Offprints from
COGNITION IN LEARNING AND MEMORY
Edited by Lee W. Gregg

CHAPTER 2

Wayne A. Wickelgren
University of Oregon

CODING, RETRIEVAL, AND DYNAMICS OF MULTITRACE ASSOCIATIVE MEMORY

Perhaps the most useful theoretical distinction in the field of memory has been the distinction between three phases of the memory process: acquisition, storage, and retrieval. I propose that a substitute trichotomy will prove more useful in the analysis of the memory process, namely: coding, retrieval, and dynamics.

Coding refers to the internal representation in memory of events and the relations between events, including such topics as: the modalities of memory, the similarity functions defined over pairs of events, the dimensions of similarity spaces and the loci of events in these spaces, the associative or nonassociative nature of any memory modality (content addressability, uniqueness of representation, contiguity conditioning), the coding of events and associations of events making use of previously learned cognitive structures (learned concept representatives and associations between concept representatives), and, finally, the number of traces within any one modality, each mediating memory for a different period of time or with other differences in their trace properties. Coding refers to the structure of memory, what its components are and how they are organized into a system. In short, coding is concerned with the question, "What is learned?"

Retrieval refers to how such memory traces are used in a variety of situations. Having specified what the different traces are in the coding assumptions of a theory of memory, it is necessary to say which of these traces are used in various situations, that is, "What is judged in retrieval?" Also, we must specify the decision rules which translate the values of the traces that are judged into responses in different types of judgment tasks.

Dynamics refers to the time course of the strengths of different traces through the various phases of the memory process. However, it seems worthwhile to distinguish four phases: acquisition, consolidation, decay, and retrieval, in place of the more conventional three phases: acquisition, storage, and retrieval. Acquisition refers to the period during which a potential trace is formed, often as a result of the input of new information. However, potential traces are assumed not to be retrievable until after they have been converted into retrievable traces by the consolidation process. In the case of long-term memory, LTM (days to years), the period of consolidation could be hours or days. In intermediate-term memory, ITM (minutes to hours), the period of consolidation could be tens of seconds. If this is so, then in either case, it would be quite important to explicitly characterize consolidation as a phase of the memory trace, distinct from both acquisition and decay. Undoubtedly, short-term memory, STM (1-20 seconds) or very-short-term memory, VSTM (generally less than 1 or 2 seconds) consolidates so quickly that the distinction between acquisition and consolidation is a mere formality. Decay refers to the period during which the retrievable trace is subject to degradation from a variety of possible sources. Retrieval, as a phase in trace dynamics, focuses on how retrieval of a trace affects its strength. For example, is retrieval destructive, nondestructive, or constructive? To characterize the dynamics of any memory trace, there must be laws specifying the form of the time function for each phase and how the parameters of these functions depend on the type of trace, the modality, and the conditions. Finally, there must be a law specifying how to combine the phases.

OVERVIEW OF MULTITRACE THEORY

Multitrace theory is a modification and extension of the theory described in Wickelgren (1969e). The basic assumptions of multitrace theory fall into three categories: coding, retrieval, and dynamics of memory traces.

Coding

There appear to be two basic types of memory structures, associative and nonassociative. In associative memories, each event or concept has a unique internal representative, and the internal representatives have different degrees of association to each other depending upon how frequently they have been contiguously activated. In nonassociative memories, there is an ordered set of locations (cells, registers, boxes, etc.) into which the internal representative of any event or concept can be coded, and sequences of events or concepts are stored in order in this ordered set of locations. A tape recorder is a good example of a nonassociative memory. From a hardware viewpoint, virtually all computer memories are also nonassociative, though with suitable programming, an associative memory can be simulated. Multitrace theory assumes that human memory is largely associative, with the exception of very short-term sensory memory (e.g., visual and auditory "afterimages" of one kind or another), which is probably nonassociative.

In specifying more precisely the nature of the internal representative of any event, it seems to be necessary to distinguish between the representation of the *attributes* of the event in whatever sensory or motor modality or modalities are relevant and the representation of some more general *concept* which is cued by the event under some particular interpretation of the event. Many events have more than one conceptual interpretation, with the frequency of each interpretation being manipulable to some extent by the conditions surrounding presentation of the event. This can complicate matters to whatever extent one wishes, but there are many situations under which particular conceptual interpretations of each of a set of events can be made overwhelmingly predominant.

For example, in visual presentation of a list of letters, there must be a visual attribute representation of these letters at one stage of the representation process. In a very young child, this may be all the representation these letters have. In an adult, the visual attribute representation of the letters is strongly associated to a conceptual representation of each letter. If the list of letters forms a word, the visual attribute representation and/or the conceptual representation of the letters is associated to the conceptual representation of the word. The conceptual representation of a letter can also be activated by auditory and tactile stimulation, and the variety of different visual, auditory, and tactile patterns that can activate the same concept representative for most concepts is astounding. Certainly, no one conjunction of attributes is common to all events which activate the representative of most concepts. Thus, we cannot identify the conceptual (semantic) modality with any sensory modality. Nor can we identify the conceptual modality with any motor modality, as there are generally representatives in at least two motor modalities (speech and writing) associated with each concept and frequently several different representatives in each of these two modalities associated with the concept (e.g., different ways of speaking or writing the same word).

The simplest, adequate way to describe the relation between attribute representatives and concept representatives is to say that concept representatives are disjunctions of conjunctions of attributes. Elsewhere, I have argued that single neurons could be concept representatives, if it is assumed that there are free neurons which are unspecified genetically and come to be specified by learning (Wickelgren, 1969d). The plausibility arguments for this position will not be repeated here, but the present version of multitrace theory does assume a process of learned representation of concepts which is consistent with this position.

Thus, multitrace theory assumes that there is a cognitive (concept) modality in addition to a variety of sensory and motor attribute modalities. Since associations apply to ordered pairs of internal representatives, the number of possibly different modalities of memory is the number of different ordered pairs of sensory, motor, and cognitive modalities.

Multitrace theory distinguishes four types of associations within or between modalities: (a) interevent associations (direct forward and backward associations between the representatives of successively presented events), (b) intraevent associations (associations between the representatives of the simultaneously presented components of an event), (c) concept associations (associations between the representatives of the simultaneously presented components of an event and a concept representative for the event), and (d) structured associations (associations between the representative of an event and cognitive structure representatives such as the familiarity representative, serial position concepts like beginning, middle, and end, subgroup labels, syntactic structures, visual images, mediators, mnemonics, rules, etc.).

Simultaneous or immediately-successive contiguity of activation of internal representatives is assumed to produce learning of the associations between these representatives in all but the initial learned specification of each concept representative. Within the present framework, concept learning requires the contiguously activated representatives to become associated to an input site of a previously unspecified representative.

Associations involving cognitive structure representatives undoubtedly play an increasingly important role in learning as the child develops, and structured associations probably play a dominant role in adult learning. Even when the task is to associate two events (A and B) in paired associate (PA) learning, it has become increasing clear that the A-B association is very frequently mediated by some cognitive-structure representative, which may be represented abstractly as A-M-B. In many cases it may be necessary to consider performance in PA learning to result from a combination of a direct A-B association and a "mediated" A-M-B association.

In addition to the different modalities of memory and the different types of associations within and between modalities, multitrace theory assumes that there are different traces mediating each type of association in each modality, namely, short-term memory (STM), intermediate-term memory (ITM), and long-term memory (LTM). STM is assumed to have a time constant in the range from 1 to 10 seconds, ITM a time constant from 2 minutes to several hours, and LTM from days to years.

Finally, mention must be made of sensory very short-term memory (VSTM), which clearly exists at least in vision (Sperling, 1960; Averbach and Coriell, 1961; Averbach and Sperling, 1961; and perceptual work on after images). It is not yet clear how many different types of visual VSTM there may be, nor are the properties of any one kind of visual VSTM completely specified. However, it does seem likely that visual VSTM is a nonassociative memory, where any pattern can be imposed on any location of an overlapping set of locations in a two (or three?) dimensional visual space. Associations are probably not formed between adjacent patterns (i.e., letters in array) in visual VSTM. Exactly such results were obtained for visual VSTM by Wickelgren and Whitman (1970).

Retrieval

There are two principal aspects to the analysis of retrieval: (a) What memory traces are judged in retrieval? and (b) What judgment rule is used? Multitrace theory designates two basic types of elementary retrieval processes, recall and recognition. Recall is a comparative judgment on a set of strengths corresponding to the alternative responses. The simplest judgment rule for recall seems to be the maximum rule (choose the alternative with maximum strength), which is the same as Thurstone's (1927) Law of Comparative Judgment. Recognition can be either a comparative judgment (as in multiple-choice tests), or an absolute judgment (as in "yes-no" recognition) on the total strength (across modalities and traces) of some association. The simplest judgment rule for absolute-judgment recognition seems to be the criterion rule (choose the one of a set of ordered responses whose criteria bracket the total strength) of Thurstonian successive intervals scaling and signal detection theory (Tanner and Swets, 1954.)

