-CD conditions that the differences were not significant. However, from a trace-strength viewpoint, differences between 97% correct recognition which Postman and Stark obtained in the AB-rest and AB-CD conditions indicate substantially greater strength than 88% correct recognition or 92% correct recognition, which they obtained in the AB-CB and AB-AC conditions, respectively. Furthermore, Postman and Stark obtained recognition of only 74% in the $AB\text{-}AB_r$ paradigm which was a statistically significant difference in retroactive interference. Thus, while Postman and Stark are probably correct that similarity-dependent unlearning is a minor component of retroactive interference on various types of recall tests, it is nevertheless an extremely reliable and theoretically significant phenomenon in the study of long-term memory.

## Relearning

According to the folklore, relearning is faster than original learning and faster than learning comparable new material, no matter how much time has elapsed since original learning and no matter how little memory strength remains. Results concerning

[2] Postman does not believe that this should have been concluded from Postman and Stark (1969).

relearning may never have been stated in such an extreme form, and, in fact, only a little empirical evidence exists to demonstrate that this result (which I have called the "folklore") is true. What is known is that relearning is invariably faster than original learning or comparable new learning, even under conditions where recall performance may be extremely low (see Nelson, 1971, for the best evidence on this).

It is an interesting consequence of the present theory that relearning might well be considerably faster, even when strength was close to zero, because resistance was extremely high. This anomalous consequence of the present theory results from the fact that there is no coupling from strength to resistance. There is a coupling from resistance to strength, but not in the reverse direction. Thus, material with close to zero strength would still be considered to be increasing in resistance. This aspect of the theory seems likely to require substantial revision when the relevant data become available. However, some less extreme form of the present theory may well be true, providing a natural mechanism for accounting for facts concerning relearning (e.g., Nelson, 1971) that do not seem to be explainable on the basis of transfer of strength alone.

Actually, the most direct test of the present theory (especially Axiom 4) in relearning is to test whether the decay rate following relearning is much lower than following original learning or the learning of comparable new material.

## Strength Retention Functions

The present section describes the principal experimental findings of the present paper. The experiments were designed to assess the form of the retention function for long-term memory over a variety of different types of materials and different conditions of the learning and retention periods. Almost the entire researchable range of long-term memory was investigated: This meant a range of approximately $10^6$ on the time dimension, from retention intervals slightly under one minute to retention intervals slightly over two years. As a check on the power of the data to test the validity of the predicted "exponential-power" form ($l = \lambda e^{-\psi \sqrt{t}}$) of the retention function, a variety of other forms of retention functions were considered and rejected on the basis of the present data. Four types of retention functions were explicitly rejected on the basis of the data presented in this section: linear decay ($l = \lambda - \psi t$), exponential decay ($l = \lambda e^{-\psi t}$), logarithmic decay ($l = \lambda - \psi \log t$), and power function decay ($l = \lambda t^{-\psi}$).

The rejection of these alternative forms of the retention function and the confirmation of the form embodied in the present strength resistance theory is valid only within a statistical decision theory framework (see Green and Swets, 1966; Wickelgren and Norman, 1966) for the retrieval-decision processes involved in recognition memory. If these assumptions are valid then the discrimination measure ($d'$) that results from this analysis is linear with underlying trace strength, as assumed in the present section. If these retrieval-decision assumptions for recognition memory are not valid, then the conclusions drawn in the present section may be in error.

At the same time, it should be mentioned that the experiments in the present section are all tests of yes–no recognition memory. No tests of multiple-choice recognition memory or recall memory are included. I think that all types of long-term memory traces, whether assessed by recognition or recall, follow the same laws of trace dynamics in storage. However, the evidence presented in the present section is only relevant to assessing the validity of this theory for "yes–no" recognition memory tasks, since no multiple-choice or recall tasks were included.

*Continuous CCCDDD Experiments*

*Procedure.* On each trial, the subject saw the trial number accompanied by a CCCDDD (consonant, consonant, consonant, digit, digit, digit) item. The subject decided whether or not the item had appeared previously in the experimental session and indicated his confidence in that "yes–no" decision on a rating scale from "1" (least) to "4" (most). The subject chose one of the eight ratings: $Y4$, $Y3$, $Y2$, $Y1$, $N1$, $N2$, $N3$, $N4$. Subjects had either 2.5, 3.5, 3.75, 7.5, or 10 sec in which to read and respond to each item on any trial.

When the rate was 7.5 sec/item, subjects also rated how likely they were to recognize the item, if it was presented again after 100 intervening items. Subjects gave subjective probability ratings for this using a 12-point scale: .5, .55, .6, .65, .7, .75, .8, .85, .9, .95, .97, .99, but the ratings were collapsed into the three intervals (different for different subjects) that gave the most nearly equal division of all the responses for each subject. In addition, when the rate was 7.5 sec/item, subjects also estimated how recently they thought each item had been presented, making the assumption for each item that it *had* been presented before. The recency responses were in terms of the number of intervening items. The recency data will be discussed in a later section of this paper.

There was no intertrial interval. The timing was accomplished by a tone presented every 2.5, 3.5, 3.75, 7.5, or 10 sec. Upon hearing the tone, the subject moved a cardboard with a "window" in it down one space on a computer printout sheet revealing the next item. The last item on a page had a red line under it, signalling the subject to turn the page when he heard the next tone. No extra time was allowed for page turning and card exchanging. Subjects were told not to rehearse (think of) previously presented items, and to think only about the current item.

*Design.* All of the trials on which an item was presented for the first time in the session make up the "new" condition (to which the correct response is "no," the item did not appear previously in the session). In addition to the "new" condition, there were 14 different "old" conditions (trials on which the item had appeared previously at one of 14 different delays). The delays were: 0, 1, 2, 3, 4, 6, 8, 10, 15, 20, 30, 50, 100, 200 intervening items. There were 400 trials in a set. One set constituted a session. Sessions were given on different days. No set was given more than once to the same subject. The first set was practice and was not included in the analyses. There was a maximum of 10 different sets per subject, but some subjects took fewer sets due to a limitation of time (many of the subjects were participating in other memory experiments to be reported in this paper). Within the limitations set by having "only" 400 trials per set and having delays as long as 100 or 200 items, an attempt was made to equate the positions of test trials for the different delays. In the first place, there were 28 initial practice trials in each set which were not counted in the analysis. Following these practice trials, there were six blocks of 62 trials each. Each block had two replications of each of the old conditions, except the 100- and 200-item delays. Tests of the latter two conditions were randomly inserted between trials 120 and 400 and 220 and 400, respectively. The generation of sets was done by computer.

After the 28 practice trials, the probability of an item having been previously presented was held relatively constant throughout the session, and this probability was close to .5. Subjects were told these facts. It was emphasized that in the practice session they were to adopt criteria for using each of the eight responses such that they used each response about equally often, in particular such that they used the "yes" responses about as often as the "no" responses. Having adopted a set of criteria, they were to attempt to maintain these same criteria throughout each session and throughout the experiment. Subjects were admonished during the practice session if they deviated markedly from equal use of the response categories at any stage of the session. In addition, throughout the experiment, subjects were shown how frequently they had used each response category in the previous session in order to maintain approximately equal frequency of each category.

*Subjects.* Subjects were undergraduates who volunteered and were paid for their services. At M.I.T., six subjects were run at the 3.5-sec/item rate, three more subjects were run at the 3.75-sec/item rate, and three more subjects were run at the 7.5-sec/item rate. At the University of Oregon, five subjects were run at both 2.5-sec/item and 10-sec/item rates.

*Analysis.* Using the statistical decision theory methods described in Green and Swets (1966), Wickelgren and Norman (1966), and Wickelgren (1968), memory strength discriminability values ($d'$ values) were obtained from memory operating characteristics (MOCs) that plotted each old-item condition against the new-item condition. MOCs were fitted by least squares on the perpendicular distance of the points to the MOC line. Actually, the $d_s$ value was determined for each MOC by the intersection of the MOC with the negative diagonal. This $d_s$ value has lower variance than $d'$ (the $x$ intercept of the MOC), but is a biased estimate of $d'$, when the slope of the MOC differs from unity. To achieve an estimate of $d'$ that had the low variance of the $d_s$ estimate and was also unbiased, a linear regression of log slope (log $m$) on $d_s$ was determined under the assumption that log slope was 0 at $d_s = 0$. In other words, the parameter $k$ was estimated in the equation: $\log m = 0 + kd_s$ ($m = e^{kd_s}$). This estimate of $k$ was used to correct all $d_s$ values (to get an estimate, $d_a$, of $d'$ that was unbiased as well as having low variance) using the following equation: $d_a = \frac{1}{2} d_s(1 + e^{-kd_s})$. The $d_s$ and $d_a$ values were also determined for MOCs with only one measurable point using numerical methods and substituting into the equation: $z_c = m(d_a + z_i)$, where $z_c$ is the normal deviate associated with the probability for an old item and $z_i$ is the normal deviate associated with the probability for a new item.

Strength retention functions were determined using $d_a$ and occasionally also $d_s$ as estimates of strength in long-term memory ($l$). Thus, according to the theory (see Eq. 3), $\log d_a$ (or $\log d_s$) should be a linear function of $\sqrt{t}$, where $t$ is the retention interval. Accordingly, such plots were made, and estimates of the degree of learning parameter ($\lambda$) and the decay rate ($\psi$) were obtained by linear regression of $\log d_a$ (or $\log d_s$) on $\sqrt{t}$ with $t$ measured in minutes. Since decay of the long-term trace was not assumed to begin until about 1 min after learning, the degree of learning will be represented by the estimated strength of the memory trace at 1 min after learning, $\lambda_1 = \lambda e^{-\psi\sqrt{t}} = \lambda e^{-\psi}$.

Although virtually all of the detailed data analyses were concerned with retention functions for individual subjects, the retention functions shown in the figures for this paper are averages across all subjects. However, the type of average that was taken was that which, according to the theory being tested, would not distort the form of the retention function. Thus, in testing whether the strength retention functions for a particular condition were linear when log strength was plotted against the square root of time, the log strengths for each subject were averaged.[3] If each individual retention function is linear on a plot of log strength against the square root of time, then the average of the logs will also be linear when plotted against the square root of time. This will not hold, if one pools the results for all subjects prior to determining operating characteristics and $d_a$'s, or if one averages $d_a$'s and then takes logs, for example.

