Annu. Rev. Psychol. 2002. 53:53–81

REMEMBERING OVER THE SHORT-TERM: The Case Against the Standard Model

James S. Nairne
Purdue University, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907-1364;
e-mail: nairne@psych.purdue.edu

Key Words short-term memory, rehearsal, decay, activation, working memory

■ **Abstract** Psychologists often assume that short-term storage is synonymous with *activation*, a mnemonic property that keeps information in an immediately accessible form. Permanent knowledge is activated, as a result of on-line cognitive processing, and an activity trace is established "in" short-term (or working) memory. Activation is assumed to decay spontaneously with the passage of time, so a refreshing process—rehearsal—is needed to maintain availability. Most of the phenomena of immediate retention, such as capacity limitations and word length effects, are assumed to arise from trade-offs between rehearsal and decay. This "standard model" of how we remember over the short-term still enjoys considerable popularity, although recent research questions most of its main assumptions. In this chapter I review the recent research and identify the empirical and conceptual problems that plague traditional conceptions of short-term memory. Increasingly, researchers are recognizing that short-term retention is cue driven, much like long-term memory, and that neither rehearsal nor decay is likely to explain the particulars of short-term forgetting.

CONTENTS

0084-6570/02/0201-0053$14.00

INTRODUCTION

How do we remember over the short term? It is clearly adaptive to keep recent information available in some kind of accessible form. It would be difficult to comprehend spoken language, which occurs sequentially, or read any kind of text without remembering the early part of an utterance or the themes relevant to a passage. Virtually all complex cognitive activities—reading, reasoning, problem-solving—require access to intermediate steps (as in adding or multiplying two-digit numbers in the head) or other situation-specific information (Baddeley 1992, Kintsch & VanDijk 1978). What then are the psychological mechanisms, or systems, that drive and control such short-term maintenance?

For many years psychologists have essentially agreed about the main mechanism controlling the temporary storage of information. The generally accepted view—referred to here as the standard model—is that short-term storage arises from *activation*, a mnemonic property that keeps information in an immediately accessible form. Permanent knowledge is activated, as a byproduct of on-line cognitive processing, and comes to reside "in" short-term (or working) memory. Short-term memory, as a whole, is simply defined as the collective set of this activated information in memory (e.g., Cowan 1995, Shiffrin 1999). Activation is assumed to be fragile, however, and it can be quickly lost—through the operation of *decay*—in the absence of rehearsal. The decay process has adaptive value because it enables us to update our memories continuously, removing recently activated information that is no longer needed (e.g., Bjork 1975). When necessary, rehearsal can counteract decay, refreshing activation, much like a juggler can temporarily defy the force of gravity by repeatedly tossing plates back up into the air.

The simplicity of the standard model is clearly a virtue, but is it empirically justified? On the surface, the standard model violates a number of well-known tenets of memory theory. For example, decay has been roundly rejected as a vehicle for long-term forgetting for decades, certainly since the seminal arguments of John McGeoch in the 1930s (McGeoch 1932). More importantly, remembering in the standard model is tied directly to activation, an inherent property of a trace; items are more recallable, or closer to the recall threshold, because they are in an "active" state (e.g., Engle et al. 1999, Lovett et al. 1999). Yet, one of the lessons of modern memory theory is that items do not have "strength," or special mnemonic properties, outside of particular retrieval environments. Remembering is always a joint function of trace properties and retrieval cues (e.g., Tulving 1983), so how can a single trace property—activation—be synonymous with remembering?

It is possible that remembering over the short term is a special case, requiring theoretical proposals that do not apply in more general arenas, but the empirical case should be strong. As I review in the next section, the empirical case for the standard model did, in fact, once seem very strong (see also Baddeley 1990). A number of empirical phenomena strongly support a time-based rehearsal plus decay model. However, as this chapter demonstrates, the case for the standard model has lost much of its secure footing over the past decade. The theoretical viability of its main assumptions, particularly those centering on rehearsal and decay, is now in question. In the last section of the chapter I consider alternative ways of conceptualizing how we remember over the short term. Increasingly, researchers are recognizing that short-term retention is cue driven, much like long-term retention, and that simple appeals to either rehearsal or decay are unlikely to explain the particulars of short-term forgetting (e.g., Brown et al. 2000; Lewandowsky 1999; Nairne 1996, 2001; Tehan & Fallon 1999).