Nonassociative VSTM only permits recall of the representative in a location, with the recalled representative being the one with the greatest VSTM strength of activation. It seems likely that any attempt to superimpose a test item on the location of the previously presented item would simply destroy the VSTM trace for the previously presented item. Presenting the test item in a location sufficiently removed to avoid this problem might well result in above-chance recognition performance, but, according to multitrace theory, subjects would be recalling the test item and answering "yes or no" depending on whether what they recalled matched the test item. This is not elementary recognition, but a more complex retrieval process based on the elementary recall process.

Associative memory permits both recall and recognition. What is judged in recall are the comparative strengths of associations from the representatives of the cue event to the representatives of the alternative responses, either direct associations or indirect associations via structure representatives.

What is judged in recognition varies greatly with the nature of the recognition task. In event recognition or recency memory, the strength of association between the event and the familiarity representative is assumed to be judged. In associative (or serial order) recognition memory, the sum of the direct and indirect associations between the representatives of the two test events is assumed to be judged.

Many retrieval processes are quite complex, such as trying to remember someone's name when it is not immediately recalled. However, according to multitrace theory, all such complex retrieval processes are, in principle, analyzable as a sequence of elementary retrieval processes.

Dynamics

The most fundamental assumption of multitrace theory regarding trace dynamics is the analysis into four phases: acquisition of potential strength, consolidation of retrievable strength, decay of strength, and retrieval of strength. With appropriate (non-)rehearsal instructions and other conditions of presentation, it is reasonable to assume that acquisition can be confined to the period of occurrence of each event prior to the presentation of the next event. Consolidation takes place perhaps partly during the occurrence of the event, but, for ITM and LTM, largely or entirely after the event and during the acquisition or other processing of subsequent events. It is reasonable to suppose that the lag from event onset or offset to the beginning or end of consolidation is greater for LTM than for ITM or STM, if indeed there is any consolidation lag at all for STM. Decay begins as soon as the trace is fully consolidated. In retrieval, the strength of the trace at the moment of retrieval is the input to the judgment processes. This terminates a complete cycle of the four memory phases, though retrieval may initiate further acquisition followed by consolidation and decay.

Each phase has its own dynamic function (strength as a function of time). The retrievable strength at any time in the acquisition period prior to consolidation is zero. During consolidation it is the product of the acquisition function and the consolidation function. During the decay period, retrievable strength is the product of all three functions. This multiplicative combination of phases is obviously valid if the phases are strictly successive and nonoverlapping. If the phases overlap, such as would occur if decay of already consolidated strength occurred during the consolidation of new strength, then multiplicative combination of phases is less obviously valid. However, one can interpret the acquisition and consolidation phases to include the effects of the decay process for already consolidated strength in addition to the effects of the true acquisition and consolidation processes on the increment in trace strength due to the last presentation of an event. Now the phases in multitrace theory are successive and strictly nonoverlapping, despite some overlap in the true phases. The advantages of a simple multiplicative combination of dynamic functions are so great that some added complexity in the acquisition and consolidation functions is a small price to pay. Also, it is not clear that there is any added complexity in the acquisition and consolidation functions, since these functions are likely to be approximations without an elegant derivation from an underlying more molecular theory.

It seems necessary to have three different acquisition functions, one for associative acquisition, one for nonassociative acquisition, and one for concept acquisition. Since there is nothing known regarding the dynamics of concept acquisition, in the present sense of the term, there is little point in stating anything but that activating a set of representatives in simultaneous contiguity increases the strength of association of each to an input site of the concept representative for which this set is a cue. Note that in order for a concept representative to be a disjunction of conjunctions of attribute representatives, the strengths of associations from attribute representatives should sum only within each conjunction. That is the purpose of specifying that a concept representative has many input sites.

The associative acquisition function is assumed to be the product of two exponentials which approach a limit as a function of the durations of the two events presented in simultaneous or immediately successive contiguity. The limit is formulated with sufficient generality to include the effects of contiguous activation of two internal representatives *i* and *j* on the association from each to the other (forward and backward associations in successive contiguity) and also the generalized increment in strength of association between any two representatives *k* and *l* resulting from contiguous activation of *i* and *j*. The accounting for generalized strength depends upon the estimation of similarity functions for any unordered pair of internal representatives in the modality involved. There is no provision for remote associations between successively activated event representatives, except indirectly via serial position representatives. This may or may not be the correct formulation, but it turns out to be simpler to handle it this way. I know of no evidence that distinguishes between the relative contributions of remote event-to-event associations and indirect effects due to generalization of associations with serial position concepts.

It should be noted that event familiarity or recency judgments can be handled by the associative acquisition function, by interpreting event familiarity and recency as determined by the strength of association between the event representative and the familiarity representative. If one assumes that the familiarity representative is always activated simultaneously with the activation of each event representative, this is a special case of simultaneous conditioning.

Provision is made in the associative acquisition function for acquisition to begin only after a lag which is assumed to be a function of the memory modality. The principal purpose of this is to account for the apparent finding that moderately rapid presentation of verbal events (0.2 to 1 second per auditory event, 0.5 to 1 second per visual event) leads primarily, though perhaps not entirely, to acquisition of associations in the phonetic attribute modality, not in the concept modality. The evidence for this is that errors and interference similarity effects tend to be primarily determined by phonetic attribute similarity, rather than by conceptual (semantic) similarity (Baddeley, 1966a, 1968; Cole, Haber, and Sales, 1968; Conrad, 1963, 1964; Conrad and Hull, 1964; Conrad, Freeman, and Hull, 1965; Hintzman, 1967; Laughery and Pinkus, 1966; Posner and Konick, 1966; Wickelgren, 1965a, b, c, d, 1966a, b). There is also some auditory and visual nonverbal attribute memory (Murray, 1968; Sperling, 1960; Wickelgren, 1969a).

However, given sufficient time to study each event, conceptual coding can be demonstrated (Anisfeld and Knapp, 1968; Baddeley and Dale, 1966; Kimble, 1968; McGeoch and McDonald, 1931; Underwood, 1965; Underwood and Freund, 1968; Underwood and Goad, 1951). The time required for conceptual coding of an event probably varies somewhat with the event and the precise determination of average lag and rate parameters requires careful control of what the subject is thinking of at each moment. However, under steady-state conditions where the subject must conceptually code each of a long series of events, the lag before any conceptual coding takes place is probably on the order of 1 second, and the time constant of the subsequent exponential approach to a limit in conceptual coding is probably also on the order of 1 second.

The nonassociative acquisition function is simply a degenerate version of the associative acquisition function where the duration of only one event needs to be considered. It should be noted that the nonassociative acquisition function has the same form as the associative acquisition function for single-event familiarity or recency judgments, making the reasonable assumption that the similarity of the familiarity representative and any event representative is zero.

The form of the consolidation function for all traces in all modalities is assumed to be some power-function approach to a limit that starts only after a lag, which is assumed to be close to zero for VSTM and STM, on the order of seconds or tens of seconds for ITM, and on the order of hours for LTM. Complete consolidation of VSTM and STM is assumed to occur essentially immediately. Complete consolidation of ITM is assumed to require seconds or tens of seconds, and complete consolidation of LTM is assumed to require hours or days.

Decay of all traces in all modalities is assumed to be exponential, though the rate of trace decay may vary with the modality and the conditions. The range for visual VSTM time constants appears to be 0 to several seconds. The range for STM time constants appears to be 1 to 10 seconds. The range for ITM time constants appears to be 2 minutes to several hours. The range for LTM time constants appears to be days to years. One point must be made clear. I use the word decay to refer to the degradation of a memory trace in storage for whatever reason. Thus, decay, as I use the term, does not exclude storage interference effects (such as unlearning), though it does exclude all retrieval interference (competition) effects. One of the primary reasons for the variability in the decay rate for any given type of trace is undoubtedly that the storage interference effects vary with the modality of the trace and the nature of the experimental conditions. However, recognition of the variability of the decay rate with the storage modality and the conditions should not overshadow the importance of the simple assumption that under constant conditions the rate of decay of a trace does not vary with its age. This is a contradiction of Jost's (1897) Second Law that the rate of decay of a trace decreases with its age. The crux of the decision between Jost's Second Law and the assumption that trace decay rate is invariant with trace age is whether trace decay curves look more like single trace decay with a continuously decelerating rate or the sum of several exponentially decaying traces.

AXIOMS OF MULTITRACE THEORY

Coding

A1 (Attribute Coding) Every event has an innate internal representation such that every pair of events i and j has similarity η_{ij} determined by their sensory and motor attributes, where $0 \leq \eta_{ij} \leq 1$.

A2 (Concept Coding) For some events, the attribute representation of event i is associated to a learned conceptual representation with strength s_i, under a given interpretation of the event. Events may have several learned conceptual interpretations. The conceptual representation is such that every pair of events i and j under particular interpretations has semantic similarity δ_{ij}, where $0 \leqslant \delta_{ij} \leqslant 1$.