*Results.* Strength retention functions for delays greater than 50 sec for all subjects at all rates of presentation were well fit by the exponential-power function derived from the theory, as expressed in Eqs. 3 and 4. There was very little difference between using $d_a$ or $d_s$ as the measure of strength in long-term memory. Deviations of points from the best-fitting straight line on a plot of $\log d_a$ or $\log d_s$ vs $\sqrt{t}$ were generally very small. The averaged results for the 3.5-sec/CCCDDD item condition are shown by the bottom function in Fig. 1 on a plot of $\log d_a$ vs $\sqrt{t}$. There is a very slight average deviation from a straight line on this plot. The deviation is consistent with an exponent of growth of resistance slightly greater than .5, but the average deviation is not enough to discard the simpler square-root hypothesis.

[3] This idea was suggested to me by Fred Attneave.

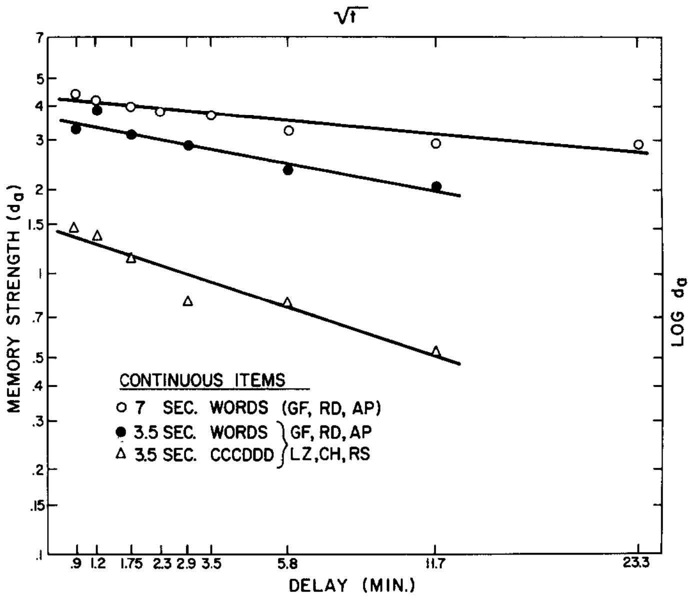


FIG. 1. Strength retention functions for the short continuous experiment, with least-squares lines determined by the exponential-power function decay hypothesis.

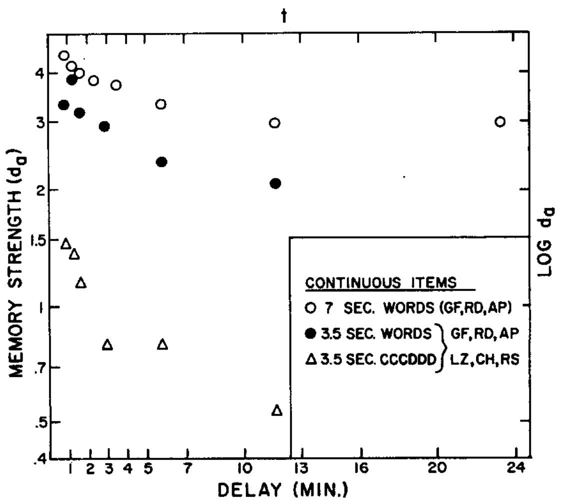


FIG. 2. Strength retention functions for the short continuous experiment on a semi-log plot (log strength vs linear time).

In contrast to the good fit of the exponential-power function (log $d_a$ vs $t^{.5}$ is a straight line), there were highly systematic deviations from both linear decay ($d_a$ vs $t$ is a straight line), and exponential decay (log $d_a$ vs $t$ is a straight line). Both of these latter two plots show the typical Jost's law finding of decreasing decay rate with increasing time. An example of the systematic deviation from exponential decay is shown on a semi-log plot in Fig. 2 for the same averaged data as that in Fig. 1. It is not possible to reject logarithmic decay ($d_a = \lambda - \psi \log t$) or power-function decay ($d_a = \lambda t^{-\psi}$) with the data from this experiment, but the results of other experiments will rule out these alternatives as well.

TABLE 1
Least-Squares Estimates of Degree of Learning ($\lambda_1$) and Decay Rate ($\psi$) in the Short Continuous Experiments
($\log d_a = \log \lambda - \psi\sqrt{t}$; $\lambda_1 = \lambda e^{-\psi\sqrt{1}} = \lambda e^{-\psi}$)

| | 7-sec Words | | | 3.5-sec Words | | | 3.5-sec CCCDDD | | |
|---|---|---|---|---|---|---|---|---|---|
| M.I.T. subjects | $\lambda_1$ | $\psi$ | $r^2$ [a] | $\lambda_1$ | $\psi$ | $r^2$ | $\lambda_1$ | $\psi$ | $r^2$ |
| GF | 3.6 | .12 | .83 | 5.0 | .18 | .66 | 1.8 | .38[b] | .93 |
| RD | 3.7 | .12 | .70 | 2.9 | .21 | .65 | 1.1 | .36 | .81 |
| AP | 5.0 | .10 | .77 | 3.6 | .18 | .60 | 1.4 | .64[b] | .84 |
| LZ | | | | 2.4 | .43 | .99 | 1.0 | .42 | .96 |
| CH | | | | 4.6 | .26 | .89 | 1.3 | .26 | .76 |
| RS | | | | 3.0 | .12 | .57 | 1.5 | .29 | .77 |
| Average | 4.1 | .11 | .87 | 3.4 | .23 | .90 | 1.3 | .39 | .91 |

| | 10-sec CCCDDD | | | 2.5-sec CCCDDD | | | 2.5-sec Words | | |
|---|---|---|---|---|---|---|---|---|---|
| Oregon subjects | $\lambda_1$ | $\psi$ | $r^2$ | $\lambda_1$ | $\psi$ | $r^2$ | $\lambda_1$ | $\psi$ | $r^2$ |
| LS | .9 | .07 | .26 | 1.2 | .27 | .96 | 3.9 | .39[b] | .98 |
| GO | .7 | .11 | .40 | .8 | .39 | .87 | 3.0 | .34 | .88 |
| DW | 1.3 | .10[b] | .73 | 1.5 | .30 | .96 | 5.3 | .39 | .99 |
| JB | 1.0 | .05 | .66 | .9 | .09 | .20 | 2.6 | .29 | .97 |
| PC | .9 | .13 | .49 | 1.0 | .45 | .74 | 4.3 | .35 | .92 |
| Average | .9 | .09 | .63 | 1.0 | .30 | .90 | 3.7 | .35 | .98 |

| | 7.5-sec CCCDDD | | | | 3.75-sec CCCDDD | | |
|---|---|---|---|---|---|---|---|
| M.I.T. subjects | $\lambda_1$ | $\psi$ | $r^2$ | M.I.T. subjects | $\lambda_1$ | $\psi$ | $r^2$ |
| DD | 1.6 | .14 | .80 | EM | .8 | .49 | .94 |
| AB | 2.3 | .21 | .77 | GR | 1.9 | .20 | .68 |
| RH | 1.7 | .20 | .90 | DS | .7 | .49 | .90 |
| Average | 1.9 | .19 | .86 | Average | 1.1 | .40 | .92 |

[a] $r^2$ = proportion of variance accounted for by theoretical function.
[b] $t$ test on difference in $\psi$ from $\psi$ for condition in center column significant beyond the .05 level.

Estimates of the decay rate ($\psi$) and degree of learning ($\lambda_1 = \lambda e^{-\psi}$) parameters for each subject at each rate of presentation are shown in Table 1. Table 1 also indicates the proportion of variance ($r^2$) accounted-for by the theoretical function. The $\lambda_1$ parameter represents the strength of the memory trace at a 1-min delay. This can be verified by examining the $d_a$ values of the best-fitting straight line in Fig. 1 at a 1-min delay.

In addition to the good fit of the exponential-power function to the strength retention data for each subject, it is also a strong point in favor of the theory that there was remarkably little variation in the estimated parameters for different subjects at the same rate of presentation. As rate of presentation decreases, the degree of learning parameter appeared to increase in the M.I.T. experiments, but remained unchanged in the Oregon experiments. However, the decay rate decreased with decreasing rate of presentation in both the M.I.T. and the Oregon experiments. Rate of decay was not proportional to presentation rate, and the theory is consistent with these results, making the reasonable assumption that $\pi$ (which is a kind of "average similarity over time" parameter) is generally greater for more rapid presentation rates.

It seemed possible that the good fit to the exponential-power function was being achieved by averaging items with vastly different decay rates and following some very different form of decay function. It was to evaluate this possibility that the subjects in the experiment with the 7.5-sec/item rate were asked to estimate their likelihood of remembering (recognizing) each CCCDDD item after a delay of 100 intervening items. It has frequently been suggested that items with a high degree of learning are forgotten more slowly than items with a lower degree of learning. If this sort of effect exists and changes the decay rate by factors of 100 or more, it could, for example, distort the apparent form of the retention functions, transforming exponential decay functions for individual items into exponential-power decay functions when all items are pooled.

The 7.5-sec CCCDDD items were divided into three groups for each subject: items given a low probability of being recognized later (low memorability), items given a medium probability (medium memorability), and items given a high probability (high memorability). The three groups were as nearly equal as possible. Strength retention functions were determined separately for each group, and the parameter estimates for all three subjects are presented in Table 2.

The results of this study of strength retention functions as affected by rated degree of memorability indicated clearly for each of the three subjects that the exponential-power function was not obtained by averaging retention functions of very different functional form. Strength retention functions showed a slight systematic deviation from exponential-power functions across the three subjects (in the direction of an exponent of growth of trace resistance greater than .5), but the effect was extremely small.