THE STANDARD MODEL: REHEARSAL PLUS DECAY

As noted above, activation is the vehicle for temporary storage in the standard model. Items, once activated, are assumed to exist in a state of immediate accessibility (McElree 1998); the amount of activation, in turn, accrues from a continual trade-off between rehearsal and decay. This account has intuitive appeal—it maps well onto phenomenological experience—and it is easily expressed with concrete metaphors. For instance, think about a juggler trying to maintain a set of four plates. Tossing a plate can be seen as a kind of activation; the height of the toss corresponds roughly to the amount of activation achieved. The juggler is able to maintain a set of activated items (plates) to the extent that each can be caught and re-tossed before gravity reduces it to an irretrievable state.

The juggling metaphor is apt because it can be naturally extended to most of the well-known phenomena of immediate memory (see Nairne 1996). For example, there is a content capacity associated with juggling—the number of plates that can be held aloft—that depends on a trade-off between the juggler's skill and the constant force of gravity. If the hand movements of the juggler are particularly skillful, content capacity increases just as the content capacity of immediate memory (memory span) appears to increase directly with the speed or effectiveness of rehearsal. Moreover, the content capacity of a juggler is not a fixed property of a limited-capacity space—air—but rather depends directly on the effectiveness of the reactivation process. A juggler can easily double his or her net content capacity by binding plates together, just as the capacity of immediate memory can be dramatically increased through semantic binding into chunks (Miller 1956).

The main assumptions of the standard model—activation, rehearsal, and decay—are prominent in a host of current theoretical accounts of short-term retention. Probably the best-known example of a "juggler" model is the working memory model of Baddeley & Hitch (1974; see also Baddeley 1992), particularly the component known as the phonological loop. The phonological loop is the subsystem of working memory that controls short-term retention of verbal material. The loop is divided into two parts: a phonological store, which is the storage location for activated information, and a rehearsal/recoding device called the articulatory control process. Information in the phonological store is assumed to decay in roughly 2 s (which is analogous to a constant force of gravity) and can be refreshed, via rehearsal, by the articulatory control process (tossing plates into the air). Capacity limitations in immediate retention—e.g., the magic number seven—are assumed to arise from trade-offs between decay and loop-based rehearsal.

There are many variants of the working memory model in the current literature (see Miyake & Shah 1999) and there are even updated versions of the Baddeley model itself. For example, Baddeley (2000) has recently added another temporary storage system to his model—the episodic buffer—to explain, in part, how verbal material can be maintained when the phonological loop is unavailable. A number of formal simulation models also mimic the standard model and Baddeley's model in particular (e.g., Burgess & Hitch 1999, Henson 1998, Page & Norris 1998). Each relies on rehearsal plus decay to explain the particulars of short-term retention, although most of the simulation models do contain additional machinery to handle cue-driven retrieval effects. I review a selection of these simulation models in more detail in the final major section of this chapter.

Finally, Cowan's (1988, 1995) conception of short-term memory also shares many, but not all, of the features of the standard model. Cowan assumes that short-term memory represents a nested subset of long-term memory; specifically, it comprises those portions of long-term memory that are currently active. Cowan further distinguishes among activation levels—e.g., some items are especially active because they sit in the focus of attention—but it is important to note that (*a*) short-term retention is generally tied to an item's activation level and (*b*) activation is assumed to be lost, normally through decay, as a direct function of time. Cowan acknowledges that factors other than rehearsal and/or decay can influence activation levels and retention, but the basic components of the "juggler" model are well represented in the Cowan model.