A3 (Types of Associations) The principal types of associations within or between modalities are: (a) interevent associations (direct forward and backward associations between the representatives of successively presented events), (b) intraevent associations (associations among the representatives of the simultaneously presented components of an event), (c) concept associations (associations between the representatives of the simultaneously presented components of an event and a concept representative for the event), and (d) structured associations (associations between the representative of an event and cognitive-structure representatives such as the familiarity representative, serial position concepts like beginning, middle, and end, subgroup labels, syntactic structures, visual images, mediators, mnemonics, rules, etc).

A4 (Modalities of Associative Memory) Every ordered pair of sensory, motor, or cognitive modalities is a memory modality. The traces involving representatives in different memory modalities may have different dynamics.

A5. (Traces in a Modality) An association in a memory modality is the sum of as many as three component traces with different dynamics: short-term memory (STM) with a time constant varying from 1 to 10 seconds, intermediate-term memory (ITM) with a time constant varying from 2 minutes to several hours, and long-term memory (LTM) with a time constant varying from days to years.

A6 (Non-Associative Very-Short-Term Memory, VSTM) Presentation of an event activates the internal representative of the event and the VSTM strength (of activation) persists for a very brief time (generally less than 1 or 2 seconds) after presentation.

Retrieval

A7 (Elementary Retrieval Processes) The principal types of elementary retrieval processes are: (a) Recall, which is a comparative judgment (1) on the STM, ITM, and LTM strengths of the associations from the representative of the cue event to the representatives of the alternative events or (2) on the VSTM activation strength of the events, and (b) Recognition, which can be either a comparative or an absolute judgment of (1) direct forward and backward event associations, (2) intraevent associations, (3) associations from event representatives to the familiarity representative,

(4) indirect event associations via a mediating structure such as serial position, or (5) any weighted additive combination of the above.

A8 (Complex Retrieval Processes) Complex retrieval processes are composed of sequences of elementary retrieval processes with final decisions being based on logical combinations of the elementary absolute and comparate judgments.

A9 (Criterion Rule for Absolute Judgment) One-dimensional n-alternative absolute judgments are based on a 1 to 1, order-preserving, mapping of an n-way, interval partitioning of the judged dimension onto the n responses. The criteria c_i ($i = 1, \ldots, n-1$) for the interval partitioning are normally distributed random variables, $c_i \sim N[\bar{c}_i, \sigma_c]$.

A10 (Maximum Rule for Comparative Judgment) One-dimensional, n-alternative comparative judgment is an n to 1 mapping from the n values on the judged dimension and their uniquely associated responses to the response associated with the maximum of these n values.

Dynamics

A11 (Phases) Memory traces have four phases: acquisition, consolidation, decay, and retrieval, in that temporal order. Activation initiates acquisition in VSTM. Contiguous activation initiates acquisition in associative STM, ITM, and LTM. Acquisition is the addition of an increment A to potential strength. Acquisition of potential strength for the same association on each trial of a sequence of n trials is A_q, $q = 1, \ldots, n$. In the consolidation phase, this potential increment in strength is converted into retrievable strength according to the function $C(t_q)$, where t_q is the delay since onset or offset of the acquisition period. In the decay phase, the activation strength decays according to the function $D(t)$, where t is the delay since the last acquisition trial (n). Retrieval is nondestructive and also can serve to produce further acquisition.

A12 (Multiplicative Combination of Phases) The memory trace available for retrieval is $M = [\sum_{q=1}^{n} A_q C(t_q)]D(t) + X$, where X is a normally distributed random variable, representing all the noise in the memory process, $X \sim N[0, \sigma]$.

A13 (Associative Acquisition) Activating two internal representatives i and j in simultaneous or immediately successive contiguity strengthens the association between the internal representatives k and l by the increment $A_q = (1 - e^{-\psi[T_i-\mu]})(1 - e^{-\psi[T_j-\mu]})(\sigma_{ik}\,\sigma_{jl}\,\alpha_f + \sigma_{il}\,\sigma_{jk}\,\alpha_b)$, where T_i is the presentation time for event i and where ψ (the rate of acquisition), μ (the lag after onset of an event before acquisition begins), σ_{xy} (the similarity of x and y in some modality), α_f (the forward association parameter), and

α_b (the backward association parameter) depend on the modality, the trace, and the state of the trace prior to presentation of events i and j on trial q. In particular, μ for the concept modality exceeds μ for any attribute modality. $\alpha_f > \alpha_b$ in successive contiguity, $\alpha_f = \alpha_b$ in simultaneous contiguity, $\alpha_{xx} = 1, \sigma_{xx} \geqslant \sigma_{xy} > 0$, and $[z] = \max(z, 0)$.

A14 (Nonassociative Acquisition) Activating an internal representative i increments the VSTM trace for item k by

$$A_k = (1 - e^{-\psi[T_i - \mu]})\, \sigma_{ik}\, \alpha$$

A15 (Concept Acquisition) Activating a set of internal representatives $\{i\}$ in simultaneous contiguity increments the strength, s_i, of association of each to an input site of the concept representative for which this set is a cue.

A16 (Delayed Consolidation)

$$C(t_q) = \begin{cases} 0 & \text{for } t_q < \tau \\ [(t_q - \tau)/(\epsilon - \tau)]^{\varphi} & \text{for } \tau \leqslant t_q \leqslant \epsilon \\ 1 & \text{for } t_q > \epsilon, \end{cases}$$

where τ, ϵ, and φ depend on the trace and the conditions. For present purposes, we can assume $\varphi = 1$ so that $C(t_q)$ is a ramp function from $t_q = \tau$ to $t_q = \epsilon$.

A17 (Exponential Decay)

$$D(t) = \begin{cases} 1 & \text{for } t < \epsilon \\ e^{-\beta(t - \tau)} & \text{for } t \geqslant \epsilon, \end{cases}$$

where β depends on the trace and the conditions.

CODING

Associative Versus Nonassociative Memory

Verbal ITM and LTM. Since the same types of arguments on this question apply to both ITM and LTM and existing data gives no reason to doubt that both ITM and LTM are associative, the two types of traces will be lumped together in the present discussion. The arguments favoring an associative theory of verbal ITM and LTM are as follows.

First, there are at least hundreds of thousands of event pairs stored with reasonable strength in LTM, and at least a thousand event pairs can be stored with reasonable strength in ITM (Wallace, Turner, and Perkins, 1957). A memory which was nonassociative in both storage and retrieval would, on the average, have to search half of all the locations in the storage system looking for the cue word in order to come up with the correct response word, stored in the adjacent location. A reasonable neurophysiological estimate of the time required to "search" a location might be on the order of 10 to 100 milliseconds, since a single synaptic delay is on the order of 1 millisecond. This yields response times for LTM that are completely absurd, and to get reasonable response times even for ITM would require some rather remarkable new discoveries in neurophysiology. Of course, the present argument only requires an associative (content addressable) retrieval scheme. Storage could still be nonassociative. However, if a system has the capacity for content-addressability in retrieval, it seems silly not to use it in storage and instead to faithfully record each occurrence of an event in a separate location. Such nonassociative storage puts great strain on the retrieval system to achieve integration of temporally distributed information concerning identical or similar events or concepts, which is achieved automatically by an associative storage system.

Second, the principal advantage of a nonassociative memory is that it could realize extremely precise temporal resolution of events. However, this appears to be an ability humans do not have, except for very brief delays on the order of seconds (Brelsford, Freund, and Rundus, 1967; Henrichs and Buschke, 1968; Peterson, 1967; Yntema and Trask, 1963). Recency judgments for delays of seconds appear to be best explained by STM associations to a familiarity representative. Recency judgments for delays of minutes or hours may make some use of ITM associations to a familiarity representative. But even for delays of minutes or hours and certainly for delays of days, weeks, and years, recency judgments are mediated by ITM and LTM associations directly to certain timing events (like the time of day or the month or season of the year) or indirectly by way of association to certain larger events (such as one's stay in a certain location) which are associated to timing events. Handling recency judgments of events by means of associations which can be quite separate from other associations involving these events accounts for the fact that recency judgments can be quite poor when other associations to the events are extremely strong. With a nonassociative memory, it is difficult to see how one could be virtually incapable of remembering when an event occurred, but could recall an event associated with the cue event. It could be explained by the retrieval system, but it is difficult to see why one would have a nonassociative storage system and then throw away by an inadequate retrieval system the precise timing information that such a system can give you. It is particularly odd when you consider that the retrieval system for associative information has to be so complex because it must operate on a nonassociative storage system.