In addition, it appears that there is a decrease in the rate of decay with increasing

TABLE 2

Effects of Memorability Rating on Degree of Learning ($\lambda_1$) and Decay Rate ($\psi$) in the Short Continuous Experiments ($\log d_a = \log \lambda - \psi\sqrt{t}$; $\lambda_1 = \lambda e^{-\psi}$)

| | | Memorability rating | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | High | | | Medium | | | Low | | | $t$ test $\psi_H$ vs $\psi_L$ |
| Experiment | Subject | $\lambda_1$ | $\psi_H$ | $r^2$ | $\lambda_1$ | $\psi_M$ | $r^2$ | $\lambda_1$ | $\psi_L$ | $r^{2\,a}$ | $p$ level |
| CCCDDD 7.5 sec | DD | 2.1 | .06 | .31 | 1.6 | .19 | .58 | 1.3 | .32 | .77 | (+) .01 |
| | AB | 3.1 | .13 | .68 | 2.8 | .29 | .78 | 1.8 | .35 | .81 | (+) .02 |
| | RH | 1.0 | .33 | .66 | 2.4 | .15 | .80 | 2.4 | .28 | .82 | (−) n.s. |
| | Average | 1.9 | .17 | .76 | 2.2 | .21 | .83 | 1.8 | .32 | .83 | |
| Words 7 sec | GF | 6.1 | .01 | .44 | 6.3 | .22 | .70 | 4.6 | .19 | .64 | (+) .02 |
| | RD | 4.8 | .06 | .59 | 2.7 | .12 | .58 | 2.3 | .19 | .90 | (+) .01 |
| | AP | 6.8 | .04 | .24 | 3.5 | .05 | .35 | 2.7 | .08 | .67 | (+) n.s. |
| | Average | 5.8 | .05 | .68 | 3.9 | .13 | .73 | 3.0 | .15 | .81 | |

[a] $r^2$ = proportion of variance accounted for by theoretical function.

degree of memorability, but subject RH apparently had an incorrect conception of what items were well learned. For RH, items rated as H had both a low degree of learning ($\lambda_1$) and a high decay rate ($\psi$). Although the results for DD and AB indicate a consistent and highly significant decrease in the decay rate with increasing degree of rated memorability, the inconsistent results of subject RH prevent any definite conclusion being reached on this matter.

However, it is clear that the retention functions for each degree of rated memorability are consistent with the theory, and the decay parameters for these exponential-power functions differ by less than a factor of 10 across the three different memorability groups. Much greater differences in decay rate across different groups of items would be required to distort seriously the form of the retention function.

### *Short Continuous Word Experiments*

*Method.* The method was identical to the Continuous CCCDDD Experiments, except that English words were substituted for the CCCDDD items as the items to be learned. The words were randomly selected from a population of the 3000 most frequent English nouns, according to the Thordike–Lorge (1944) word count. The rates of presentation were 2.5, 3.5, and 7 sec/item. The same six subjects participated in the 3.5-sec/word experiment as participated in the 3.5-sec/CCCDDD experiment, and three of these same subjects also participated in the 7-sec/word experiment. The 3.5-sec/word sessions were randomly alternated with the 3.5-sec/CCCDDD sessions. The 7-sec/word experiment was run after the 3.5-sec/item experiments were finished.

Subjects made memorability ratings and recency judgments of words in the 7-sec/word experiment, in the same manner as in the 7.5-sec/CCCDDD experiment. The recency data will be discussed in a later section of this paper. The same five subjects who participated in the 2.5- and 10-sec/CCCDDD experiments participated in the 2.5-sec/word experiment.

*Results.* Strength retention functions for delays greater than 50 sec were well fit by exponential-power functions for all subjects and rates of presentation (see Fig. 1). There was very little difference between using $d_a$ or $d_s$ as the measure of strength in long-term memory. Deviations of points from the best-fitting theoretical lines were generally very small and not very systematic across subjects. Again, there was a very systematic deviation from both linear and exponential decay (see Fig. 2).

Estimates of the decay rate and degree of learning parameters are shown in Table 1. Again, the range of parameter variation across subjects was quite modest (less than a factor of 4). As in the CCCDDD experiment, the degree of learning tended to increase and the rate of decay tended to decrease with increased study time, but there were reversals on degree of learning.

Comparison of the results in Table 1 for the same subjects indicates clearly that the degree of learning was lower for CCCDDD items than for words with the same study time, but this is hardly surprising. What is of more interest is that the decay rate for CCCDDD items was greater than the decay rate for words at presentation rates of 3.5 sec/item or slower, but there was no systematic difference at the faster rate of 2.5 sec/item. The most obvious interpretation of this result is that, at a fast rate, words and CCCDDD items have the same degree of intraclass similarity ($\pi$), but words have greater potential for distinctive encoding (lower $\pi$) than CCCDDD items with increased study time. Three seconds may be a critical time, in this regard.

When the data for the 7-sec/word experiment were subdivided into three memorability groups for each of the three subjects, the resulting nine strength retention functions were still exponential-power functions as shown by the averaged results in Fig. 3, not linear or exponential decay functions. Degree of learning increased with increased rated memorability, and decay rate decreased, as shown in Table 2. The overall difference in decay rate between the low- and high-rated words for the three subjects (GF, RD, and AP) was significant at the .01 level.

### *Long Continuous Word Experiments*

*Method.* The basic procedure was identical to that for the short continuous recognition memory experiments, except that each subject worked continuously (without a break) at the recognition memory task for 3½ hr in the morning, then took a 2-hr lunch break and worked for 3 more hours in the afternoon.

The words were selected from a population of 9915 most frequent English nouns, verbs, adjectives, and adverbs, according to the Thorndike–Lorge (1944) word count.

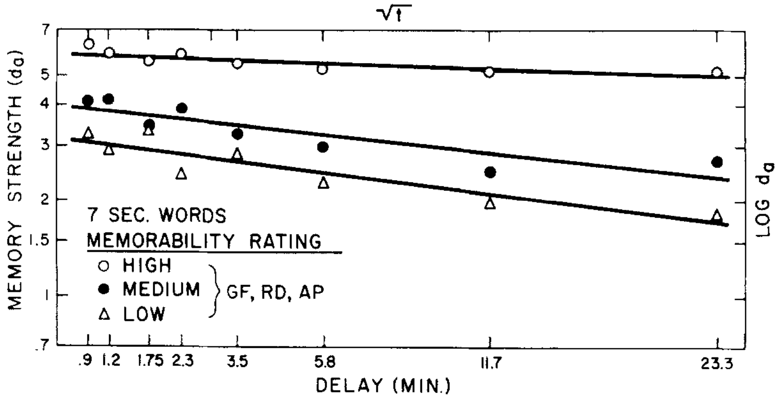


FIG. 3. Strength retention functions as a function of rated memorability of words, with least-squares lines determined by the exponential-power function decay hypothesis.

The rates of presentation were 16 words per min, 4 words per min, and 2 words per min, for three subjects, and 8 words per min, 4 words per min, and 2 words per min, for the other three subjects. The six subjects were EM, GR, DS, DD, AB, and RH, the same subjects as indicated by those initials in the CCCDDD experiments. These subjects participated in both the word and CCCDDD experiments in a randomly intermixed order. Furthermore, the order with which the subjects took the sessions for the continuous word experiments at the three different rates was also randomized differently for each subject. Subjects made memorability ratings and recency judgments of words in the same basic manner as in the short continuous word experiments. The first half-hour of the continuous sessions was considered practice and was not counted in the data analysis. The delays in the main portion of the experiment were 5, 10, 15, 20, 30 min and 1, 2, 3, 4, 5, 6, and 7 hr. Delays of 3 hr or more always crossed the 2-hr lunch break and included the 2 hr of the lunch period as part of the delay in determining trace resistance. Delays of 2 hr or less never crossed the lunch break. All the words presented in any continuous learning session were presented again either one day or seven days later, randomly mixed with the same number of new words to test even longer-term retention of the words.

*Analysis.* Separate strength retention functions were determined for each subject for each presentation rate for delays of: (a) 5 min to 2 hr during the continuous session, (b) 3 hr to 7 hr during the continuous session, (c) 1 hr to 3 hr during the lunch break, and (d) delays of 1 to 7 days during the subject's normal waking and sleeping routine. The best-fitting strength retention functions for delays of 5 min to 2 hr during the continuous session were determined in the same way as previously indicated for the short-continuous experiments. The retention functions during the continuous session for delays of 3–7 hr were also done in the same manner. Decay rates determined over the two sets of retention intervals, namely, 5 min to 2 hr and 3 hr to 7 hr, cannot be lumped together to form a single retention function from 5 min to 7 hr, because all delays of 5 min to 2 hr *do not* include the lunch break, whereas delays from 3 to 7 hr always *do* include the lunch break. Thus, since it seems likely that the $\pi$ parameter and the decay rate over the 2-hr lunch break are different from the $\pi$ parameter and the decay rate during the continuous session, we cannot consider both the short and long delays to be part of the same retention function. However, when separate decay functions are estimated for both short and long delays, one expects the decay rate parameter to be identical for both short and long delays, with the only difference being in the degree of learning parameters. The reason for this is that the time between the 5-min and 2-hr delay is composed entirely of time during the continuous session and also the time between the 3-hr and the 7-hr delay is composed entirely of time during the continuous session.

The decay rate over the 2-hr lunch break is assessed by taking the best fitting $d_a$ value estimated for the 1-hr delay from the best-fitting function for the short delays and the $d_a$ value for the 3-hr delay estimated from the best-fitting function for the long delays. These 2-point retention functions are quite variable, but they do provide some approximate idea of the rate of decay over the lunch break for comparison to the rate of decay during the continuous sessions, as estimated separately by both short and long delays.

If the theory is correct, decay rate should be at least somewhat faster at faster rates of presentation because the "average similarity over time" parameter, $\pi$, ought to be sensitive to some extent to the presentation rate. Decay rate should be substantially reduced over the lunch break, since the subject can be presumed to be learning less similar material and at a less rapid rate than during the continuous learning session.

