Single-trace fragility theory of memory dynamics*

WAYNE A. WICKELGREN

University of Oregon, Eugene, Oregon 97403

In single-trace fragility theory, forgetting is produced by two factors, time and interference. Memory traces are assumed to have two partially coupled dynamic properties, strength and fragility. Strength determines the probability of correct recall and recognition, while fragility determines the susceptibility of the trace to the time-decay process but not to the interference process. Consolidation is assumed to be a continual reduction in the fragility of the memory trace rather than any change in strength or availability. Decreasing fragility accounts for the continually decreasing forgetting rate, the temporal character of retrograde amnesia and recovery therefrom, and the type of internal clock necessary for nonassociative recency judgments. Data are presented to indicate that interference is independent of the interval between original and interpolated learning, that nonassociative recency discriminability approaches a limit at about 30 min, and that the decay rate of long-term retention in amnesic patients is the same as in normal Ss.

For some years now it has been fashionable to assume at least two dynamically distinct memory traces or processes, short term or long term. An enormous number of memory phenomena have been interpreted as providing support for the two-trace hypothesis. I think that all of these phenomena are equally consistent with a single-trace hypothesis, though this single trace must be viewed as undergoing a consolidation process that progressively reduces its rate of forgetting and its susceptibility to amnesic agents with increasing trace age.

Wickelgren (1973) discussed all the frequently cited psychological evidence for distinguishing two dynamic types of memory traces and demonstrated that all of this evidence was equally consistent with a single-trace theory, with three exceptions. The three phenomena alleged to support the distinction between short-term and long-term memory were (a) the different forms of retention function, (b) the different effects of similarity on storage interference (as measured by recognition tests), and (c) evidence from amnesic patients such as the classic H. M. of Scoville and Milner (1957). I now believe that a single-trace theory can account for even these three phenomena and one such theory is described in this paper. Instead of two traces, the theory proposes two mechanisms that produce decay of the memory trace: (a) a storage interference process dependent on the processing of subsequent similar material and (b) an interference-free time-decay process. The theory also contains a new hypothesis regarding consolidation based on the assumption that memory traces have two partially coupled dynamic properties, strength and fragility. Greater strength implies more accurate recognition and recall. Greater fragility implies greater

susceptibility of the trace to ~~interference~~ *time decay*. At the time of learning, the acquisition process establishes a memory trace with an initial degree of strength and an initial degree of fragility, both of which decline over time. According to the present theory, consolidation does not involve a growth in trace strength or availability over a period of time following learning. Rather, consolidation is the reduction in trace fragility (and the consequent reduction in time-decay rate) with increasing trace age. The theory is as follows.

THEORY

Let m be the strength of the long-term memory trace and let f be its fragility. Let t be the delay since learning (retention interval), $\pi > 0$ be the similarity of interpolated material to the originally learned material, $\eta > 0$ be the rate parameter for the time decay of strength, μ be the rate parameter for the decay of trace fragility, $\lambda > 0$ be the initial degree of learning (initial strength of the trace), and $\rho > 0$ be the initial level of fragility (just after learning). The following pair of partially coupled differential equations appears to provide a good fit to available dynamic memory phenomena:

$$\frac{dm}{dt} = -\eta fm - \pi m, \quad m(t = 0) = \lambda > 0 \qquad (1)$$

$$\frac{df}{dt} = -\mu f^2, \quad f(t = 0) = \rho > 0 \qquad (2)$$

Note that in Eq. 1 two factors produce decay of the memory trace. Both are proportional to the strength of the trace (m). But the first, the time-decay process, is proportional to trace fragility and independent of the similarity of interpolated learning to original learning, while the second, the interference process, is proportional to similarity and independent of fragility. The rate of trace fragility (Eq. 2) is assumed to be

proportional to the square of fragility because preliminary data fitting indicates that this works. One consequence of Eq. 2 in conjunction with Eq. 1 is that the rate of time-dependent forgetting slows down due to the decrease in fragility.

Under the boundary condition that $f(t = 0) = \rho > 0$, Eq. 2 can be solved to obtain trace fragility as a function of time:

$$f = \frac{\rho}{1 + \beta t}, \text{ where } \beta = \mu\rho > 0 \qquad (3)$$

Plugging Eq. 3 into Eq. 1 under the boundary condition $m(t = 0) = \lambda$ yields the solution for trace strength as a function of delay, namely:

$$m = \lambda(1 + \beta t)^{-\psi} e^{-\pi t}, \text{ where } \psi = \eta/\mu > 0 \qquad (4)$$

Using the methods of statistical decision theory, an interval-scale measurement of the memory strength is possible for recognition memory, making only rather weak and plausible assumptions concerning the decision process that translates strength into yes-no decisions (Wickelgren & Norman, 1966; Wickelgren, 1970, 1972). According to this retrieval decision theory, the difference between (d_m) the trace strength for an old item tested after a delay of t sec and the strength of a new (not previously presented) item has the same form of retention function as Eq. 4.