Third, retroactive (Bugelski and Cadwallader, 1956; Gibson, 1941; McGeoch and McDonald, 1931; Osgood, 1949, 1953) and proactive (Blankenship and Whitely, 1941; Melton and Von Lackum, 1941) interference in recall is a function of the similarity of the two lists. At a minimum, this requires an associative retrieval system. However, recently, there is a growing body of evidence to indicate that similarity-dependent retroactive interference is not solely a retrieval phenomenon in verbal ITM and LTM. Numerous studies have demonstrated that interpolated A-C learning causes greater storage interference (unlearning or inhibition of consolidation) for previously learned A-B association than does interpolated C-D learning (Houston, 1967; McGovern, 1964; Postman, 1965). Similarity-dependent storage interference is inconceivable with a nonassociative storage system. In a nonassociative memory, there is just no reason why storing A-C in one pair of adjacent locations should cause greater decay of A-B in another pair of locations than should the storage of C-D in the prior pair of locations. In an associative memory, it is perfectly reasonable to suppose that the connection capacity of any single internal representative is limited, so that increasing an A-C association will tend to weaken an A-B association.

Fourth, some position must be taken in the nonassociative theory as to whether repetition of a sequence of events can strengthen the traces for the events in the same sequence of locations or whether new occurrences of events are always stored in new locations. If it were possible to increase the probability of correct recall as much by strengthening the trace in a particular sequence of locations as by replication and if the subject were deliberately practicing a particular sequence, one would assume that strengthening would be the storage method of choice. Improving memory by replication causes problems in retrieval, particularly in the time required to come up with the correct response and particularly in learning by the part method. If one is learning the sequence ABCDEF in two parts, ABC and DEF, replication leads to a memory structure like the following ABCABCABCABCDEFDEFDEFDEF, whereas strengthening leads to a memory structure that parallels the sequence structure, namely, ABCDEF. Nevertheless, the results of experiments on learning by the part method make it clear that, if humans had a nonassociative memory, they would have to be assumed to be using replication not strengthening learning. This is because substantial practice is required to connect the parts after each part has been learned separately and subjects make CA type errors in attempting to connect the parts, as if the end of a part had been associated to the beginning of the part in learning the part (see McGeoch and Irion, 1952, for a review of whole vs. part learning). However, the nonassociative theory with replication has its own problems. How does the subject ever learn to connect the parts? How could he ever stop making CA errors, if the previous part learning is to be of benefit to him, as it clearly is? The simple linear structure of nonassociative memory is inadequate to account for the more complex topological (connection) structure of human memory.

An associative memory is completely consistent with all the facts of whole vs. part learning. It explains why the whole method is often superior to the part method in efficiency of learning, despite the fact that subjects strongly prefer the part method and receive more immediate reward in the form of correct performance with the part method. The advantage of the whole method is that it secures the greatest efficiency ratio of direct event-to-event associations rehearsed to events rehearsed, namely $n-1/n$, where n is the length of the list. The part method, particularly in the combination of the parts, has a very much lower efficiency ratio. Furthermore, an associative memory predicts that combining the parts should be particularly inefficient, and also predicts that there should be frequent errors involving the end of a part being associated to its beginning, if rehearsal of a part involved close temporal contiguity of the end and the beginning of the part.

Verbal STM. Since the capacity of verbal STM is small in comparison to ITM and LTM and since work on STM began in earnest about the time that computers were being developed, many people interested in STM assumed explicitly or implicitly that STM was nonassociative (buffer storage). The earliest focus of my research on STM was to show that this was false and that STM was associative, just as is ITM and LTM.

First, RI and PI in short-term recall are similarity dependent, when the phonetic similarity of the interfering list to the list to be remembered are manipulated (Wickelgren, 1965a, 1966b).

Second, in at least one case intralist phonemic similarity facilitated item memory while severely depressing order memory, exactly as predicted by an associative theory of STM (Wickelgren, 1965e). Taking as an example the phonemically similar list of consonant-vowel digrams, *na fa ta*, versus the phonemically different list, *na fo ti*, with only the consonants having to be recalled, the argument is essentially as follows: in an associative memory, all of the consonant representatives will be associated to the single vowel representative, *a*, in the similar list. Since *a* is very frequently activated, it will be certain to be recalled producing a very good item-recall for all of the consonants, but it will give no information about the order of the consonants. Considering only the direct forward associations between the components of adjacent CV digrams, order information is carried only by the single association between the adjacent consonants in the similar lists, but is carried by all four associations from and to adjacent consonants and vowels in the different lists. Thus, an associative theory of STM predicts that order memory might well be better for different lists, while item memory is better for similar lists. This was what was found, and it is difficult to see how a nonassociative theory could account for such a dissociation between item and order memory.

Third, an associative theory of STM predicts that the topological structure of the associations between item representatives will be different for lists with repeated items than for lists with no repeated items. A nonassociative theory of STM predicts no such differences, though of course one can construct *ad hoc* hypotheses to graft onto the nonassociative memory structure to account for some of the same repeated item phenomena predicted by an associative memory structure. A rather large number of differences in STM for lists with and without repeated items have been found (Wickelgren, 1965f, 1966c), perhaps the simplest of which to describe is the associative intrusion phenomenon. Lists which have the form...*AB*...*AC*...show frequencies of substituting *B* for *C* and *C* for *B* that significantly exceed the frequencies of substituting items in these positions when they are not preceded by a repeated item. The reason for such associative intrusions is obvious with an associative theory of memory, the representative of *A* is associated to both *B* and *C* and the differentiating serial position cues do not completely swamp the effect of the ambiguous prior item cue. In a nonassociative theory, there is no logical reason to expect such associative intrusions, at least not in STM for relatively short lists.

Fourth, in STM experiments on the effects of rehearsal grouping, subjects give evidence of the ability to cross-classify items by beginning, middle, or end group and by beginning, middle, and end position within the group (Wickelgren, 1964, 1967a). Several phenomena in rehearsal grouping experiments support this interpretation, the most definitive of which is that errors tend to be within the group or in the same position of a different group. In an associative memory, an item can have associations to a group label and a position within a group label, but it is very difficult to see how this could be accomplished by a one-dimensional nonassociative memory.

Visual VSTM. Wickelgren and Whitman (1970) have obtained experimental evidence which indicates that visual VSTM (e.g., Sperling, 1960) is nonassociative. Similar experiments to those used to show that verbal STM is associative can probably be done to show that visual VSTM is nonassociative, namely, storage of a visual pattern in each of some array of locations on the retina, with no associations between the item representatives in each location.

Attribute Versus Concept Representation

Verbal ITM and LTM. Since the same phonetic and visual attributes characterize an enormous number of different verbal events, it is not likely that interevent associations in ITM and LTM would be primarily based on ITM and LTM associations between the phonetic or visual attribute representatives of the events. The retrieval and storage interference effects of subsequent events and the retrieval, and possibly storage, interference effects of prior events would simply be too great at the phonetic attribute level of coding. Thus, it is not surprising that errors and interference effects in verbal ITM (and presumably also LTM) for words are primarily based on conceptual similarity as shown by studies in serial learning (McGeoch and McDonald, 1931; Underwood and Goad, 1951), paired associate learning (Baddeley, 1966b; Baddeley and Dale, 1966), and recognition learning (Anisfeld and Knapp, 1968; Kimble, 1968; Underwood, 1965; Underwood and Freund, 1968) with delays on the order of minutes or tens of minutes.

The learning of verbal lists (nonsense syllables or words) generally does show interference effects due to phonemic or graphemic (letter) attribute similarity, often called "formal" similarity, (e.g., Baddeley, 1966b; Cohen and Musgrave, 1966; Dallett, 1966; Feldman and Underwood, 1957; Horowitz, 1962; Runquist, 1966; Underwood, 1953). However, the influence of attribute similarity may be largely due to a substantial STM component in precriterion performance on standard verbal learning tasks. When 16 second delays are introduced between learning trials and a rehearsal-preventing task is used to fill these delays, the effect of attribute similarity on verbal learning disappears (Baddeley, 1966b).

This does not imply that there is no ITM or LTM involving attribute representatives. There must be LTM for the associations from attribute representatives that specify concept representatives. Furthermore, in the present theory, the distinction between attribute representation and concept representation of an event must be considered to be relative to the event, with the attributes of an event being its immediate constituents. Thus, the immediate constituents of a word or other multiphonemic item might be an ordered set of its syllables, phonemes, or graphemes (letters). Alternatively, the immediate constituent of a word might be an unordered set of its context-sensitive allophones (Wickelgren, 1969c). According to the context-sensitive allophonic theory of the coding of speech at the phonetic level, there is an internal representative for every ordered triple of phonemes that appears in any word in one's language. Thus, *stop* is coded as the unordered set, $/_{\#}s_t, {}_st_o, {}_to_p, {}_op_{\#}/$. It is a remarkable feature of context-sensitive allophonic coding of words that the ordered set can always (to my knowledge) be recovered from the unordered set. This recovery of the ordered set from the unordered set is accomplished largely by the use of long-term associations between the allophone representatives. The context-sensitive coding theory solves the problem of serial order in the articulation of speech at the word level and also solves the two most celebrated problems in the recognition of speech at the word level, namely, segmentation (partitioning the acoustic signal for a word into segments) and the immediate left and right context-conditioned variation in the acoustic features of a phoneme. Thus, there is considerable support for a theory which requires the assumption of LTM associations between the immediate constituents (attributes) of a word.