Decay rate should be slowest of all during the 1–7-day period since this period includes both normal waking activities (similar to the lunch break) and also sleep periods. Following Jenkins and Dallenbach (1924), it is reasonable to assume that $\pi$ (and therefore $\psi$) would be lower during sleep than during wakefulness. Thus, the decay rate ($\psi$) for the 1–7-day period ought to be somewhat, though not drastically, lower than during the lunch break. Finally, separate analyses were performed for the 1–7-day delays for words that had been presented only once during the continuous session vs words that had been presented twice. If $\pi$ is lowered by greater distinctiveness of encoding of a word, we might assume that, over the 1–7-day period, the decay rate for items presented twice would be somewhat lower than for items presented once. Furthermore, if distinctiveness of encoding increases with slower presentation rate during the continuous session, then decay might be somewhat slower over the 1–7-day period and over the lunch break for words presented at a slower rate.

*Results.* Strength retention functions for both the 5-min to 2-hr delays and the 3-hr to 7-hr delays were fit extremely well by exponential-power functions for all

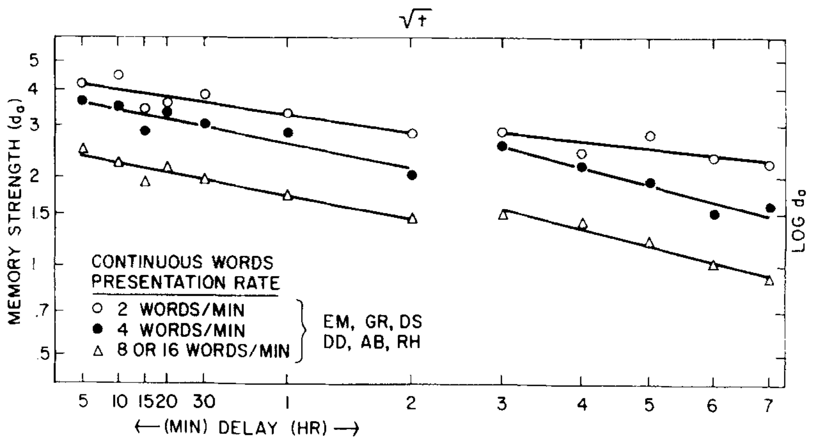


FIG. 4. Strength retention functions as a function of rate of presentation at both short and long delays in the long continuous word experiment. Least-squares lines are for the exponential-power function decay hypothesis.

subjects and rates of presentation. Figure 4 shows the average strength retention functions on a plot of log $d_a$ vs $\sqrt{t}$.

Degree of learning tended to increase with decreasing rate of presentation (increased study time), but there were a number of reversals. The reversals were probably due to the fact that only one continuous session was run at each rate, making the results confounded by interday variation. Decay rate tended to be relatively constant for presentation rates from 4 to 16 words per min but was about 40% lower at the 2-words-per-min rate. The estimates of the degree of learning and decay rate parameters are shown in Tables 3 and 4.

Present results indicate that decay rate is not proportional to presentation rate, that is, decay is not strictly a function of the number of intervening items. Rather, decay rate seems to depend upon presentation rate in a manner quite consistent with the limited aims of the theory in this regard. Namely, there is some dependence of decay rate on the density of presentation of new material to be learned during the retention interval, but this dependence can only be an estimated parameter at present.

The most critical prediction the theory makes concerning decay rate is that the decay rate should be invariant for the short delays from 5 min to 2 hr and the long delays from 3 hr to 7 hr. As indicated in Table 3, there was no systematic difference between the decay rates over the short delays vs the long delays, confirming the theory. By contrast, a power-function decay theory ($\log d_a = \log \lambda - \psi \log t$) when fitted to the present data yielded much faster decay over the 3-hr to 7-hr period than over the 5-min to 2-hr period. This is indicated by the plots shown in Fig. 5. By and large, in both this experiment and other experiments reported in the present paper, there was negligible difference in the goodness of fit of a power-function decay to the data vs

TABLE 3

Least-Squares Estimates of Degree of Learning ($\lambda$) and Decay Rate ($\psi$) in the Long Continuous Experiments ($\log d_a = \log \lambda - \psi\sqrt{t}$; $\lambda_5 = \lambda e^{-\psi_S\sqrt{5}}$, $\lambda_{180} = \lambda e^{-\psi_L\sqrt{180}}$)

| Subjects | Rate | Short delays ($t$ = 5–120 min) | | | Long delays ($t$ = 180–420 min) | | |
|---|---|---|---|---|---|---|---|
| | | $\lambda_5$ | $\psi_S$ | $r^2$ [a] | $\lambda_{180}$ | $\psi_L$ | $r^2$ |
| EM | 16 | 1.3 | .044 | .39 | .8 | .12 | .71 |
| | 4 | 2.9 | .051 | .78 | 1.9 | .11 | .30 |
| | 2 | 4.0 | .025 | .22 | 3.0 | .095 | .64 |
| GR | 16 | 3.0 | .038 | .38 | 2.2 | .050 | .33 |
| | 4[b] | 5(?) | | | 4(?) | | |
| | 2 | 4.3 | .072 | .76 | 2.0 | .037 | .56 |
| DS | 16 | 1.0 | .048 | .75 | .8 | .13 | .76 |
| | 4 | 2.5 | .023 | .56 | 1.6 | .016 | .20 |
| | 2 | 3.9 | .035 | .76 | 3.2 | .067 | .66 |
| DD | 8 | 2.9 | .065 | .91 | 1.8 | .046 | .40 |
| | 4 | 5.4 | .084 | .79 | 3.6 | .044 | .43 |
| | 2 | 5.1 | .041 | .72 | 3.0 | −.023 | .07 |
| AB | 8 | 5.7 | .097 | .86 | 3.4 | .069 | .87 |
| | 4 | 4.3 | .093 | .88 | 3.1 | .13 | .98 |
| | 2 | 4.9 | .042 | .80 | 3.5 | −.012 | .07 |
| RH | 8 | 2.7 | .050 | .85 | 1.8 | .024 | .98 |
| | 4 | 3.7 | .046 | .58 | 3.2 | .083 | .88 |
| | 2 | 3.1 | .050 | .52 | 2.5 | .022 | .24 |
| Average | 16 & 8 | 2.4 | .057 | .91 | 1.6 | .073 | .98 |
| | 4 | 3.6 | .059 | .86 | 2.6 | .077 | .93 |
| | 2 | 4.2 | .044 | .77 | 2.8 | .031 | .61 |

[a] $r^2$ = proportion of variance accounted for by theoretical function.

[b] Most of the operating characteristics had no plottable points, indicating very high $d_a$ values ($\geqslant 4$) that were not measurable with this sample size.

TABLE 4

Average Degree of Learning ($\lambda$) and Decay Rate ($\psi$) in the Long Continuous Experiments, Lunch Break, and the 1 vs 7 Day Tests ($\log d_a = \log \lambda - \psi\sqrt{t}$; $\lambda_5 = \lambda e^{-\psi\sqrt{5}}$, $\lambda_{60} = \lambda e^{-\psi\sqrt{60}}$, $\lambda_D = \lambda e^{-\psi\sqrt{1440}}$)

| Presentation Rate | Continuous | | Lunch break | | 1 Learning trial | | 2 Learning trials | |
|---|---|---|---|---|---|---|---|---|
| | $\lambda_5$ | $\psi$ | $\lambda_{60}$ | $\psi$ | $\lambda_D$ | $\psi$ | $\lambda_D$ | $\psi$ |
| 16 & 8 | 2.8 | .065 | 2.0 | .016 | 1.0 | .013 | 1.6 | .012 |
| 4 | 3.9 | .068 | 3.2 | .003 | 1.8 | .015 | 2.6 | .010 |
| 2 | 4.2 | .038 | 3.3 | .026 | 1.7 | .011 | 2.5 | .008 |
| Average | | .057 | | .016 | | .013 | | .010 |

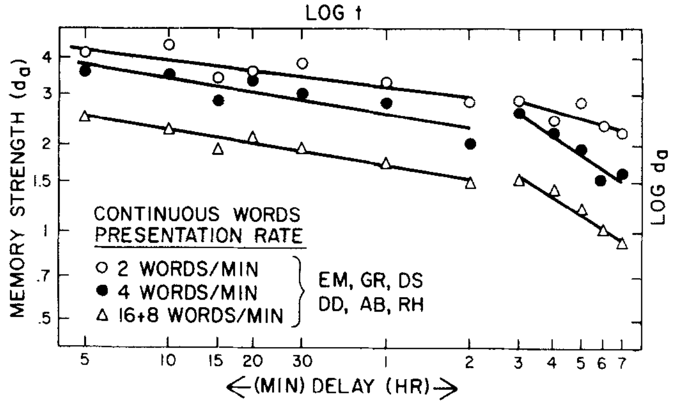


FIG. 5. Strength retention functions as a function of rate of presentation at both short and long delays in the long continuous word experiment. Least-squares lines are for the power function decay hypothesis.

the fit of the exponential-power function. However, the lack of invariance of the decay rate parameter for short vs long delays ruled out the power-function decay hypothesis.

Since the 3-hr to 7-hr delays were all tested in the afternoon, while the 5-min to 2-hr delays were tested in both morning and afternoon sessions, it could be objected that the afternoon session might have had a greater decay rate than the morning session. If this were so, then the criterion of invariance of the decay-rate parameter in this study would not be a good one for ruling out the power-function decay hypothesis (linear growth of trace resistance). To check on this, the 5-min to 2-hr retention functions were determined separately for the morning and afternoon sessions for the 8/min and 16/min conditions. There was no significant difference in decay rate at either rate for the morning vs afternoon sessions, and the slight difference that did exist was in the wrong direction for this hypothesis. Thus, the evidence from this experiment strongly disconfirms the theory in which trace resistance grows linearly as a function of time, in favor of some theory with an exponent of growth of trace resistance less than unity. In this experiment, an exponent of .5 (the square-root hypothesis) is ideal, though most of the other experiments reported in this paper would be very slightly better fit by assuming an exponent of .6 or .7.

Furthermore, in testing the power-function decay theory, there were frequent reversals of expected relations between the decay rates ($\psi$ parameters) of different conditions. These reversals of expectations occurred precisely when the lower $\psi$ values were being expected at longer delays.

As expected, decay rate during the lunch break was lower than during the continuous sessions, and decay rate over the 1–7-day period was lowest of all. Decay rate was lower for items that had two learning trials than for items that had only one learning trial, also confirming theoretical expectations.