FORM OF RETENTION FUNCTIONS

Empirical strength retention functions for a variety of materials under a variety of conditions can be well fit by some type of power function at delays from 10 or 20 sec to at least 2 years (Wickelgren, 1972). By contrast, exponential decay functions provide good fit for certain retention data at delays under 10 or 20 sec (Wickelgren & Norman, 1966; Wickelgren, 1970). I had previously considered this difference in the form of the retention function for long vs short retention intervals to support a distinction between two dynamic types of memory (Wickelgren, 1973). However, it is a property of Eq. 4 that, depending upon the relative values of the rate parameters for time decay (ψ) and interference (π), the retention function can be dominated either by the power function characteristic of time decay $[(1 + \beta t)^{-\psi}]$ or by the exponential function characteristic of interference $[e^{-\pi t}]$. According to the theory, π is dependent upon the similarity of interpolated material to originally learned material. It is precisely in those tasks where exponential decay is observed that the π value for the interference process would be assumed to be highest. In these tasks the interpolated material is of the same character and is learned under exactly the same conditions as the original material. Also, the materials have been drawn from small populations of semantically homogeneous items (e.g.,

letters or digits), where phonetic coding plays a substantial role.

By contrast, with slower presentation of more semantically diverse material such as words, word pairs, or sentences, where encoding is primarily in a vastly larger capacity (lower similarity) semantic memory system, the form of the retention function appears to be some type of power function. This is precisely what ought to happen, according to Eq. 4, with π very close to 0.

SIMILARITY AND STORAGE INTERFERENCE

As discussed in Wickelgren (1973), the similarity of interpolated material to original material appears to have no effect on the allegedly short-term memory tasks (exhibiting simple exponential decay), while such similarity does affect long-term memory studies (exhibiting power function decay). However, this apparently qualitative difference in the effects of similarity on interference can be given a single-trace explanation.

According to the present single-trace theory, greater similarity of interfering learning produces greater storage interference under all conditions. However, the magnitude of the similarity effect depends upon the degree of difference between the high- and low-similarity conditions. In the "long-term" experiments, characterized by power function decay and semantic encoding, the difference in similarity between an AB-AC condition, on the one hand, and an AB-CD condition, on the other hand, is very large, because the semantic similarity (π) of almost all word pairs (A and C) is very low. In fact, it is approximately zero according to parameter estimates obtained so far for the single-trace theory. Thus, one is comparing zero similarity for the AB-CD condition with substantial (50%?) similarity for the AB-AC condition. This is obviously a substantial manipulation on the similarity dimension.

By contrast, in the supposedly "short-term" experiments, characterized by phonetic encoding and exponential decay, each item has a rather high degree of phonetic similarity on the average to any other item from populations of letters and digits. The difference between the high-similarity characteristic of AB-AC, on the one hand, and the unknown, but still quite high, similarity characteristic of AB-CD, on the other hand, will be much less. Thus, a much smaller effect of similarity should be obtained for the "short-term" studies than for the "long-term" studies. Hence, failure to find a significant effect of similarity in the "short-term" studies cannot be taken to provide any convincing evidence for two dynamically different traces.

Furthermore, an AC interpolated learning trial occurring only a few seconds after an AB learning trial may induce an S to go back and rehearse the AB pair, contrary to instructions. Such differential uncontrolled

rehearsal could easily compensate for a small negative storage interference effect. Therefore, these interference phenomena are totally unconvincing as evidence for a two-trace theory.