Articulation Rate and Span

Empirically, the assumptions of the standard model receive their most convincing support from studies showing a systematic relationship between overt articulation rate and memory span. Articulation rate is assumed to correlate with the speed of internal rehearsal (Landauer 1962), and it turns out to be related in a roughly linear fashion to remembering over the short-term (Baddeley et al. 1975, Schweickert & Boruff 1986). The faster the overt articulation rate, the higher the recorded memory span. One of the best-known examples of this relationship is the word-length effect: Lists of short words (defined in terms of spoken duration) are remembered better than lists of long words, holding constant factors such as letter length and number of syllables (Baddeley et al. 1975).

This relation between articulation rate and memory span is well documented and noncontroversial, although some questions remain about the exact form of the function—e.g., linear or possibly quadratic (Brown & Hulme 1995, Schweickert et al. 1990). The relationship holds at the level of aggregate data as well as for individual subjects. Individual differences in immediate memory performance, for both children and adults, are predicted well by individual articulation rates (Baddeley et al. 1975, Tehan & Lalor 2000). One can even predict differences in digit span cross-culturally by knowing how rapidly digits can be pronounced in a given language. If digit names can be spoken relatively quickly, as in English or Chinese, memory span tends to be higher than if the corresponding digit names take longer to pronounce (e.g., Welsh or Arabic) (Ellis & Hennelly 1980, Naveh-Benjamin & Ayres 1986).

Additional evidence comes from studies measuring retention in the absence of rehearsal. Rehearsal can be prevented, or at least disrupted, by requiring subjects to engage in articulatory suppression during the study and recall of a memory list (saying "the" or "bla" repeatedly aloud); alternatively, one can test young children who lack the ability to rehearse strategically. In the former case—with articulatory suppression—immediate memory performance declines significantly and the word-length effect disappears (Baddeley et al. 1984). In the latter case, young children who fail to rehearse consistently often show little sensitivity to spoken duration in immediate retention, at least when material is presented visually (Gathercole & Hitch 1993). As children age, their mean articulation rate increases and their memory span grows in a roughly proportionate manner (Hitch & Halliday 1983, Hulme et al. 1984, Nicolson 1981).

Collectively, these data seem to offer compelling support for the standard model. Articulation rate, and inferentially the speed of internal rehearsal, varies in a more or less direct way with memory span. In the absence of rehearsal, immediate memory performance declines and no longer shows the clean connection with spoken word duration. However, as we shall see, the empirical case loses much of its vigor on closer inspection. Over the past decade, a number of important exceptions to the data patterns listed above have appeared: For example, it turns out that articulation rate and span are not always closely related, and spoken duration effects sometimes emerge in the absence of rehearsal. Moreover, the theoretical assumptions of the standard model, when examined closely, turn out to contain an unsettling amount of ambiguity. In the sections that follow, I examine the discrepant data—first for rehearsal and then for the process of decay—and discuss the conceptual problems that remain unresolved.

PROBLEMS WITH REHEARSAL

I begin by discussing some of the empirical difficulties associated with the rehearsal arm of the standard model. At the outset, it is worth noting that proponents of the standard model (e.g., Baddeley 1992) rarely, if ever, specify the dynamics of the rehearsal process in any kind of systematic way. As Brown & Hulme (1995) observed, "if rehearsal does exert an important causal influence on memory span, it is clearly necessary to know exactly when such rehearsal is taking place" (Brown & Hulme 1995, p. 599). Yet, the protocol of covert rehearsal is likely to change with a host of variables, such as list length, modality of presentation, presentation rate, lexicality, or interitem similarity. Does rehearsal occur only during stimulus input, or does it also occur during response output? Does it matter whether one rehearses based primarily on sound, or is it necessary to access meaning in order to maintain the activation of permanent knowledge? The idea of the standard "rehearsal + decay" model seems simple enough, but its actual implementation is likely to be quite complex.