The decay rate over the 1–7-day period tended to decrease slightly with decreasing presentation rate (increased study time). This is in accord with the hypothesis that increased study time tends to increase distinctiveness of encoding (decreasing similarity $\pi$, and therefore decay rate $\psi$). However, there was no consistent variation in the decay rate over the lunch break with presentation rate. Possibly, this lack of effect of presentation rate on decay rate over the lunch break was due to the extremely unreliable manner in which the decay rate over the lunch break was measured.

Determining separate retention functions for low, medium, and high degrees of rated memorability did not change the form of the retention function. Once again, decay rate tended to decrease with increasing degree of rated memorability, suggesting a more distinctive encoding (and therefore lower $\pi$ value) for items rated more highly memorable. However, in the case of the long continuous word experiment, the degree of learning was not consistently greater for items with high-rated memorability than items with lower-rated memorability. Thus, the presumably more distinctive encoding was not increasing the starting point of the retention function in this case, but was having a significant influence on its decay rate. It should be noted that the theory does not require any direct relation between degree of learning and decay rate. It is only because, very frequently, more distinctive encoding results in both higher degree of learning and slower decay rate that we expect to see any correlation between the two. However, in this experiment and in the Russian–English long-term memory experiment to be discussed later, the absence of a positive correlation between degree of learning and decay rate is not contradictory to the theory. For example, it is quite reasonable to expect that when a subject exerts effort to encode an item in a distinctive way, he might sometimes achieve a smaller degree of initial learning of the item as compared to when he simply rehearses the item over and over again in a more rote fashion. However, the decay rate would be very different in the two cases, being much lower for the more distinctive encoding.

### *Five-Hour–Two-Week Word Pair Experiment*

*Method.* This experiment used a study-test design. During an initial learning session of a little over three hours, each subject went through a set of 1500 word pairs attempting to form an association between the members of each pair. Subjects were instructed in the use of visual image and verbal mnemonic techniques including embedding the pairs in sentences or thinking of a mediator, etc. Subjects had only 7.5 sec to study each word pair, form some association between the members of the pair, and rate each pair for its degree of memorability. Subjects were instructed that they should divide the items into six categories for degree of memorability and base this rating on how distinctive or vivid or memorable a conceptual unit they were able to form for the pair. The first 60 word pairs were for practice and were not later tested. The remaining 1440 pairs from each session were divided into high, medium, and low memorability categories based on the subject's rating of each pair. Thus, it was necessary to collapse the subject's six memorability rating categories into three categories, attempting to make the three categories as nearly equal in number as possible. Each of the three memorability categories was then subdivided into 15 subcategories for testing at each of 15 different delays. At each delay, half of the word pairs were presented correctly paired and half were presented incorrectly paired (response members of each pair permuted so that the word pairs consisted of the same words that had been learned previously, but incorrectly paired). During the retention test, the subject decided whether each pair was correctly or incorrectly paired and stated his confidence in that decision on a scale from 1 (least) to 4 (most). During the retention test, pairs were also presented at the rate of 7.5 sec per pair. The fifteen different delays that were used varied slightly from subject to subject depending on his schedule, but a typical set was as follows: 5, 9, 12, 22, 26, 30, 34 hr, 2, 3, 4, 5, 7, 10, 12, 14 days. The basic pattern of one learning session and fifteen retention tests spread over a two-week period was repeated three times with each subject in order to achieve an adequate sample size, but due to an error the first third of the data was lost for most subjects. There were six subjects in the experiment who were paid volunteers recruited from the Student Employment Office at the University of Oregon. New word pairs were used in each of the three learning and test sequences. These subjects later participated in some of the previously reported short continuous experiments. (They can be identified by their initials in Table 1.)

*Results.* Strength retention functions were well fit by exponential-power function decay, though the points were somewhat more variable than in the continuous experiments (probably due to a smaller sample size). Averaged results for high, medium, and low degrees of rated memorability are shown in Fig. 6. As Shepard and Teghtsoonian (1961) originally noted, one of the advantages of the continuous design is that the conditions under which one measures memory strength at each delay are as close as one can imagine to being a steady state. By contrast, when one uses a study-test design, the measured strength at each point is subject to interday variability. Furthermore, the sample size for each degree of rated memorability in the present experiment was the absolute minimum because of the cost and difficulty of running even three sets of two-week sessions with each subject. Finally, due to a miscue in the experimental procedure, many of the results for the first two-week period could not be included in the final tabulation of results. Despite the greater variability, it is still possible to say that the exponential-power function fits the results without major systematic deviations.

Estimates of degree of learning and decay rate for each degree of memorability for each subject in the experiment are shown in Table 5. As indicated in Table 5, degree of learning increased with rated memorability and decay rate decreased. This is precisely what one would expect on the assumption that pairs rated more memorable

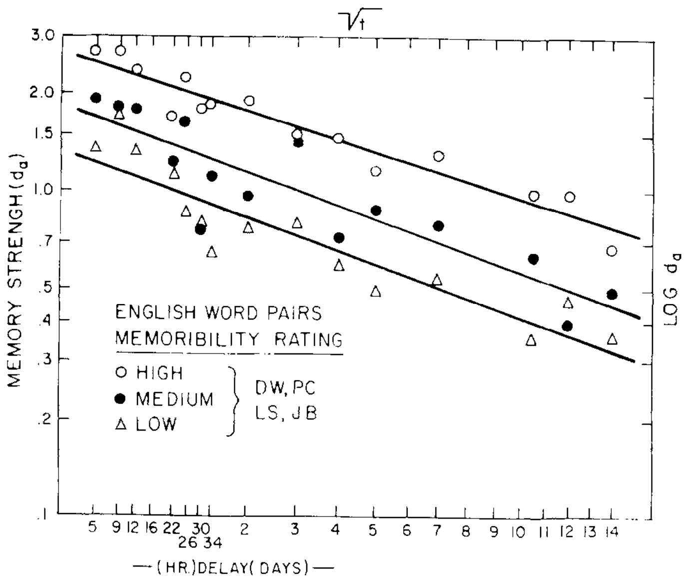


FIG. 6. Strength retention functions as a function of rated memorability of English word pairs, with least-squares lines for exponential-power function decay.

TABLE 5

Degree of Learning ($\lambda$) and Decay Rate ($\psi$) in the 5-Hour to 2-Week Word Pair Experiment ($\log d_a = \log \lambda - \psi\sqrt{t}$; $\lambda_{300} = \lambda e^{-\psi\sqrt{300}}$)

| | Memorability rating | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | High | | | Medium | | | Low | | |
| Subjects | $\lambda_{300}$ | $\psi$ | $r^2$ [a] | $\lambda_{300}$ | $\psi$ | $r^2$ | $\lambda_{300}$ | $\psi$ | $r^2$ |
| DW | 5.2 | .013 | .74 | 3.8 | .019 | .61 | 1.3 | .019 | .72 |
| PC | 2.2 | .013 | .73 | 1.5 | .012 | .52 | 1.1 | .010 | .47 |
| LS | 1.8 | .008 | .27 | .8 | .007 | .36 | .8 | .008 | .73 |
| JB | 2.0 | .003 | .15 | 1.8 | .004 | .14 | 2.0 | .006 | .43 |
| MB | 2.8 | .018 | .84 | 1.3 | .023 | .45 | .7 | .100 | .91 |
| GO | 1.2 | .004 | .20 | .7 | .007 | .15 | .3 | .059 | .17 |
| Average | 2.5 | .010 | .92 | 1.6 | .012 | .79 | 1.0 | .034 | .83 |

[a] $r^2$ = proportion of variance accounted for by theoretical function.

have a more distinctive encoding and thus a lower average-similarity-over-time parameter ($\pi$).

The range of decay rates obtained for the word pairs in the present experiment (as shown in Table 5) is approximately the same as the range of decay rates for single words over the 1–7-day period following the continuous sessions (as shown in Table 4). There is no way of telling, *a priori*, whether the similarity parameter $\pi$ ought to be greater for word pairs or for single words. However, it does seem likely that they should be of approximately the *same magnitude*, and this expectation is confirmed by the present findings.

When the results for this word pair experiment were analyzed within the context of the logarithmic decay theory ($\log d_a = \lambda - \psi \log t$) a strong positive correlation was observed between rated memorability, degree of learning, and decay rate. Thus, decay rate was faster for items rated more memorable (which also had higher degrees of learning). The same tendency to have faster decay with higher degree of learning is consistently observed across all the present experiments when analyzed within the logarithmic decay theory. This inverse correlation seems so dysfunctional and contrary to what one's intuition would suggest that the logarithmic decay theory must be considered highly improbable on the basis of the present findings.

### *Two-Week to Two-Year Russian–English Word Pair Experiment*

*Method.* In these experiments subjects were attempting to memorize Russian–English word pairs. Rather large populations of word pairs were used ranging from 240 pairs to 1000 pairs. Learning 240–1000 Russian–English vocabulary pairs to any reasonable degree of learning requires many sessions spread out over many days or weeks. Following a period of several weeks of learning, the population of Russian–English word pairs was subdivided, for testing at different delays following learning. Half the pairs tested at each delay were correctly paired and half were incorrectly paired, as in the previous experiment. Thus, a study-test design was basically used, but with the study periods spread out over several weeks.

According to the Axiom 4, the trace resistance transfers completely to subsequent increments in the trace due to multiple learning trials. Thus, in computing delays from the standpoint of the theory, one should compute the delay from the time the subject *first* encountered the word pair during learning, rather than from the *last* time he encountered the word pair during learning. This was done. Evidence to support the validity of this assumption will be presented later in the paper. However, whether one figures delay times from the beginning of learning ($t_b$) or from the end of learning ($t_e$) makes only a small difference for the present results.

During learning, subjects used a recall-anticipation procedure: looking at the Russian word and writing down an English response. Then, they turned over the card and viewed the Russian word paired with the correct English word, at which time they wrote down the correct English word and then went on to the next trial. Subjects proceeded at their own rate but were supposed to spend about 10 sec with each word pair during learning. This was controlled in an approximate way by keeping records of how long the subject took to go through an entire pack of cards.