It is probably the case that the ITM and LTM traces involving attribute representatives play a small role in most adult memory tasks because of the interference factors. Nevertheless, according to multitrace theory, it ought to be possible to find situations in which they play a large enough role to demonstrate their existence.

Verbal STM. Errors and interference effects in verbal STM for rapidly presented multiphonemic items such as words, nonsense syllables, letters, and digits in many studies appear to be primarily dependent upon phonemic similarity (Conrad, 1963, 1964; Conrad, Freeman, and Hull, 1965; Conrad and Hull, 1964; Posner and Konick, 1966; Wickelgren, 1965a, b, c, e, 1966b). Errors in verbal STM for rapidly presented single vowel or consonant phonemes in many studies appear to be primarily dependent upon distinctive feature similarity (Cole, Haber, and Sales, 1968; Hintzman, 1967; Wickelgren, 1965d, 1966a).

This does not mean that there is no STM in the concept system. Generally small, but statistically significant, effects of conceptual similarity have been found in several types of situations usually thought to be primarily measuring STM (Corman and Wickens, 1968; Dale and Gregory, 1966; Henley, Noyes, and Deese, 1968; Turvey, 1967; Wickens and Eckler, 1968; Wickens and Simpson, 1968). Even those STM studies which failed to find statistically significant effects of conceptual similarity have found effects in the same direction (Baddeley and Dale, 1966; Dale, 1967). All of these demonstrations of small effects of conceptual similarity are open to the charge that they are assessing the conceptual coding, not of STM, but of a small ITM component which may be present in the situations studied. The study by Turvey (1967) supports the interpretation of the conceptual similarity effects as being due to ITM, since Turvey found effects of conceptual similarity with delays of 24 seconds but not with delays of 12 seconds. Henley, Noyes, and Deese (1968) also found greater effects of conceptual similarity with delays of 16 seconds than with delays of 0, 4, and 8 seconds. However, until one has the ability to assess conceptual similarity effects in memory situations which are known, by some independent criterion, to be mediated entirely by STM, or until we have a well established quantitative analysis of trace components in some situation, it will not be possible to definitely conclude whether there is STM facilitation of associations in the concept system.

Nonverbal Auditory STM. Errors in STM for the pitch of pure tones follow a generalization gradient which has one maximum at the pitch of the tone to be remembered (standard tone), at least within a single octave (Wickelgren, 1969a). Although this has not been established quantitatively to my knowledge, the accuracy of the ability to recognize a difference between the standard and comparison tones with short delays between them is probably vastly greater than the accuracy of one's ability to sing a tone of a given pitch. Furthermore, the ability to recognize pitch differences extends beyond the range of the human voice. Therefore, it is likely that the memory is in the auditory system, rather than being in a vocal motor system. A variety of unpublished attempts to attribute part or all of this pitch memory to verbal memory for absolute or comparative judgments of pitch have all failed to provide any support for the presence of verbal memory components when the number of alternative standard tones is 10 or more. Thus, I think it is quite certain that there is nonverbal attribute STM for pitch and undoubtedly for many other psychophysical attributes in sensory, and probably also motor, modalities.

Visual VSTM. Visual VSTM presumably is coded in some visual attribute system, and indeed it is very susceptible to visual interference, but not very susceptible to verbal or nonverbal auditory interference (Sperling, 1963). An unsystematic error analysis in one situation which probably is mediated in part by visual VSTM and in part by verbal STM suggested the presence of both visual and phonetic coding (Sperling, 1960). Presumably the visually similar errors were due to the visual VSTM component and the phonetically similar errors were due to the verbal STM component, but it would be nice to have this shown in a precise quantitative manner.

Rote Versus Structured Learning

Rote learning refers to the strengthening of a direct association between the internal representatives of two actually occurring events or components of a single event, without use of any associations to or from other (cognitive structure) representatives which are not representatives of the occurring events or event components. Structured learning refers to every type of learning that involves the strengthening of associations to, or from, internal representatives which are not the representatives of currently occurring events or event components. Such representatives are called *cognitive structure* representatives. As such, structured learning includes the direct association of an event representative to a cognitive structure representative and the indirect association between two event representatives (*A* and *B*) mediated by a chain of associations involving cognitive structure representatives. Such a chain presumably starts with associating event *A* to one or more cognitive structure representatives which may be associated to other cognitive structure representatives and so on until some terminal set of cognitive structure representatives is elicited which is associated to event *B*.

If we summarize the sequence of sets of cognitive structure representatives by M (mediator), getting A-M-B, we have the conventional diagram for mediated association. In those cases where M is a single internal representative unadorned by the presence of cognitive operations (logical, syntactically structured, etc., thought processes, whatever they are), there seems to be little reason to complain with this formulation. Where M hides a multiplicity of cognitive operations, the formulation is open to the charge that it does just that, namely, hides a multiplicity of cognitive operations without any attempt to understand them. However, that is just the point, to study learning and memory, it would be extremely desirable to be able to give cognition its due in relation to learning and memory, without having to explain everything in cognition as well.

The current formulation of multitrace theory adopts the mediated-association approach to the influence of cognition upon memory. Following Jarrett and Scheibe (1962), multitrace theory takes the further step in the quantitative formulation of mediation theory of assuming that one can represent a mediational chain of associations, A-M-B, by a single strength. This mediated strength of association is just like the strength of a direct association, A-B, and can be added to any direct association to determine the total strength of association between A and B. Whether this simple approach to the effect of cognition on memory will be successful remains to be seen, but the determination of its success depends upon how adequately it accounts for learning and memory phenomena, not upon how adequately it accounts for thinking and language phenomena.

Although I am not, at present, in a position to make a quantitative evaluation of the mediational approach to structured learning, the need for some incorporation of structured learning is clear and will be described in the following sections. From an introspective viewpoint, most of adult learning involves association of the representatives of occurring events to a cognitive structure, and very little is rote. Multitrace theory says that all of the basic principles of rote learning apply to structured learning. Nevertheless, it seems to me to be important for an adequate theory of learning and memory to make the distinction between rote and structured learning and to describe the important classes of structured learning, because this is important in and of itself, not because it interacts with any other part of the multitrace theory of memory.

Verbal Paired Associates Learning. Subjects very frequently report the use of verbal mediators in paired associate (PA) learning, and there is now rather substantial evidence that PA learning is facilitated by good verbal mediators whether supplied by the subject or the experimenter (Dallett, 1964; Davidson, 1964; Jarrett and Scheibe, 1962; Jensen and Rohwer, 1963a,b; Kiess and Montague, 1965; Wood and Bolt, 1968). Good mediators also facilitate retention (Adams and McIntyre, 1967; Montague, Adams, and Kiess, 1966; Runquist and Farley, 1964; Reed, 1918) and reduce retroactive interference of A-B pairs due to subsequent A-C learning (Adams and Montague, 1967).

Good mediators can be defined theoretically as those mediating concepts which have strong previously established associations from the stimulus event and to the response event, such as the example given by Dallett (1964) of *bacon* as a good mediator between the stimulus trigram *bac* and the word response *eggs*. Poor mediators have low levels of previously established associations from the stimulus event and to the response event. Empirical measures of the goodness of mediators can be derived from free association data, subjects' or experimenters' ratings, or even from whether the mediator is or is not recalled on some later trial or retention test. Poor mediators can inhibit PA learning (Dallett, 1964), and generally fail to facilitate retention or possibly inhibit it (Adams and McIntyre, 1967; Montague, Adams, and Kiess, 1966). However, as one might expect, most mediators chosen by subjects are good mediators and so the net effect of natural language mediation by subjects is facilitatory in both learning and retention.

In addition to verbal mediation, it appears that combining the stimulus and response events into a single visual image facilitates PA learning and memory (Bugelski, 1968; Epstein, Rock, and Zuckerman, 1960; Paivio, 1969). The mediator in this case is a visual image that includes both the stimulus and response events in some sense, and this seems somewhat different from a natural language mediator, which does not necessarily include the stimulus and response event in any sense. However, multitrace theory treats both verbal and visual mediation in the same manner, namely, by means of a mediated strength of association between the stimulus and response events.

Serial Learning. To the extent that serial learning (SL) in STM, ITM, or LTM is achieved by facilitation of associations between adjacent item-representatives (Postman and Stark, 1967; Shuell and Keppel, 1967; Wickelgren, 1965f, 1966c), verbal or visual mediation of the type just described ought to be facilitatory in SL, and indeed there is some evidence for this (Houston, 1964). However, the interesting new type of structured learning that appears with SL is mediation by the use of serial position concepts.