Dissociating Rehearsal and Span

In its simplest form, the standard model assumes that most of the variability in immediate retention is attributable to rehearsal. Decay is presumably fixed, like gravity, so effective remembering over the short term hinges on one's ability to maintain activation in the face of decay. Such an account is clearly undermined by the fact that recall differences can emerge when articulation rates are held constant. For example, Schweickert et al. (1990) measured pronunciation rates for lists containing phonologically similar and dissimilar items. Similarity had no significant influence on pronunciation rate (similar = 3.01 items/sec; dissimilar = 2.92 items/sec), but produced substantial memory span differences (similar = 5.62 items; dissimilar = 7.06 items). Comparable results were obtained by Hulme & Tordoff (1989) in a study using children, although small differences in articulation rate favoring the dissimilar items were found in one of their experiments.

Dissociations between articulation rate and span have been reported frequently over the past decade. Hulme et al. (1991) found that words produced significantly higher memory spans than nonwords, even when articulation rates were matched for the item types; in fact, the articulation rates for nonwords were actually slightly higher than for the words, yet the words consistently produced the higher memory spans (Hulme et al. 1999). Similar results have been found recently for high- and low-frequency words (Hulme et al. 1997, Roodenrys et al. 1994) and for concrete versus abstract words (Walker & Hulme 1999). In each case, clear immediate memory differences emerged that were independent of articulation rate.

Other researchers have found that manipulations of item duration sometimes fail to produce concomitant changes in immediate memory performance. Caplan et al. (1992) used lists containing short- and long-duration words matched for number of syllables and phonemes and found no memory advantage for the short-duration words; in fact, the word length effect actually reversed when duration was manipulated by using "lax" vowels of short duration (e.g., carrot) and "tense" vowels of long duration (e.g., spider). Caplan et al. (1992) pointed to the phonological structure of a word, rather than its spoken duration, as the major determinant of the magnitude of the word length effect. Service (1998) reached a similar conclusion in a study using Finnish stimuli, which allow one to vary duration by manipulating combinations of the same acoustic and articulatory features. Lastly, Lovatt et al.

(2000) varied spoken duration in disyllabic words, holding constant a host of potentially confounding factors (e.g., frequency, familiarity, number of phonemes) and failed to find any advantage for short-duration words across several experiments; word duration effects emerged only when the original word pools used by Baddeley et al. (1975) were employed as stimuli.

Ironically, Baddeley et al. (1975) suggested decades ago that certain variables such as meaningfulness, list length, and interpolated delay might affect span without significantly changing reading or articulation rate (Baddeley et al. 1975, p. 583). The data just described confirm this basic intuition, as does the companion finding that items of differing spoken duration need not lead to measurable changes in immediate memory performance. These results are inconsistent with any simple form of the standard model, unless one is willing to assume that decay rates are variable across materials or that additional storage assumptions (e.g., the episodic buffer) can be added to handle the troublesome data. One could explain the findings of Schweickert et al. (1990), for example, by assuming that items decay faster in an environment high in phonological similarity (i.e., Posner & Konick 1966). One might also assume that decay rates are relatively faster for nonwords, low-frequency words, and abstract words, but such assumptions detract from the elegant simplicity of the rehearsal plus decay idea.

Other evidence can be marshaled in support of articulatory processes in immediate retention, although the implications of the data are unclear. For example, Wilson & Emmorey (1998) recently found a duration-based word length effect using word lists generated from American Sign Language. Lists composed of short signs were remembered better than lists of long signs under conditions in which "length" could be attributed uniquely to the duration of the signing process. Of course, one can question the relevance of these data to immediate retention and rehearsal of verbal lists, although Wilson (2001) argues that memory for sign language follows the same general principles as memory for spoken lists. In addition, Cowan et al. (1997b) were able to produce an artificial word length effect by instructing subjects to pronounce words either quickly or slowly at presentation and during recall. Given that the same word set was used in the "short" and "long" condition, these data suggest a role for articulation time (see also Cowan et al. 2000). Again, however, the experimental conditions are unusual and it is difficult to know what effect the instructional manipulation truly had on the subject's retention strategies (see Service 1998).