Finally, the present single-trace explanation for when large effects of similarity will be found and when they will not actually accounts for the one failure to find an effect of similarity in long-term retention, namely, the study of Bower and Bostrom (1968). Bower and Bostrom described their study as applying to short-term memory because the design and the materials (letter-digit pairs) were similar to those used in short-term memory experiments. But their average retention interval of 35 sec was actually sufficiently long that most experimental psychologists would consider the study to be concerned with long-term retention. Bower and Bostrom failed to find a significant effect of similarity on storage interference. This is the only contradictory study, as against at least 16 studies supporting the negative effects of similarity on storage interference in long-term retention (Birnbaum, 1973; Postman & Stark, 1972; see Wickelgren, 1972 for a listing of 14 earlier studies). It should be noted that virtually all of these 16 studies employed nonsense syllables (occasionally words or single letters) as stimuli and always used words as responses. These materials, in combination with multiple learning trials, should produce a substantial degree of semantic encoding. Since the materials and design used by Bower and Bostrom are of the type likely to produce largely phonetic encoding, the single-trace explanation is consistent with the Bower and Bostrum results, while the dual-trace explanation is inconsistent. Thus, there is actually some evidence favoring the single-trace explanation over the dual-trace explanation of the effects of similarity on retroactive interference in recognition memory.

RETROGRADE AMNESIA

In a manner analogous to that specified in the previous trace resistance theory (Wickelgren, 1972), the present theory explains retrograde amnesia following concussion-type head injuries in terms of the trace fragility property. Clinically, such retrograde amnesias cover a period of time terminated by the head injury, and the traces lost are the most recent traces, irrespective of strength (Russell, 1959). When the memories are recovered, as they often are, the recovery is also temporally defined. There is a gradual shrinkage of the retrograde amnesia, with the oldest memories being recovered first. Thus, a property of the memory trace other than strength is needed in order to explain the specificity characteristic of retrograde amnesia and that property must be monotonic with trace age, as is trace fragility. The explanation of the selectivity of retrograde amnesia is that the lost traces are those having the greatest fragility, with increasing severity of head injury producing loss of progressively less fragile traces. Similarly, in recovery from amnesia the lost traces that recover first are those having the least fragility.

Note that, although fragility is considered a dynamic property of the memory trace important for storage dynamics, the recoverability of retrograde amnesia indicates that this loss can be considered to be a retrieval phenomenon in some very broad sense. However, the temporary as opposed to permanent character of retrograde amnesia is not predicted by fragility theory and is completely irrelevant to the evaluation of fragility theory. Even though retrograde amnesia can be thought of as a retrieval phenomenon, its temporal selectivity appears to require an explanation in terms of a dynamic property of the memory trace, such as trace fragility, that increases monotonically as a function of trace age.

RECENCY JUDGMENTS

Also analogous to the earlier trace resistance theory, the current theory accounts for the ability to make certain recency judgments. In ordinary experience, a great number of recency judgments are based on direct or indirect associations to time concepts and require no accounting by any other mechanism. For example, I remember how long ago I graduated from high school, because I have direct associations to the date I graduated and to the current date. No other type of internal clock is necessary to explain such recency judgments.

By contrast, my ability to determine how long I have spent dictating this portion of the paper seems difficult to account for by association to time concepts, since I do not know the present time or the time I began. Nevertheless, under these and other such circumstances, human beings can make recency judgments that are far beyond chance (though far less accurate than a physical clock).

It is interesting to note that the decreasing fragility of the memory trace for an event will automatically provide a limited capacity for judging its recency, since fragility decreases monotonically with trace age. Assuming that recency judgments are based on fragility, one can determine the function for discriminating events presented at various times (t) in the past from a standard condition where the event was presented a fixed time ago (t_0). It is necessary to predict the discriminability of recencies for two past events because the methods of statistical decision theory provide only interval-scale measurement. The equation for recency discriminability of an event t sec in the past from an event t_0 sec in the past is

$$d_r = \alpha \left(1 - \frac{1 + \beta t_0}{1 + \beta t} \right), \text{ where } \alpha = \frac{\rho}{1 + \beta t_0}, t_0 < t \quad (5)$$

Thus, a recency discriminability function provides a direct interval-scale measurement of trace fragility, independent of strength. The theory has two important qualitative properties that can be experimentally tested by data on recency judgments. First, according to the theory, the ability to discriminate recencies approaches a limit. Second, according to the theory, the quantity

being judged, fragility, is high for short delays and low for long delays. Since variance typically increases with the mean, the theory predicts that the variance in the judged dimension for recency will be high at short delays and low at long delays. In addition, one can assess the quantitative fit of Eq. 5 to empirical recency discriminability functions for items presented at various delays.