There were actually three different Russian–English word pair experiments run over about a five-year period at M.I.T. In the first (pilot) experiment, subjects simply learned 240 Russian–English word pairs to a high criterion of 6 correct anticipations of the English response to each Russian stimulus word and then were tested at a short delay of either 3 or 4 weeks following learning ($t_b \approx$ 7 or 8 weeks)[4] and a long delay of about 12 weeks following learning ($t_b \approx$ 15 weeks). To ensure greater equality of degree of learning of each word pair, subjects only went through 120 pairs at each learning session, and the 120 pairs they were given were selected from the set of pairs that they had recalled least often. At the end, it was necessary to run learning sessions involving less than 120 pairs in order to equate the number of times that the English response had been anticipated correctly. An item was not included in a learning session when it had been anticipated correctly 5 times. When all 240 pairs reached a criterion of 5 correct anticipations, two terminal learning sessions of 120 pairs each were given with the expectation that the subject would recall every English response item correctly. This was very nearly the case. Four subjects were run in this experiment (AH, KR, RC, and CD). The basic purpose of the experiment was to determine the decay rate, and CD had to be dropped from the experiment because she performed perfectly at the first retention test (2 weeks following learning). For the two initial retention tests in this experiment subjects were tested on totally different pairs at each delay. However, following the second test delay, the pairs used at the first delay were given to the subject to determine whether they would be remembered better or worse as a result of repeated testing than the pairs which had not been tested at the earlier delay. The subjects were also brought back for two later test sessions with delays of about 65 and 118 weeks following the first trial of learning and were tested on the entire set of 240 word pairs at each delay.

In the second experiment, there were two sets of 240 Russian–English word pairs: set A was learned to the same high criterion as in the previous experiment (6 correct anticipations), but set B (which was learned after the subject had 40% of the words correct once from set A) was presented for only 5 trials on each word pair. The initial delays were 3 days and 9 weeks plus 3 days after the end of learning for set B ($t_b$ = 3–4 weeks and 12–13 weeks) and 12 weeks and 30 weeks after learning for set A ($t_b$ = 16 weeks and 34 weeks). At the longer of each pair of delays, the previous set of pairs tested was tested subsequently to determine whether the previous no-feedback retention test had any influence on memory strength. Two of the subjects (JL and JY) were brought back for a later test of the pairs learned at both high and low degrees, 86 or 100 weeks following the first trial of learning.

[4] $t_b$ is the time since the *first* learning trial (beginning of learning).

In the third experiment, subjects learned 1000 Russian–English word pairs to various degrees of learning that depended on how fast they learned and how many learning sessions we were able to run with them. Each subject's degree of learning can be roughly assessed by the number of items he had correctly anticipated once out of the possible 1000 at the end of learning. This was 295 out of 1000 for DR, 337 out of 1000 for KW, 351 out of 1000 for PK, 566 out of 1000 for SM, 625 out of 1000 for GH, and 737 out of 1000 for RS. In the final learning session, subjects rated the degree of memorability of each Russian–English word pair and the results were analyzed separately for low vs high degrees of rated memorability. Subjects were tested at each of five different retention intervals with the 1000 word pairs being divided into five groups of 200 each: 100 being correctly paired and 100 being incorrectly paired at each delay. Pairs tested at each delay that were correctly paired or incorrectly paired were composed of an equal number of low vs high memorability pairs. All subjects had these tests of *different* word pairs at each retention interval. In addition, for two of the subjects (KW and RS), after each test of a new set of 200 pairs at any given delay, all of the previously tested pairs were tested again, to see if previous no-feedback tests had enhanced or interfered with their memory strength.

*Analysis.* In plotting retention functions for the Russian–English word pair experiments the time delays used were the times since the first learning trial, rather than the time since the last learning trial for a pair, following Axiom 4 of the theory.

*Results.* Retention functions were well fit by exponential-power decay functions as illustrated by the averaged results in Fig. 7. In addition, there was no systematic effect of repeated testing on memory strength. Thus, for all cases, the results were combined for all tests at a given delay regardless of whether the pairs were being tested for the first time or second, third, or fourth time.

Estimates of degree of learning (at the shortest delay used in the experiment) and the decay rate ($\psi$) are shown in Table 6. Degree-of-learning parameters were assessed always at the shortest delay for each subject in each condition. This was a constant 4 weeks after learning for all subjects and all degrees of memorability for SM, DR, GH, KW, and RS in the third experiment. For PB, JM, JL, and JY in the second experiment, degree of learning for the pairs with a low number of study trials was assessed .5 weeks after learning while degree of learning for pairs with a high number of study trials was assessed 12 weeks after learning. Despite this, it is still evident in Table 6 that degree of learning increases with both the number of study trials and rated memorability. There seems to be no systematic effect of degree of learning or rated memorability on the decay rate. Decay rate is considerably lower than in any of the previous experiments using English words. This is what one would expect, given that the stimulus items were Russian words, which would be very dissimilar from the material the subject would be processing during the retention interval.

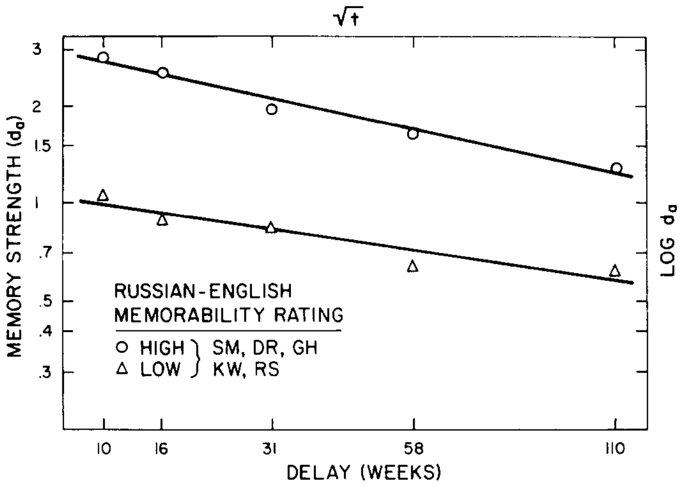


FIG. 7. Strength retention functions as a function of rated memorability of Russian–English word pairs, with least-squares lines for exponential-power function decay.

TABLE 6

Degree of Learning ($\lambda$) and Decay Rate ($\psi$) in the 3-Day to 2-Year Russian–English Word Pairs Experiments ($\log d_a = \log \lambda - \psi\sqrt{t_b}$)

| Delay (weeks) | Subjects | Memo-rability | $\lambda_{4W}$ | $\psi$ | $r^2$ | Delay (weeks) | Subjects | Study trials | $\lambda_S$ | $\psi$ | $r^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4–104 | SM | High | 2.2 | .0017 | .89 | 12–30 | PB | High | 3.9 | .0021 | * |
| 4–104 | | Low | .5 | .0005 | .75 | .5–9.5 | | Low | 3.0 | −.0014 | * |
| 4–104 | DR | High | 2.9 | .0011 | .96 | 12–30 | JM | High | 2.2 | .0010 | * |
| 4–104 | | Low | 1.0 | −.0002 | .06 | .5–9.5 | | Low | 1.7 | −.0011 | * |
| 4–104 | GH | High | 3.3 | .0013 | .87 | 12–82 | JL | High | 2.5 | .0012 | * |
| 4–104 | | Low | 1.2 | .0019 | .96 | .5–96 | | Low | 1.9 | .0012 | * |
| 4–104 | KW | High | 2.2 | .0007 | .65 | 12–82 | JY | High | 2.0 | .0009 | * |
| 4–104 | | Low | .6 | .0004 | .10 | .5–82 | | Low | .9 | .0006 | * |
| 4–104 | RS | High | 3.4 | .0005 | .41 | 4–115 | AH | High | 2.8 | .0014 | .95 |
| 4–104 | | Low | 2.0 | .0005 | .71 | 3–112 | KR | High | 2.9 | .0014 | 1.00 |
| 4–104 | Aver. | High | 2.7 | .0011 | .98 | 4–113 | RC | High | 4.0 | .0011 | .95 |
| 4–104 | | Low | 1.0 | .0007 | .89 | | | | | | |

Note: $r^2$ = proportion of variance accounted for by theoretical function.
$\lambda_{4W}$ is the degree of learning measured at a 4-week delay following learning.
$\lambda_S$ is the degree of learning measured at the shortest delay following learning.
$t_b$ is the time since the first learning trial on the pairs.
*$r^2$ is a meaningless 1.00, since only 2 points on each retention function.

## Multiple Learning Trials and Spaced Practice

According to the theory, the rate of decay of a long-term memory trace following the last of multiple learning trials will be slower than after a single learning trial, and it will be slower the greater the temporal separation between the first and last learning trial. According to the theory, the resistance of a trace starts increasing after the first learning trial and all subsequent increments to learning contributed by later learning trials start off with the resistance determined by the first learning trial. It should be noted that this does not necessarily predict slower forgetting following distributed vs massed practice, if more learning trials were allowed for massed than for distributed practice (in order to reach the same criterion of learning, for example). Most of the earlier studies of retention following massed vs spaced practice used learning to a criterion and the time between the first and last learning trial was not too different for massed vs spaced practice in most cases. Little or no effect of spacing on retention was observed in these studies, while an enormous effect of spacing was obtained by Keppel (1964) using a fixed number of learning trials and huge spacings (24 hr) between pairs of trials. Keppel's massed-practice conditions were quite superior to his spaced-practice condition at the shortest delay and grossly inferior at the longer delays. The retention functions crossed. Hence, it is difficult to argue that Keppel's results were obtained due to some correlation between degree of learning and decay rate.

All of the previous work on retention following massed and spaced practice must be considered to provide largely qualitative results because subjects are pooled to get group retention functions, very few different retention intervals are employed in any one condition, and the dependent variable (probability correct) has no theoretical justification regarding its linearity with the underlying memory trace. However, the difference in rate of decay found by Keppel (1964) was so large that it undoubtedly transcends all these objections. In addition, the following quantitative study with spacings from 1 min to 20 min in a continuous design confirms Keppel on an appropriately smaller scale and fits the theory very well.