Eliminating Rehearsal

As noted earlier, the fact that item-duration effects can be eliminated under articulatory suppression is usually seen as strong support for the rehearsal arm of the model. Baddeley et al. (1975) found that the word length effect was eliminated under articulatory suppression when lists were presented visually but not aloud. Later work in the same laboratory found that the word length effect could be eliminated under auditory presentation as long as the subject was required to continue the articulatory suppression throughout the recall period (Baddeley et al. 1984). Why the pattern depends on the modality of presentation has never been adequately explained, although the standard model can handle the finding by assuming that the dynamics of rehearsal depend on presentation modality. For example, Baddeley et al. (1984) suggested that there is a "greater compatibility of an articulatory response to auditory material than to visual," perhaps as a consequence of language learning.

More recent evidence, though, further complicates the picture. LaPointe & Engle (1990) found that articulatory suppression eliminates the word length effect only when list items are drawn from a restricted set. If different words are used on every trial, the word length effect remains under articulatory suppression even for visually presented items. Similarly, Caza & Belleville (1999) tested immediate serial recall of words and nonwords under articulatory suppression (during input and the output period) and still found a large advantage for the words (see also Bourassa & Besner 1994, Tehan & Humphreys 1988). Thus, the presence of active rehearsal is not needed to obtain item-based differences in immediate retention.

Word length effects also occur under conditions of extremely rapid visual presentation. Coltheart & Langdon (1998) presented lists of either short or long words at rates approximating 100 msec/item. Presenting an entire list in well under a second, it was argued, should make normal rehearsal virtually impossible. Yet, a highly reliable word length effect was found; moreover, the word length effect was eliminated, under these same rapid presentation conditions, when subjects engaged in concurrent articulatory suppression. Word length effects are also found in patients suffering from serious articulatory difficulties. Baddeley & Wilson (1985) studied patients with dysarthria—that is, patients who lack the capacity to control their articulatory muscles. These individuals cannot speak overtly, although more central aspects of speech planning may remain intact, yet show significant word length effects in immediate recall (see also Bishop & Robson 1989, Vallar & Cappa 1987).

At face value, these data seem troubling for the standard model. In the absence of rehearsal, word length effects can remain. However, the standard model does not necessarily predict that the word length effect should be eliminated under these conditions: Long words, by definition, are associated with longer presentation durations. This means that one would expect the word length effect to disappear in the absence of rehearsal only when the total presentation time has been equated for short and long lists. Suppose it takes 1 s to present a long word and only 0.5 s to present a short word. For pure lists of any length, then, the retention interval for a given long item in the list, with the exception of the last list item, will be longer than for the comparable item in the short list (e.g., for lists of 6 items, 5 s will elapse before the first long item can be recalled versus only 2.5 s for the first short item). Assuming decay, we would expect a word length effect under these circumstances, even without considering rehearsal as a factor (or by assuming that all words are rehearsed only once). In many experimental studies of the word length effect, item presentation rates are controlled (e.g., 1–2 s per item) so the total passage of time elapsing during presentation is, in principle, controlled. However, as reviewed in the next section, the main locus of the word length effect appears to be at output—during recall, when output timing is rarely controlled experimentally.

Output Effects

There is considerable evidence to suggest that word length effects originate, at least in part, during recall output. First, as discussed above, when list items are presented aloud, articulatory suppression eliminates the word length effect only if it occurs during output. Importantly, Baddeley et al. (1984) controlled for output times in these studies by allowing subjects to write down long words in an abbreviated form (e.g., hippo for hippopotamus). Thus, assuming control over the passage of time, these data are largely consistent with the predictions of the standard model (assuming some rehearsal occurs at output). Second, several investigators have found that the word length effect is reduced or eliminated when retention is tested using a probe task rather than serial recall (Avons et al. 1994, Henry 1991). In a probe task, subjects are required to recall only one item at test from a randomly chosen serial position, thus sharply reducing the potential for output-based memory effects.

Converging evidence comes from studies employing mixed lists of short and long words. Cowan et al. (1992) varied whether short and long words occurred early or late in a presentation sequence—i.e., there were lists that began with short words and ended with long words and vice versa. At the point of test, subjects were cued randomly to recall the items in either a forward or a backward direction. The important finding was that recall level was determined by which kind of word was output first. When short words were recalled first (i.e., forward recall of short-long lists or backward recall of long-short lists), performance was better than when long words were output first. Note that at the point of recall presentation time was equated for all the list types, so the performance differences must have arisen during output.