These predictions were tested in a continuous recognition memory experiment where words were presented at a rate of 10 sec/word. The Ss made three judgments concerning each word: (a) a recognition judgment concerning whether or not the word had been presented before, with confidence indicated on a 6-point scale, (b) a recency judgment concerned with how long ago the word had been presented on a 24-point scale ranging from 20 sec to 3 h, and (c) a memorability rating concerned with how likely the S was to recognize the word should it be presented 1 h later on a 12-point scale from a probability of .5 to .99. A memory drum that used IBM computer paper presented the words, with no indication of the trial number. There were no clocks, watches, or other time-keeping devices available to the Ss.

The delays between presentations of the same word were 1, 1.5, 2, 3, 5, 7, 10, 15, 20, 30, 45, 60, 75, 90, 105, and 120 min. There were 60 practice trials followed by 1200 regular trials, so that a session lasted 3½ h. Five Ss had 1 practice and 12 regular sessions. Recency discriminabilities (d_r) were assessed by operating characteristics using the 1-min delay condition as a standard against which all other conditions were compared (see Wickelgren, 1972).

The results of the experiment confirmed both qualitative and quantitative predictions. The recency discriminability functions approached a limit (by about a 30-min delay). For each S the variance in the judged recency dimension was high at short delays and decreased systematically with increased delay (slopes of
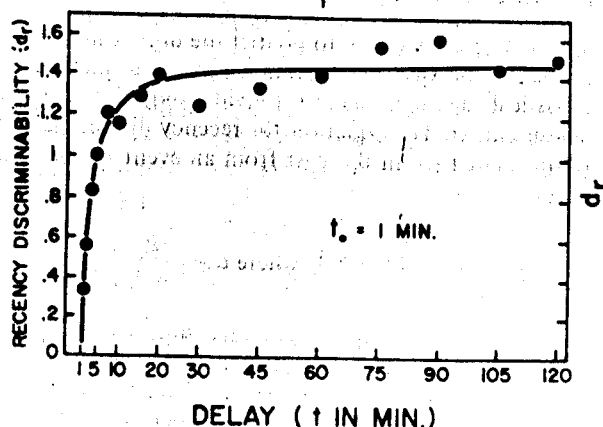

Fig. 1. Recency discriminability of an event occurring t sec ago vs an event occurring at a standard delay of t_0 = 1 min ago.

operating characteristics were consistently greater than unity and increased with increasing delay). Finally, the quantitative fit of Eq. 5 to the empirical recency discriminability function over delays from 1 min to 2 h was extraordinarily good, accounting in the averaged data for 96% of the variance, with α = 1.50 and β = .022 (t in seconds). The fit of the function to the averaged experimental data is shown in Fig. 1.

INTERFERENCE INDEPENDENT OF DELAY

The earlier trace resistance theory was a pure interference theory of forgetting in long-term memory, and it assumed that increasing the resistance of the memory trace made it progressively less and less susceptible to interference as a function of time. The resistance theory, thus, made the prediction that highly similar interpolated learning would cause greater interference if presented early in the retention interval than if presented later in the retention interval.

Previous studies using conventional recall and modified-modified free recall (MMFR) by Archer and Underwood (1951), Houston (1967), Howe (1969), and Newton and Wickens (1956) have yielded inconsistent results. But none of them found any evidence in support of the prediction that similar learning causes greater interference when interpolated earlier in the retention interval. However, such recall measures are sensitive to a variety of response competition, response suppression, and other retrieval interference factors, in addition to any storage interference factor. Thus, a recognition measure is a more appropriate test of the theory. However, Howe (1969) also used a recognition matching test and found results inconsistent with the theoretical prediction.

Nevertheless, I performed two "yes-no" recognition memory experiments to test the relation between interference and delay of interpolated learning. In the first, 12 Ss studied word pairs for 7.5 sec each under mnemonic instructions (embed the word pair in a sentence, think of a mediator, or construct a visual image). The Ss were given some preliminary training in the use of these mnemonic techniques, followed by a 5-min practice period in which word pairs were presented for 2½ min and then tested over the next 2½ min. Following this, 600 word pairs were presented over a learning period of 75 min, followed by "yes-no" recognition testing of the pairs over the next 75 min. In the test phase, Ss were also given 7.5 sec to make a "yes-no" decision concerning whether the word pair had been previously presented as a pair, stating their confidence on a six-point scale. There was a constant retention interval of 75 min between each correct pair and its test. Incorrect pairs were obtained by permuting response members maintaining an exact 75-min retention interval for the stimulus member and as close to a 75-min interval as possible for the response member (to prevent discriminating correct and incorrect pairs on some nonassociative basis using relative recency). The