There is evidence for the use of serial position concepts in STM studies of serial learning. Melton and Von Lackum (1941), Conrad (1959, 1960) found that there was a significant tendency for intrusion errors in immediate recall of a list to come from the same position in the previous list. Wickelgren (1964, 1967a) in a study of rehearsal grouping in groups of different sizes found a significant tendency for intrusion errors to be other items in the same group or to be items in the same relative position in other groups. It is difficult to see how the latter within-position errors could have occurred above chance, except if subjects were using a two level cross-classification of serial position concepts for groups and for positions within a group. Furthermore, the error patterns and the optimum rehearsal group size at around 3 or 4 suggested that subjects in this situation can use on the order of 3 distinct serial position concepts, beginning, middle, and end. However, the ordering power of these 3 serial position concepts can be greatly extended by the human capacity for two-level use of these concepts through grouping.

Intermediate term memory studies of SL also indicate the importance of serial-position-to-item associations. Evidence for this comes from a variety of sources. First, positive transfer between a serial list and a PA list derived from adjacent pairs of items in the serial list is far from 100 percent (Jensen and Rohwer, 1965; Postman and Stark, 1967; Shuell and Keppel, 1967; Young, 1959, 1961, 1962; Young and Casey, 1964; Young and Clark, 1964), though it can sometimes be substantial on early transfer trials. This indicates that some other cues besides the direct forward item-to-item associations play a role in SL, and while this does not prove that the other cues are serial position cues, they are the most likely candidates. Second, when the serial list is learned by varying the starting position on each learning trial, the transfer from SL to a compatible PA list is significantly enhanced (Shuell and Keppel, 1967). Since varying the starting position in SL must primarily reduce the reliance upon serial position cues, this effect indicates that serial position cues are important in ordinary SL. Third, SL with variable starting position is more difficult than ordinary SL (Bowman and Thurlow, 1963; Ebenholtz, 1963a; Keppel, 1964; Saufley, 1967; Shuell and Keppel, 1967; Winnick and Dornbush, 1963). Fourth, there is positive transfer from SL to a PA list which pairs the items from the serial list with numbers referring to their previous serial positions (Jensen and Rohwer, 1963). Fifth, there is positive transfer from one SL task to another SL task for items that are common to the two lists and maintain the same serial position (Ebenholtz, 1963a). Sixth, presenting a list of items in a spatial order without temporal ordering can be just as efficient for learning a serial list as presenting the list in temporal order (Slamecka, 1967), though under other conditions, conventional SL has been found to be somewhat faster (Ebenholtz, 1963b). Seventh, in the Ebenholtz (1963b) study there was heavy positive transfer between conventional SL and the spatial position learning.

RETRIEVAL

Event

Recognition Memory. Event recognition memory is measured by presenting a test event, such as a verbal item (letter, digit, nonsense syllable, word, etc.), and asking the subject to say whether or not that event has occurred in some period of time. The only reason for using the term *event* rather than *item* is to include motor and nonverbal sensory events which are not so naturally referred to as *items*.

Most studies of event recognition memory have been concerned with judgments of the occurrence or nonoccurrence of the test event in a period of time that could be considered to end with the presentation of the test event. Under such circumstances it is easy to think of event recognition memory judgments as being based on the familiarity of the test event, where familiarity is based on the strength of the STM, ITM, and/or LTM association from the representative of the test event to the familiarity representative. Recently presented events have STM and/or ITM facilitation of these associations on the hypothesis that the familiarity representative is always activated in contiguity with the representative of any occurring event.

However, when a subject is asked to recognize whether a test word was presented previously in a 1-hour session that occurred a week ago, the ITM or LTM strength of the association from the test word to the familiarity representative is not likely to be the primary basis for making the judgment. Such associative strengths would not distinguish between words presented in the critical period and words presented before or after in the subject's daily life. Thus, it is undoubtedly necessary to assume that, if it is familiarity that is being judged in this type of event recognition, it is the familiarity of the compound event consisting of the test event and the situation in which it occurred. This might be called associative recognition memory because it seems reasonable to assume that the strength of association of the representative of a compound event to the familiarity representative must be determined largely or completely by the strength of the associations between the representatives of the components of the compound event.

In either case, multitrace theory can treat the recognition of an event as being based on a single strength characterizing the familiarity of that event. Judgments in event recognition memory are assumed to be made by the criterion rule, namely, respond yes, if and only if the strength (familiarity) of the test event exceeds a criterion, and no otherwise. Further partitioning of the strength scale can be achieved by having the subject give confidence ratings in addition to the yes-no response. A yes-no decision combined with n levels of confidence yields $2n$ possible responses, which are assumed to be determined by the placement of $2n-1$ criteria that partition the strength scale into $2n$ not necessarily equal intervals. This makes event recognition memory an absolute judgment on trace strength.

Under constant experimental conditions, this strength may vary due to uncontrolled noise in the memory process, acquisition, consolidation, decay, and retrieval. In multitrace theory, the presence of such noise is handled by the addition of a random variable to the retrieved strength with zero mean and a variance that is often constant over the different conditions in an experiment, but could be assumed to vary with the conditions, if that were necessary.

In addition to noise in the memory process, there is undoubtedly a very large source of noise in a subject's criterion placements. Under the assumption that memory noise and criterion noise are normally distributed, the two can be combined into a single normally distributed noise. The combined noise can then be considered to add to a real-valued memory strength, converting it into a random variable whose mean is the same as the real-valued strength, and whose variance is the variance of the noise distribution. This is exactly the same trick used by Hull (1943) known by the name of *behavioral oscillation.* The advantage of this way of handling noise is that it permits one to have a real-variable memory theory almost everywhere, converting real variables to random variables only at the end of all computation regarding the passage of memory traces through different phases. To the extent that retrieval and criterion noise are large in relation to the noise in acquisition, consolidation, and decay, there can be no better way of handling the uncontrolled stochastic aspects of memory. With appropriate experimental control of coding and rehearsal, it seems reasonably likely that criterion noise overshadows all other noise, so this is a particularly attractive way of handling noise. However, even if the noise in acquisition, consolidation, or decay were substantial, the present approach could still be the simplest way of handling the stochastic aspect of memory, though, of course, it is somewhat *ad hoc*.

Like any strength theory, multitrace theory has the desirable property of allowing one to use the judgment rule of the theory, in combination with the noise assumptions, to convert response probabilities in event recognition memory into interval scale measures of the total memory strength of the event. The fact that measurement is on an interval scale means that one only measures the difference in the strength of two types of test events, for example, an event presented i seconds ago and an event that was not presented at all in the period in question. Furthermore, the unit of measurement is the standard deviation of the noise in the strength of one of these events. If the noise affecting the measured strength of each event is the same for all types of events in an experiment, then this is no problem. If there is some reason to suspect that the noise may vary for different events, then it may be necessary to take special care to measure all strength differences with the same unit of measurement. The latter problem is discussed at length in Wickelgren (1968a).

Operating Characteristics. In most cases, it suffices to estimate strength differences and the ratios of the standard deviations of the noise in two different strength distributions by means of a special plot called an operating characteristic (OC). The present section explains how to plot and interpret OCs for an event recognition memory experiment using confidence ratings.

Imagine that you have presented a list of items followed by a test item which is sometimes an item from the previous list (correct item) with i intervening items and sometimes an item not from the previous list (incorrect item). Call these conditions i and $*$, respectively. On each trial a subject answers yes or no and states his confidence on a scale from 1 (least) to 4 (most). We consider the subject's yes-no and rating responses to be a choice of one of 8 responses ordered on a single dimension of sureness that the test item was in the list: $Y4$, $Y3$, $Y2$, $Y1$, $N1$, $N2$, $N3$, $N4$.

Ideally, we have enough data for single subjects to do all analyses separately for each subject. This avoids any possibility of being misled as to the form of various functions due to averaging subjects with very different values of various parameters. More important, it gives us some idea of the range of individual variation and whether this variation can be accounted for by different values of parameters.

From our experiment we obtain the frequency of each of the 8 responses in each condition. OCs are generated from the data for two conditions, say for i and *. The first step in generating an OC is to convert the frequency of each response in each condition into an empirical cumulative probability of a response with a certain degree of sureness or greater sureness (that the test item was in the list). This means that for condition i you divide the frequency of $Y4$ responses by the N_i for condition i, then you divide the frequency of $Y3 + Y4$ responses by N_i, then you divide the frequency of $Y2 + Y3 + Y4$ responses by N_i, ..., then you divide the frequency of $N3 + N2 + N1 + Y1 + Y2 + Y3 + Y4$ responses by N_i. Thus, you obtain 7 empirical probabilities (which must be monotonic nondecreasing) for condition i; call them p_{1i}, $p_{2i}, \ldots, p_{7i}$. In an identical manner, you obtain the 7 probabilities for condition *, $p_{1*}, p_{2*}, \ldots, p_{7*}$.

The second step is to plot on normal-normal probability paper, the 7 pairs of points: (p_{j*}, p_{ji}) for $j = 1, \ldots, 7$, where p_{j*} is the horizontal coordinate and p_{ji} is the vertical coordinate.