Of course, pinpointing the locus of word length effects at output does not mean that rehearsal is responsible. Again, the raw passage of time could be critical: Recalling long words first imposes more of an output delay on the remaining words than recalling short words first. In fact, when spoken recall protocols are analyzed in microscopic detail, it turns out that both the duration of the recall response and the pauses that occur between recall responses are importantly related to span. A number of studies have found that interword pause durations correlate significantly with span (e.g., Cowan et al. 1998, Tehan & Lalor 2000), yet these durations are far too short for any useful covert rehearsal (ranging from 100 to 300 msec). Moreover, output pauses are not affected by word length, nor do they correlate significantly with articulation rate (Cowan et al. 1998). Instead, they tend to correlate with retrieval factors such as memory search rate or the ease of cue-driven redintegration (see Hulme et al. 1999). It is difficult to see from these data how covert rehearsal processes could be involved.

Finally, if the durations of the output pauses are too short for rehearsal, then when can rehearsal be occurring during output? There are two remaining possibilities: Rehearsal could be occurring during the actual output responses or during the preparatory interval preceding recall. The former case seems unlikely—in fact, one could argue that rehearsal at any point after output has begun is likely to have a detrimental effect because one would need to coordinate two sets of items, the output set and the rehearsal set (Avons et al. 1994). It is more likely that some rehearsal occurs during the preparatory interval immediately preceding recall. However, these intervals usually show no correlation with either memory span or articulation rate (e.g., Cowan et al. 1998, Hulme et al. 1999, Tehan & Lalor 2000). From this perspective, then, there is little support for the proposal of output-based rehearsal processes.

Summary

The widely held belief that internal rehearsal plays a large role in remembering over the short term, refreshing activation, rests primarily on the well-documented relationship between articulation rate and span. However, as the preceding section indicates, there are many reasons to question whether the correlation between articulation and span is actually caused by rehearsal. Perhaps most telling are the repeated failures to replicate item duration effects when other potentially confounding factors are controlled (e.g., Caplan et al. 1992, Service 1998, Lovatt et al. 2000). Short and long items differ on a number of dimensions other than spoken duration (e.g., familiarity, frequency, phonological complexity) and any of these factors (or their combination) could potentially contribute to memory success. Early reports indicated that there might be a pure effect of time when these other factors are controlled (Baddeley et al. 1975), but more recent research suggests otherwise.

It is also well documented that many item-based differences in immediate retention remain when (*a*) rehearsal is disrupted or blocked and (*b*) articulation rates are held constant. In the former case, the passage of time could be held responsible—that is, through differential decay—but there is no simple way for the standard model to explain the large span differences found at matched articulation rates. Thus, the span advantage seen for words over nonwords, high-frequency over low-frequency words, or concrete over abstract words must be attributable to factors other than those contained in the rehearsal plus decay assumptions of the standard model. What then accounts for the overall relationship between measures of articulation and immediate retention? One possibility is that there is some unspecified third factor that is associated with both rapid articulation ability and memory skill. Cowan (1999) recently suggested that "the ability to read or extract information from phonological memory quickly" might be involved; Tehan & Lalor (2000) agree, proposing that lexical access is a more potent predictor of individual differences in span than articulation rate. Others have suggested that the length of a word, and thereby its complexity, affects articulation rate as well as the ease of interpreting or "deblurring" the retrieval cues needed for immediate retention (e.g., Neath & Nairne 1995). I return to interpretations of this type in the section Reconceptualizing Short-Term Memory.

Finally, what is the role of rehearsal in remembering over the short-term? The subjective reality of rehearsal seems clear enough, and there is evidence from neuroimaging work that internal rehearsal is correlated with distinct neural activity (see Awh et al. 1996, Smith & Jonides 1997). It seems likely that rehearsal could serve as a mechanism for re-presentation of the stimulus material—that is, discrete rehearsals might produce copies of an item in memory (as in overt recall) that enhance later memory. Multiple presentations would be expected to increase the variability of contextual encoding and effectively shorten the functional retention interval (see Tan & Ward 2000). Importantly, however, rehearsal produces its benefits in this scenario not because it refreshes a decaying trace, as in the juggler model, but rather because it promotes the success of a cue-driven retrieval process.