principal manipulation involved whether or not an A-B pair was followed by a single A-C pair or only C-D pairs. If there was a similar interfering learning trial (A-C), the interval between original and interpolated learning was either 15 sec, 2 min, or 16 min. In the interference conditions, both original and interpolated learning pairs were tested and the discriminability of correct and incorrect such pairs was assessed using the d_m measure. The average d_m was 2.23 for the control condition and 2.11, 2.17, and 1.90 for the 15-sec, 2-min, and 16-min conditions of testing the A-B pair when followed by an A-C pair. Using a Wilcoxon matched-pairs signed-ranks test, the difference between the control condition and the average interference condition was significant ($p < .05$, one-tailed test), but there were no significant effects of the interval between original and interpolated learning. Tests of the A-C interpolated learning pair yielded d_ms of 1.72, 1.81, and 1.88 for the 15-sec, 2-min, and 16-min conditions, respectively. The average of these conditions differed from the control condition by the Wilcoxon test at well beyond the .01 level. All Ss showed this proactive interference effect (whether due to acquisition, storage, or retrieval being, of course, not determinable in this experiment).

A more conventional continuous recognition memory experiment for word pairs with retention intervals varying from 1.25 to 75 min was also performed to investigate the same question and distinguish any differences in acquisition vs decay rate as a function of delay between original and similar interpolated learning. Word pairs were again presented for 7.5 sec/pair using the same mnemonic instructions and training. The delay between original and similar interpolated learning was 15 sec, 1 min, or 3.75 min. Once again there was a significant superiority of the control condition to the interference conditions, but no consistent effect of the interval between original and interpolated learning. Furthermore, the interference effect of the similar A-C interpolated learning trial appeared to be entirely restricted to the period of A-C interpolated learning, as the subsequent retention function for the A-B learning trial (and the A-C interpolated learning trial) was parallel to the control condition. That is to say, the decay rate was about the same for all interference conditions as for the control condition, but the A-C interpolated learning trial produced a significantly greater reduction in the strength of the A-B pair than did the C-D interpolated learning trial at the same place in the control condition.

Thus, it appears that similar interpolated learning produces greater interference than does less similar interpolated learning, but the effect is independent of the interval between original and interpolated learning. There appears to be no way to account for this finding within a pure interference theory and yet preserve the increasing trace resistance so important in accounting for the decreasing rate of forgetting, the temporal nature of retrograde amnesia, and recency judgments. It was primarily because of the results of these similarity

inference experiments that I developed the two-process theory, in which the decreasing trace fragility affects the time-decay process but not the interference process. It is easy to demonstrate mathematically that the current theory predicts a negative effect of similar interpolated learning but an effect that is independent of the point of interpolation in the retention interval.

ACQUISITION AMNESIA

For some time it was thought that certain types of amnesic patients provided evidence for the distinction between two dynamically different types of memory traces, short-term and long-term memory. These are patients such as those described by Milner (1966) who have a very specific deficit in the ability to establish new long-lasting cognitive memory traces, while exhibiting little or no deficit in short-term memory span and the retention of previously established long-term memory.

However, it is now well established that such amnesic patients have relatively normal perceptual and motor skill long-term learning and memory despite their severe deficits in the establishment of new cognitive memory traces (Baddeley & Warrington, 1970; Corkin, 1968; Warrington & Weiskrantz, 1970). These findings indicate that such amnesic patients have a modality deficit in the formation of cognitive (semantic) memories rather than a general deficit in the establishment of "long-term" memory traces. The relative lack of impairment in memory span and certain other short-term memory tasks is, therefore, explained by patients having relatively unimpaired capacity for encoding in a phonetic modality, while having severely impaired capacity for new encoding in the semantic modality. Under the conditions of most verbal learning studies, phonetic traces will be rapidly interfered with and, therefore, lost. Hence, amnesic patients will show severe deficits at long retention intervals. Since these amnesic patients are not generally impaired in the ability to form new long-term memory traces, the modality explanation seems clearly more attractive at the present time than the dual dynamic trace explanation.