*Method.* A continuous design was used with single words being presented at the rate of 16 words/min. Words were presented once, twice with intervals of either 1, 2, 5, or 10 min between the two learning trials, or 3 or 4 times with an interval of 5 min between each learning trial. Following the last learning trial, items were tested for retention at delays of 1, 2, 5, 10, 20, 30, or 60 min. Not all delays were tested for all conditions of number and spacing of learning trials, but the principal difference between the conditions was that a 1-min delay was used only for the one learning trial condition, so the shortest delay in all the other conditions was 2 min. The only other exceptions were that there was no 30-min delay for the 2-1 condition (2 learning trials with 1-min spacing) and no 30-min delay for the 2-10 condition (2 learning trials with 10-min spacing). Of course, to the subject, second, third, and fourth learning trials were retention tests with 1-, 2-, 5-, or 10-min delays since the last presentation, and such results are lumped together with appropriate terminal test conditions. Seven paid M.I.T. students participated in the experiment. As in previous conditions, the subjects made yes–no recognition judgments along with their confidence from 1 (least) to 4 (most). Sessions lasted 3½ hr with the first half-hour not being counted. To get an adequate sample size, continuous sessions with different words were run on the same subjects on three different days. Some of the subjects were the same as those used in previous continuous experiments, and these subjects can be identified by their initials.

*Results.* Strength retention functions were again well fit by exponential-power decay functions, with averaged results shown in Fig. 8. In accordance with Axiom 4 of the theory, the delay time is $t_b$, which is the time since the first learning trial.

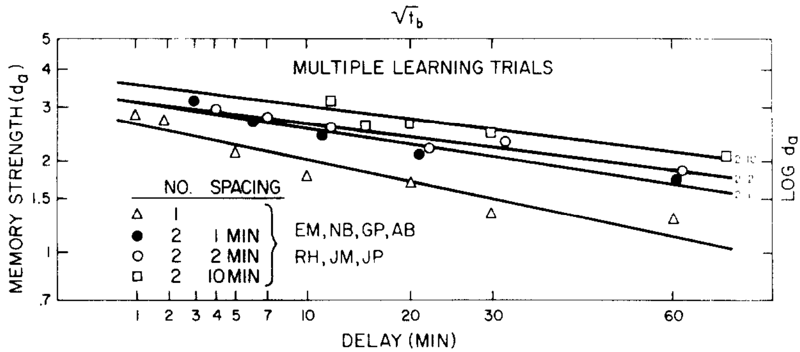


FIG. 8. Strength retention functions as a function of number and spacing of learning trials in a continuous word experiment.

There appear to be three natural alternative theories with regard to the strength retention functions folllowing multiple learning trials. First, as assumed by the present theory, the resistance of a memory trace initiated by the first learning trial might transfer completely to the increments in degree of learning contributed by subsequent learning trials. Second, the opposite might occur: Subsequent learning trials might add increments to the degree of learning, but the resistance of the entire trace (contributed by all previous learning trials as well as the present learning trial) might have to start over again at zero. Alternatively, each increment might have a resistance of its own, and one would have to add together the increments in order to get the total strength at any time following learning. In this case, each increment would have to be analyzed as having a retention function of its own.

With such small differences in spacing between learning trials as were used in the present experiment, it is not really possible to determine which of these theories provides the best fit to the data simply by examining the strength retention functions for goodness of fit. However, it is parsimonious to assume that the distinctiveness of encoding (the similarity parameter $\pi$) would be the same for two learning trials at 1-, 2-, 5-, and 10-min spacings between learning trials. Making this assumption one can ask which of the three theories yields decay rates for these two-learning-trial conditions that are independent of spacing. The theory which accomplishes this is the theory incorporated in Axiom 4, that resistance transfers completely to subsequent increments (resistance is determined entirely by the time since the first learning trial). Estimates of degree of learning and decay rate for these conditions, as well as the other conditions are shown in Table 7.

Note in Table 7, that the degree of learning tends to increase with the number of learning trials, while the decay rate decreases with increasing number of learning trials. This is interpreted as indicating that the distinctiveness of encoding generally increases with study time, which always decreases the decay rate and usually (but not always) increases degree of learning. However, with the theory that resistance transfers completely to subsequent increments, no systematic difference is obtained in decay rate across the different degrees of spacing for the various two-learning-trials conditions. This is relatively weak evidence for Axiom 4, but the enormous decrease in decay rate observed by Keppel (1964) for large spacings (1 day) between learning trials vs massing the learning trials within a few minutes provides additional confirmation for this assumption.

TABLE 7

Degree of Learning ($\lambda$) and Decay Rate ($\psi$) As a Function of Number of Learning Trials and Spacing ($\log d_a = \log \lambda - \psi \sqrt{t_b}$)

| | 1 | | 2-1 | | 2-2 | | 2-5 | | 2-10 | | 3-5 | | 4-5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Subjects | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ | $\lambda_{2e}$ | $\psi$ |
| EM | 1.7 | .12 | 1.9 | .07 | 2.1 | .08 | 2.0 | .07 | 2.0 | .10 | 2.1 | −.03 | 2.2 | .03 |
| NB | 3.4 | .13 | 3.9 | .11 | 4.5 | .15 | 4.1 | .10 | 3.9 | .06 | 5.2 | .09 | 4.1 | .01 |
| GP | 2.4 | .12 | 2.7 | .09 | 3.1 | .10 | 2.8 | .09 | 2.7 | .06 | 3.4 | .07 | 3.6 | .06 |
| AB | 4.1 | .13 | 4.7 | .09 | 4.8 | .09 | 4.9 | .07 | 5.1 | .11 | 5.6 | .06 | 4.8 | .00 |
| RH | 2.6 | .08 | 2.4 | .04 | 2.3 | .02 | 2.5 | .04 | 2.6 | .05 | 2.6 | .03 | 2.6 | .02 |
| JM | 1.0 | .14 | 1.2 | .09 | 1.0 | .03 | 1.2 | .12 | 1.2 | .09 | 1.3 | .04 | 1.1 | .01 |
| JP | 4.6 | .13 | 6.1 | .14 | 5.2 | .06 | 5.7 | .08 | 5.3 | .05 | 6.7 | .11 | 5.5 | .03 |
| Average | 2.5 | .12 | 3.0 | .09 | 3.0 | .08 | 3.3 | .08 | 3.4 | .08 | 3.7 | .05 | 3.3 | .02 |
| $r^2$ | .91 | | .92 | | .93 | | .85 | | .89 | | .80 | | .55 | |

Note: $r^2$ = proportion of variance accounted for by theoretical function.
$t_b$ is the time since the first learning trial on an item.

## Recency Judgments

People are not only capable of remembering that they have experienced some event before and what other events were associated with the cued event, they can also remember when the event took place to some extent. Introspectively, it seems clear that much of the ability to judge the recency of events that occurred weeks, months, or years ago is due to direct and indirect associations to time concepts. Thus, one can recall some of the context in which the memory was established, and these context cues directly or indirectly are associated to days, months, seasons, periods in one's life, etc., which have associations to calendar concepts. Some of one's recency memory for events occurring within the present day are likely also to be the result of direct or indirect associations to times during the day, e.g., getting dressed, eating breakfast, lunch, dinner, etc.

However, in a long, homogeneous memory experiment, with no external clock in the room, the ability to make recency judgments suggests the use of some kind of internal clock (e.g., "time tags" as suggested by Yntema and Trask, 1963). It is of interest to note that if trace resistance were to some extent a retrievable property of the memory trace, it would automatically provide a certain degree of time tagging that could mediate recency judgments.[5]

As mentioned previously, a number of the subjects in both the short and long continuous experiments made recency judgments, as well as judgments of the presence or absence of a word in the preceding set of words. That is to say, the subjects judged how long ago they thought each word had occurred, making the assumption that it had occurred previously in the list (whether or not they thought it had, as indicated by their yes–no recognition judgment). The 3 subjects in the 8/min CCCDDD short continuous experiment made these judgments The 3 subjects in the 7-sec short continuous word experiment made recency judgments. All 6 subjects in the long continuous word experiments made these judgments over a time scale from 5 min to 7 hr in the 2/min and 4/min conditions. The 3 subjects in the long continuous word experiment at the 8/min rate also made recency judgments over a range from 5 min to 7 hr. Operating characteristics were determined from the subject's judgments of recency on a rating scale from shortest to longest by plotting each condition against the shortest delay (5 min) in the long continuous word experiment and against delays of 1 min in the 8/min CCCDDD experiment and 0.93 min in the 7-sec word experiment. Results for each condition where the delay was longer than this minimum delay were scaled for their discriminability from the shortest delay in standard deviation units of the noise in the judgments of recency for that shortest delay.

The resulting growth of discriminability is plotted against real time on a linear scale in Fig. 9 for the three subjects in the 7-sec word experiment and in Fig. 10 for

[5] This idea was suggested to me by Douglas Hintzman.

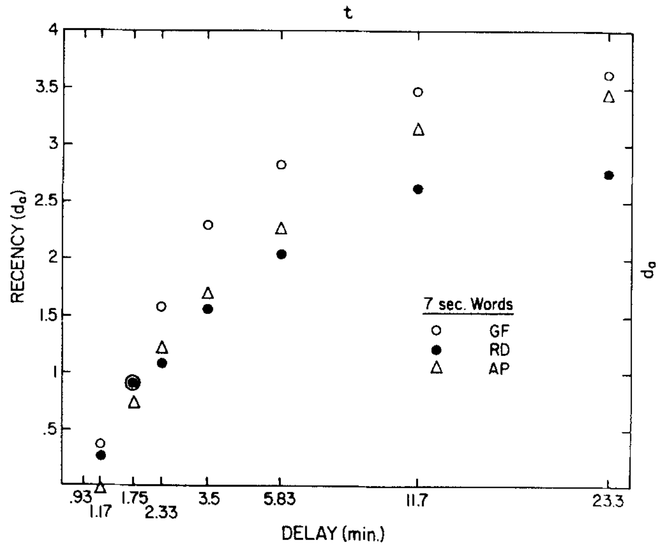


FIG. 9. Growth of recency discriminability for the 3 subjects in the 7-sec short continuous word experiment. The discriminability ($d_a$) values are scaled with the zero at the 0.93-min condition and the unit being the standard deviation of the noise in the quantity ($R$) determining recency judgments at the 0.93-min delay.