The third step is to fit a straight line to the points, assuming that the points do not deviate systematically from a straight line in this and other OCs for the same subject in similar conditions (say the OCs for each of the various i conditions against the * condition). The best fitting straight line is our empirical estimate of the true OC. The absence of any systematic deviation of the empirical points of the various OCs from straight lines is evidence that the noise in both i and * conditions is approximately normally distributed.

The fourth step is to determine the slope and horizontal intercept of the best-fitting straight line. The slope provides an estimate of the ratio of the noise in the * condition to the noise in the i condition, σ_*/σ_i. The horizontal intercept provides an estimate of the difference between the trace strengths of i and * test items in units of σ_*, namely, $D(i, *) = (M_i - M_*)/\sigma_*$.

Empirical Adequacy. The only direct test of the retrieval assumptions for event recognition that is currently available is the fit of the OCs to straight lines on normal-normal plots. In almost all cases, there is no systematic deviation, but there can be a lot of unsystematic variation. Appropriate goodness of fit tests are now being developed (Dorfman and Alf, 1968), so this test can be made more powerful in the future. However, this is mainly a test of the normal distribution assumption. As long as the strength distributions are unimodal, slight deviations from normality, no matter how significant statistically, are of little significance theoretically. Hence, I think that the results of such tests of the fit of OCs to straight lines on normal-normal plots are of very limited value in determining the accuracy of the retrieval assumptions of multitrace theory.

The principal evidence for the validity of the retrieval assumptions for event recognition appears to be largely indirect, namely, how elegant are the laws for the dynamics of memory traces that result when these retrieval assumptions are made. So far the laws of trace dynamics for event recognition memory appear to be rather elegant, but it is far too early to be sure of this.

Associative Recognition Memory

Although, as mentioned in the previous section, event recognition memory may sometimes or always be merely a special case of associative recognition memory, it is clearer, and perhaps safer, to distinguish them at the present time than to lump them together. Associative recognition memory refers to the subject's memory for the temporal contiguity or temporal order of events, namely, his ability to discriminate correctly and incorrectly paired events in PA recognition memory and his ability to discriminate direct forward associations from backward or remote associations in SL using a recognition test of serial order memory. In each case, multitrace theory assumes that it is the strength of association between the two test events that is judged in a recognition test, with the forward association between the test events being weighted more heavily than the backward association in the serial order case at least.

In the case of associative recognition memory, it is possible to devise a very strong test of the central property of the retrieval assumptions for recognition memory. This central property is called *independence from irrelevant associations* because it asserts that only the strength of the test association is judged in recognition memory. For example, if the test pair is A-B and subjects are to decide whether B followed A in the previous list, only the strength of the A-B association is assumed to enter into the judgment. The strengths of presented or nonpresented A-C pairs are assumed to have no effect on the A-B recognition judgment. Another way to refer to this is to say that recognition tests are free of retrieval interference.

Independence from irrelevant associations has been tested in STM studies of SL and PA by comparing recognition memory judgments for A-B pairs in cases where the lists that were presented contained both A-B and A-C pairs, and in cases where the lists that were presented contained only A-B pairs with X-C pairs used in place of the A-C pairs. No differences in recognition memory performance were obtained (Bower and Bostrom, 1968; Wickelgren, 1967b and unpublished data), confirming the central property of the multitrace retrieval assumptions for recognition memory.

Multiple-Choice Recognition

In multiple-choice recognition, subjects must choose which of n events occurred in some previous list. The events are frequently compound events, $A - X_i$, $i=1, \ldots, n$, where the subject's task is to choose the X_i which was paired with A in some previous list. Multitrace theory assumes that only the strengths of the test associations are judged in multiple-choice (independence from irrelevant associations) and that the maximum decision rule is employed, namely, choose the event with the greatest strength of association to the cue event. If these retrieval assumptions are correct, it should be possible in many cases to predict performance on a multiple-choice test with m alternatives from performance on a multiple-choice test with n alternatives. Such predictions have been reasonably successful (Kintsch, 1967; Norman and Wickelgren, 1969).

Recall

Although the mechanism of recall with even a small known set of n alternatives may be different from n-alternative multiple-choice, they may both be describable by the maximum decision rule applied to the strengths of the associations from the cue event to the n alternatives. Evidence for different mechanisms was obtained by Norman and Wickelgren (1969) who found recall with 4 alternatives to be substantially faster than 4-alternative multiple-choice recognition. No systematic differences were found in the probabilities of correct choice in 4-alternative recall and multiple-choice, but large unsystematic differences were found. No definite conclusion regarding recall from small specified populations can be drawn at the present time. Recall from large populations of events is largely an unexplored frontier in the testing of multitrace theory.

However, some plausible speculations can be made regarding recall. First, it may be a simultaneous search through the alternatives, while multiple-choice is a sequential search. This is suggested by my introspection and by the latency differences obtained by Norman and Wickelgren (1969). Of course, I am referring to an elementary recall process, not a complicated recall task where a subject is allowed a lot of time to consider and reject alternatives. The latter task may have lots of elementary recall processes arranged in a series interspersed with elementary recognition processes. Of course, both a simultaneous and a successive search through the alternatives could be described by the maximum rule, but there might be some minor or major differences, for example, greater noise in the retrieval of associative strengths with the simultaneous search, or greater opportunity for decay of associative strengths during retrieval with the successive search.

Second, the total strengths of the alternatives judged in recall and multiple-choice may have somewhat different components. For example, subjects can be guaranteed that both items in a pair were in the previous list and they are only to decide whether the second item immediately followed the first in the previous list. Formally, this makes item strengths irrelevant to the judgment, and it may make them psychologically irrelevant. In recall, the item strengths of the different alternative response items may be important in addition to the strengths of association from the cue to the response items. Furthermore, in multiple-choice recognition, the backward association may help to discriminate correct and incorrect pairs, whereas in recall it is less likely that the backward association plays any role.

Predicting Recall and Multiple-Choice from Recognition

Another way of testing all of the retrieval assumptions of multitrace theory is to attempt to predict recall and multiple-choice with different numbers of alternatives from recognition memory data. The prediction is sensitive to hordes of minor assumptions (see Norman and Wickelgren, 1969 and Wickelgren, 1968a for a discussion), but, surprisingly enough, the prediction has been reasonably successful in at least two cases (Green and Moses, 1966; Kintsch, 1968). In a third case (Norman and Wickelgren, 1969), there appeared to be some minor complications in the prediction, but there was no reason to fault the basic retrieval assumptions of multitrace theory.

DYNAMICS

Two-Phase Studies of Verbal STM

Two-phase studies of verbal STM involve presenting a list of items followed immediately by a probe item or pair of items, with the subject required to say whether the probe was in the previous list or not. When the probe is a single item, this is an example of event recognition memory. When the probe is a pair of items and the subject is guaranteed that both items were in the list, though not necessarily as a pair in the presented order, this is an example of associative recognition memory. The two independently manipulatable experimental phases are the acquisition-consolidation-decay phase on the one hand and the retrieval (decision) phase on the other hand. This does place some limitations on the power of the two-phase design in testing theories of memory, by comparison to designs with more independently manipulatable phases. The principal example of the latter is the three-phase design of Brown (1958) and Peterson and Peterson (1959) where a third type of activity is interpolated between the presentation of the list to be remembered and the test.

However, the two-phase design does have an important advantage, which is that it appears to produce the least amount of active STM trace maintenance (rehearsal) of earlier items and the least amount of ITM difference between items presented and not presented on a trial. The evidence for the absence of ITM is that the two-phase STM strength decay curves are virtually always simple exponentials, whereas, three-phase STM strength decay curves always require the assumption of two traces (STM and ITM) decaying at very different rates (Wickelgren, 1970a). Furthermore, the STM component of three-phase strength decay curves is decaying somewhat more slowly than the STM trace in two-phase studies (Wickelgren, 1969b, 1970b).

Simple exponential decay of trace strength in STM has been obtained with two-phase studies on: (a) event recognition memory for three-digit numbers (Wickelgren and Norman, 1966) and single letters (Wickelgren 1970b), (b) associative recognition memory for the serial order of digits in SL (Wickelgren, 1967e) and of letter-digit pairs in PA (Wickelgren, unpublished), and recall of single digits (Norman, 1966). All of these studies carefully instructed subjects to think of the current items only and not try to rehearse previous items. This may be of critical importance in obtaining simple exponential decay in two-phase studies of STM.

One possible explanation of the absence of ITM in the above two-phase studies is that the presentation rate was too fast and the number of items to be learned was too great to permit the consolidation of appreciable amounts of ITM. This seems unlikely to account for the apparently complete absence of ITM in the strength decay curves. More likely, the explanation is, at least in part, as follows. The ITM traces are decaying at a rate which is slower than the decay rate of the STM traces by a factor of about 10^2 (Wickelgren, 1970a). With short lists constructed from a small population of items and with a short intertrial interval, the ITM traces for all items, whether correct or incorrect for that trial would be relatively constant. Thus, they would contribute nothing to the strength discriminability of correct and incorrect test items and would leave the strength decay curve reflecting only the STM strength discriminability.