The results of a continuous recognition memory study I did some years ago at MIT with the amnesic patient H. M. of Scoville and Milner (1957) lends some general support to the single-trace position and to the present theory. In the experiment, words were presented at the rate of 3.5 sec/word, with the S instructed to decide "yes" or "no" regarding whether each word had been presented earlier in the list. The retention interval varied from a few seconds to 11.7 min. [See Wickelgren (1972, 1974) for a more detailed description of the experiment conducted previously with a group of six normal Ss.] The results for H. M. and for the six normal Ss are presented in Fig. 2 along with the predictions of the single-trace theory (Eq. 4) under the assumption that $\pi = 0$ and $\beta = .022$ (same β as found for recency discriminability). In addition, the decay rate was constrained to have the same value ($\psi = .29$) for H. M. as
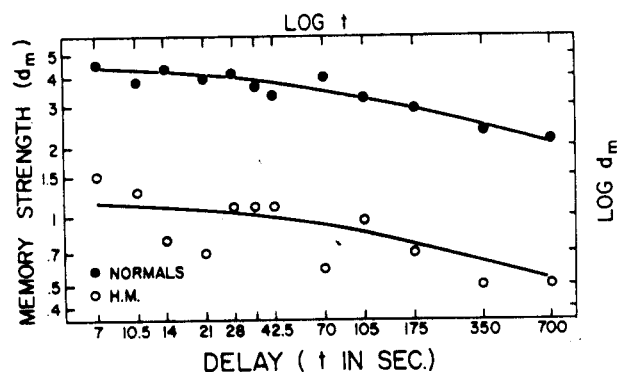
Fig. 2. Strength retention functions for recognition memory for normal Ss and for the amnesic patient H. M.

for normals. The only free parameter estimated for H. M. was the degree of learning, $\lambda = 1.2$, contrasted with $\lambda = 4.6$ for normals.

Although H. M. showed a very slight deviation in the direction of greater decay rate than normal Ss, any difference in decay rate between H. M. and normal Ss is clearly very minor. It seems parsimonious to conclude that amnesic patient H. M.'s deficit is in the initial degree of learning (by a factor of about four), with no difference in subsequent consolidation or storage.

Moreover, there is good reason to consider this acquisition difference between amnesics and normals to be a cognitive modality deficit rather than a general dynamic deficit of long-term memory. Thus, the findings for amnesic patients do not appear to favor a dual-trace theory over a single-trace theory.

REFERENCES

Archer, E. J., & Underwood, B. J. Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. Journal of Experimental Psychology, 1951, 42, 28-290.

Baddeley, A. D., & Warrington, E. K. Amnesia and the distinction between long- and short-term memory. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 176-189.

Birnbaum, I. M. Retroactive inhibition in two paradigms of negative transfer. Journal of Experimental Psychology, 1973, 100, 116-121.

Bower, G. H., & Bostrom, A. Absence of within-list PI and RI in short-term recognition memory. Psychonomic Science, 1968, 10, 211-212.

Corkin, S. Acquisition of motor skill after bilateral medial temporal-lobe excision. Neuropsychologia, 1968, 6, 255-265.

Houston, J. P. Retroactive inhibition and point of interpolation. Journal of Verbal Learning & Verbal Behavior, 1967, 6, 84-88.

Howe, T. S. Effects of delayed interference on list 1 recall. Journal of Experimental Psychology, 1969, 80, 120-124.

Milner, B. Amnesia following operation on the temporal lobes. In C. W. M. Whitty and O. L. Zangwill (Eds.), Amnesia. London: Butterworths, 1966.

Newton, J. M., & Wickens, D. D. Retroactive inhibition as a function of the temporal position of the interpolated learning. Journal of Experimental Psychology, 1956, 51, 149-154.

Postman, L., & Stark, K. On the measurement of retroactive inhibition in the A-B, A-D paradigm by the multiple-choice method: Reply to Merryman. Journal of Verbal Learning & Verbal Behavior, 1972, 11, 465-473.

Russell, W. R. Brain, memory, learning. Oxford: Oxford University Press, 1959.

Scoville, W. B., & Milner, R. Loss of recent memory after bilateral hippocampal lesions. Journal of Neurosurgery & Psychiatry, 1957, 20, 11.

Warrington, E. K., & Weiskrantz, L. Amnesic syndrome: Consolidation or retrieval? Nature, 1970, 228, 628-630.

Wickelgren, W. A. Time, interference, and rate of presentation in short-term recognition memory for items. Journal of Mathematical Psychology, 1970, 7, 219-235.

Wickelgren, W. A. Trace resistance and the decay of long-term memory. Journal of Mathematical Psychology, 1972, 9, 418-455.

Wickelgren, W. A. The long and the short of memory. Psychological Bulletin, 1973, 80, 425-438.

Wickelgren, W. A. Strength-resistance theory of the dynamics of memory storage. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), Contemporary developments in mathematical psychology. San Francisco: Freeman, 1974.

Wickelgren, W. A., & Norman, D. A. Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 1966, 3, 316-347.