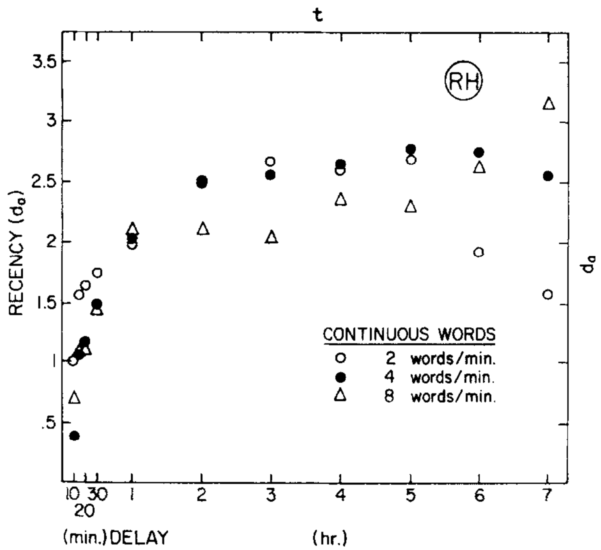


FIG. 10. Growth of recency discriminability for RH as a function of rate of presentation in the long continuous word experiment. The discriminability ($d_a$) values are scaled with the zero at the 5-min delay and the unit being the standard deviation of the noise in the quantity ($R$) determining recency judgments at the 5-min delay.

the three rate conditions for subject RH in the long continuous word experiment. Although the recency judgments are starting from 1 min in Fig. 9 and 5 min in Fig. 10, the form of the function is very similar in both cases.

Since trace resistance is assumed to increase monotonically, strictly as a function of time, in the present theory, without being influenced by all of the other factors that influence strength, trace resistance seems like an ideal memory property on which the subject could base a judgment of recency in the present experiments.

The simplest theory of the relationship between resistance and recency judgments would be that recency judgments were based on a quantity which was linear with trace resistance. This would predict, as indicated in Axiom 3, that resistance would grow in a manner proportional to the square root of time. Although the trace resistance functions such as those shown in Figs. 9 and 10 grew in a manner which was increasing as a negatively accelerated function of time, they were not well fit by a square-root function.

The only hypothesis concerning the relation between recency judgments and trace resistance that I have been able to think of so far that provides a good fit to the recency discriminability time functions is the following: the quantity judged for recency ($R$) is equal to the log of trace resistance, namely, $R = \log r$. Furthermore, to fit the data for the short continuous word experiments and the short CCCDDD experiment requires one to assume that trace resistance does not begin to grow until about 30 sec after learning. To fit the long continuous word experiment, one has to assume that here trace resistance does not begin to grow until about 3 min after learning. While this particular logarithmic hypothesis is possible, it is somewhat complex and certainly *ad hoc*.

A somewhat more attractive possibility is that the experimental design was faulty in the following manner. On each trial of these continuous memory experiments, subjects saw both the word and the trial number (mod 50). Thus, it is possible that an association of some small strength might have been formed between the trial number and the word. Since trial numbers repeated over and over again in a cycle from 1 to 50, this association to a time concept could not aid recency memory at the long delays following learning. However, this association to the trial number might contribute somewhat to recency discriminability at short delays following learning. This was precisely the nature of the deviation from the square-root hypothesis. Thus, a repetition of the experiment in which the subjects did not have access to the trial number might yield results consistant with the square-root hypothesis.

## Conclusions

The concept of trace resistance is surely the most novel contribution of this paper. A need to assume some kind of resistance property for long-term memory traces that is increasing with time, is indicated by a variety of qualitative phenomena. First, long-retrograde amnesia seems to require the assumption of a trace resistance property which is distinct from the trace strength property. Second, the present findings indicate that the decay of long-term memory traces is decreasing as a function of time since learning. While these results were obtained for yes–no recognition memory using the $d_a$ measure derived from statistical decision analysis of recognition memory process, the same conclusion regarding the deceleration of forgetting in long-term memory has been drawn by virtually every previous investigation of long-term memory retention functions using other experimental paradigms and dependent variables. Jost's second law, which embodies this qualitative statement, was formulated in the late 1800s and has stood the test of time to the present date. Third, the slower decay following spaced learning trials is also very nicely handled by assuming that traces established by spaced practice have greater resistance than traces formed by the same number of massed learning trials. The concept of trace resistance may also account for the ways in which retention following relearning differs from retention following original learning, but the data most relevant for evaluating this hypothesis are not yet available. Finally, the same trace resistance property may often mediate recency judgments, since the logical, qualitative properties of the growth of both trace resistance and recency discriminability as a function of time are identical. However, at present there is no satisfactory quantitative theory relating trace resistance and recency discriminability.

The specific quantitative theory of the dynamics of storage in long-term memory presented in this paper may well require modification at some future time. However, the evidence from a considerable range of studies of long-term memory as reported in the present paper suggests that the present theory may not be far off. If the trace resistance approach is qualitatively valid, a power-function growth of trace resistance with an exponent in the vicinity of .5 is clearly indicated by the present findings. Any dynamic function for trace resistance in which resistance approaches some upper bound will fail to fit results of the present study, unless that bound is not reached until two years or more following learning. Furthermore, to fit the results of long retrograde amnesia in which patients can sometimes lose memory for the last 10 or 20 years of their lives, this upper bound would have to be assumed not to be reached for 20 or more years following learning. Clearly, it is simpler to assume some type of unbounded function for the growth of trace resistance. At the same time it is clear from the present findings that this growth of trace resistance must be assumed to be decelerating over time. A linear growth of trace resistance will not fit the facts either.

The need to assume that the rate of decay is proportional to trace strength as embodied in Axiom 2, has only a small degree of support. This support derives from the fact that, if decay rate is assumed to be invariant with trace strength, then, in the present series of studies, decay rate is always observed to be greater for greater degrees of learning. To avoid this counterintuitive result, the theory incorporated the assumption that decay rate is proportional to trace strength.

By contrast, the assumption that decay rate is proportional to $\pi$ (the similarity of interpolated to original learning) is strongly supported by a large variety of previous studies and by the present series of experiments as well.

Qualitative evidence for Axiom 4 derives from studies on retention following spaced practice and the possibility that retention following relearning is different from retention following original learning. Quantitative support for Axiom 4 is very modest.

## References


Delprato, D. J., and Garskof, B. E. Associative unlearning of A-B following A-C or A-$B_r$ interpolation. *Journal of Experimental Psychology*, 1968, **78**, 685–686.

Delprato, D. J., and Garskof, B. E. Specific associative unlearning in the AB-CD paradigm. *Canadian Journal of Psychology*, 1969, **23**, 402–409.

Garskof, B. E. Unlearning as a function of degree of interpolated learning and method of testing in the A-B, A-C and A-B, C-D paradigms. *Journal of Experimental Psychology*, 1968, **76**, 579–583.

Garskof, B. E., and Bryan, T. M. Unlearning as a function of degree of original learning and retention test. *Psychonomic Science*, 1966, **6**, 391–392.

Garskof, B. E., and Sandak, J. M. Unlearning in recognition memory. *Psychonomic Science*, 1964, **1**, 197–198.

Garskof, B. E., Sandak, J. M., and Malinowski, E. W. A. Controlling the "fate" of first list associates. *Psychonomic Science*, 1965, **2**, 315–316.

Goggin, J. Retroactive inhibition with different patterns of interpolated lists. *Journal of Experimental Psychology*, 1968, **76**, 102–108.

Goggin, J. Retroactive interference with multiple interpolated lists. *Journal of Experimental Psychology*, 1969, **80**, 483–488.

Goulet, L. R., and Bone, R. N. Interlist form-class similarity and interference effects. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 92–97.

Green, D. M., and Swets, J. A. *Signal detection theory and psychophysics*. New York: John Wiley and Sons, Inc., 1966.

Houston, J. P., and Johnson, O. Unlearning and the unreinforced evocation of first-list responses. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 451–453.

Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley and Sons, Inc., 1951. Pp. 613–689.

Jenkins, J. G., and Dallenbach, K. M. Obliviscence during sleep and waking. *American Journal of Psychology*, 1924, **35**, 605–612.

Keppel, G. Facilitation in short- and long-term retention of paired associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 91–111.

Keppel, G., and Zavortink, B. The effect of retroactive inhibition on the response-integration stage of paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 234–239.

McGovern, J. B. Extinction of associations in four transfer paradigms. *Psychological Monographs*, 1964, **78**, Whole No. 593.

Nelson, T. O. Recognition and savings in long-term memory: Related or independent? *Proceedings of the 79th Annual Convention of the American Psychological Association*, 1971, 15–16.

POSTMAN, L. Unlearning under conditions of successive interpolation. *Journal of Experimental Psychology*, 1965, **70**, 237–245.

POSTMAN, L., AND STARK, K. Role of response availability in transfer and interference. *Journal of Experimental Psychology*, 1969, **79**, 168–177.

RUSSELL, W. R. *Brain · memory · learning*. Oxford: University Press, 1959.

SHEPARD, R. N., AND TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

THORNDIKE, E. L., AND LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

WHITTY, C. W. M. The neurological basis of memory. *Modern Trends in Neurology*, 1962, **3**, 314–335.

WICKELGREN, W. A. Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, 1968, **5**, 102–122.

WICKELGREN, W. A., AND NORMAN, D. A. Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, 1966, **3**, 316–347.

WICKELGREN, W. A., AND BERIAN, K. M. Dual trace theory and the consolidation of long-term memory. *Journal of Mathematical Psychology*, 1971, **8**, 404–417.

WILLIAMS, M. Memory disorders associated with electroconvulsive therapy. In C. W. M. Whitty and O. L. Zangwill (Eds.), *Amnesia*. London: Butterworths, 1966. Pp. 134–149.

YNTEMA, D. B., AND TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

RECEIVED: January 10, 1972