This interpretation is confirmed by an unpublished two-phase study with words from a rather large population of 1000 words where a modest ITM component appears in the strength decay curve. One might well ask why about the same amount of ITM did not appear in the Wickelgren and Norman (1966) study which used a population of a little under 700 three-digit numbers. The reason is undoubtedly that three-digit numbers are much more similar to each other in their internal representation than are words, leading to a much greater amount of ITM strength generalization for three-digit numbers than for words.

Obtaining simple exponential decay of strength in STM is a very important simplifying result for multitrace theory. It would be very messy to have trace decay rate depend on trace age. However, one further simplification does not hold, namely, STM decay rate is not invariant over the conditions in the delay interval. At present, the primary variable that is known to affect the temporal decay rate of STM is the density of items to be learned in the delay interval. The evidence for this is a study by Wickelgren (1970b) in which rate of decay for letters (with time measured in seconds) was a linear function of rate of presentation of the list, with a positive intercept. The positive intercept, obtained for all subjects in all conditions, disproves the hypothesis that decay rate is invariant over rate of presentation if time is measured by the number of intervening items. Furthermore, the extrapolated rate of decay of STM at a hypothetical rate of 0 items to be learned per second is of the same order of magnitude as the rate of decay of the STM component in three-phase studies where the material that fills the delay interval does not have to be learned (Wickelgren, 1970b). The interpretation placed on the dependence of decay rate on the rate of presentation of new items to be learned is as follows. There is an active STM trace maintenance process which is partially counteracting a passive temporal decay process. However, the STM trace maintenance process requires some of the same neural apparatus as the STM trace acquisition and consolidation processes and thus cannot maintain previous STM traces as well during a period of around .25 seconds/letter in which the STM trace for each new letter is established. During any blank time between letters, the STM trace maintenance process operates much more effectively in counteracting decay. According to this hypothesis, rehearsal is the conscious top of the unconscious trace-maintenance iceberg. The simple assumptions formulated by Wickelgren (1970b) to account for this STM trace maintenance process may very well be approximately valid only for cases where the frequency of conscious rehearsal is kept minimal by nonrehearsal instructions.

Three-Phase Studies of Verbal STM and ITM

Three-phase studies involve presenting a short list of events to be remembered, followed by a delay which is filled with a task involving material that usually does not have to be remembered, followed by a test of the first list. The test may be event or associative recognition or recall and may be either a complete recall or recognition of the list or a probe recall or recognition of an individual event in the list. Strength decay curves for such three-phase studies (e.g., Peterson and Peterson, 1959; Murdock, 1961; Helleyer, 1962; Melton, 1963) deviate from a simple exponential decay in a manner which suggests the presence of a significant amount of ITM (Wickelgren, 1970d). This presumably results from the relatively greater temporal separation between repetitions of an item in these studies. However, studies with longer delays are needed to determine whether the decay curve is well fit by the sum of two exponentially decaying traces with very different time constants. Furthermore, according to multitrace theory, it ought to be possible to achieve a simple exponentially decaying trace with the three-phase design by modifications that decrease the temporal separation between repetitions of an item, for instance, reducing the population of items, increasing the number of items in a list, requiring learning of the interpolated material and using interpolated material from the same population as the list.

Continuous Studies of Verbal ITM

Continuous recognition studies (e.g., Donaldson and Murdock, 1968; Melton, Sameroff, and Schubot, 1963; Shepard and Teghtsoonian, 1961) involve presenting a long list of events, each of which is a new event to learn, a delay filling event, and a test event. In the multiple-choice version of continuous recognition memory (e.g., Shepard and Chang, 1963), the subject sees two (or more) events on each trial, must learn both, and must choose the one that has been presented before. In continuous recall studies (e.g., Atkinson, Brelsford, and Shiffrin, 1967; Brelsford, Shiffrin, and Atkinson, 1968), a trial has two parts: a test phase followed by a study phase. In the study phase, one pair of events is presented to be learned. In the test phase, a cue event is presented for which the subject is to recall the correct response event.

Continuous recognition studies are one-phase designs and continuous recall studies are two-phase designs. However, the continuous design has an enormous advantage over other designs, namely, the efficiency with which really long delays can be studied. In addition, it is easy to have short delays also, so one can easily study the memory trace over a very wide range of delays. Finally, the number of events to be remembered is often (but not always) so large that one can be sure that conscious rehearsal is of negligible significance, except perhaps at very short delays in the absence of non-rehearsal instructions.

When probabilities of correct recall are transformed to strengths in the continuous recall studies of Atkinson, Brelsford, and Shiffrin (1967) and Brelsford, Shiffrin, and Atkinson (1968), very nice fits to simple exponential decay of ITM are found over a range from 10 seconds to 3 minutes. Strength decay curves for continuous recognition memory over roughly the same period of time are also well fit by a simple exponential decay (Melton, Sameroff, and Schubot, 1963; Shepard and Teghtsoonian, 1961; unpublished data of mine). The decay rates for these different studies are all within the same order of magnitude (time constant of around 3 minutes), but they are not invariant over the different conditions. Thus, just as for STM, the functional decay rate for ITM must be assumed to vary over a range. What conditions affect the functional decay rates for ITM are not yet clear, but a strength analysis of the Melton, Sameroff, and Schubot (1963) study suggested that the number of intervening items to be learned is irrelevant, provided the total delay time is constant. This must be replicated in future studies with longer delays, but, if it is true, it is an important difference between ITM and STM. Within the same theoretical framework used to explain the dependence of STM decay rate on the rate of presentation of items to be learned, one would conclude that ITM has no active trace maintenance process or at least not one which interacts with the trace acquisition process.

All this would make a fairly neat story if it were not for the presence in some unpublished data of mine on continuous recognition memory (one example of which is shown in Wickelgren, 1970a) of yet another, more slowly decaying ITM trace, which appears when delays of 3 to 12 (or presumably more) minutes are used. One ITM is bad enough without two of them. However, at present, the continuous recognition decay curves from 10 seconds to 12 minutes do look somewhat more like the sum of two exponential decays than like a monotonic deceleration of decay rate (Jost's Second Law). Obviously, much more data are needed before the number of ITM traces and the relations between them are clear.

Verbal LTM

Unpublished studies of mine of recognition memory for Russian-English word pairs over a 2-year retention interval yielded decay curves which were not well fit by a single exponentially decaying trace. However, the experiment was not well suited to the determination of the shape of the decay function, being intended only to determine the order of magnitude of the decay rate in verbal LTM for overlearned verbal materials. Thus, further experiments will be needed to determine whether the decay function for LTM is a single exponential, the sum of two exponentials, or some version of Jost's Second Law.

The time constant of the LTM decay curve for overlearned Russian-English word pairs was on the order of 2 years. This means that there is a factor of around 10^7 or 10^8 between the decay rates of the fastest decaying verbal STM and a very slowly decaying verbal LTM. This is truly a staggering range of decay rates for human memory. This enormous range suggests both that several different types of traces may be required and that each trace may have a range of possible decay rates over a factor of 10 or 100.

Successive Comparison of Pitch

All psychophysical successive comparison studies involve memory for the standard (S) stimulus which is judged in relation to the comparison (C) stimulus. Successive comparison studies are recognition memory studies, and decay curves can be determined by plotting the decline in the discriminability of the trace strength for C stimuli identical to or different from the S stimulus as a function of the delay between the two stimuli. In the case of a one-dimensional judgment, such as pitch over intervals less than one octave, there are a large variety of possible C stimuli that could be chosen differing from the S stimulus by different amounts on the relevant dimension.

In the case of pitch judgments for S tones, Wickelgren (1969a) has investigated the discriminability of C tones identical to the S tone versus C tones that are 10, 20, 30, 40, or 50 Hz. different from the S tone. The decay curves from 0 to 3 minutes are well fit by the sum of two exponentially decaying traces, with the STM and ITM components having about the same decay rates as the STM and ITM components over the same delays in verbal memory experiments. Acquisition of the STM trace approaches a limit (approximately exponentially) as a function of S tone duration, and the rate of decay of STM is invariant over the degree of acquisition manipulated in this way. The rate of decay of STM for pitch is also invariant with the distance of the C tone from the S tone over the range from 10 to 50 Hz. Although I have felt it was necessary to fill the delay interval between the S and C tones with another tone to control rehearsal of the S tone, the intensity of this interpolated tone and its similarity to the S and C tones (beyond 50 Hz.) are irrelevant to the decay of STM trace for pitch.

CONCLUSION

Multitrace theory requires much more extensive testing before any definite conclusion can be reached regarding its suitability as a general theory of memory. Very likely, future testing will indicate that modification or extension of the theory is required, especially to specify the nature of the dependence of the parameters of trace dynamics upon various conditions. However, the theory already has considerable generality, and there is every reason to hope that future modifications and extensions can be made within the same basic theoretical framework.