PROBLEMS WITH DECAY

The second arm of the standard model is *decay*, defined as the loss of trace information exclusively as a function of time. Activation, the vehicle for short-term storage, is believed to decay spontaneously, much like a plate tossed into the air falls spontaneously from the force of gravity. Of course, the fact that forgetting proceeds with time is self-evident—but is it really the passage of time that causes the forgetting? Time is correlated with forgetting, but it may be the events that happen in time that are actually responsible for the loss. The famous analogy used by McGeoch (1932) was of an iron bar left out to rust in the open air—the accumulation of rust develops over time, but it is oxidation, not time per se, that is actually responsible.

With the exception of short-term memory environments, decay is rarely, if ever, used by memory theorists seeking to explain forgetting. There are both empirical and theoretical reasons to reject the concept. First, long-term retention can stay constant, or even improve, with the passage of time (e.g., spontaneous recovery and reminiscence). Second, the rate and extent of forgetting often depend on the specific activities that occur during the retention interval. For example, the similarity between original and interpolated learning determines how much forgetting one sees for the original material (e.g., Osgood 1953). Neither of these empirical results can be explained through a simple appeal to decay. Importantly, as discussed below, both of these results apply to remembering over the short term.

Theoretically, the concept of decay is equally troubling. To propose that memories are lost spontaneously with time ignores the potential contribution of the retrieval environment (e.g., Tulving 1983). In the standard model remembering is essentially tied to a property of the trace—activation—and no claims are made about the effectiveness of retrieval cues. One could assume that short-term memory is also influenced by the availability of retrieval cues (e.g., Nairne 1990a, 2001), but this undercuts the explanatory power of the decay concept. Trace features might be lost over time, in a process akin to radioactive decay, but the effects on memory will depend on the cues present at the time of test. In principle, losing trace features could produce a trace that is actually more compatible with the cues present at time 2 than at time 1, producing little loss or even improved retention. Such a view is widely accepted in the study of long-term memory, in which the passage of time per se is rejected as a sufficient condition for forgetting, but similar analyses are rarely, if ever, made in the short-term memory literature.

Why does decay remain popular? Part of the reason may be phenomenological: We have all experienced the rapid loss of information from consciousness, although the specific cause of the loss is not consciously apparent. More likely, though, the reason is historical. In the original Brown-Peterson experiments (Brown 1958, Peterson & Peterson 1959) requiring subjects to count backward, or read a list of digits, after presentation produced rapid loss in memory for letters. This result was considered striking for two reasons: First, the length of the to-be-remembered series was well below span (e.g., three consonants); second, the to-be-remembered information (letters) and the distractor activity (counting backward) were highly dissimilar. Similarity was believed to be the major determinant of interference at the time, so it was difficult for researchers to explain the rapid loss. Of course, as I discuss below, subsequent work showed that it is not interference from the distractor task per se that is the major cause of forgetting; instead, it is the interference produced by memories from prior trials (e.g., Keppel & Underwood 1962, Nairne et al. 1997).

Dissociating Time and Forgetting

In fact, it is relatively easy to falsify simple versions of decay theory. If time passes and there is no memory loss, or perhaps an improvement in retention, then factors other than decay must be operating. In the standard model rehearsal counteracts decay through reactivation of the trace. However, a number of studies have found little, or no, forgetting in contexts in which rehearsal is unlikely. For example, Greene (1996) found no evidence for loss in the traditional Brown-Peterson paradigm when distractor length was manipulated between subjects (see also Turvey et al. 1970). In the more traditional application, in which retention interval is manipulated within subjects, Keppel & Underwood (1962) reported essentially no forgetting on the first trial. Comparable results have been found by others (e.g., Gorfein 1987), although Baddeley & Scott (1971; see next section) did report small amounts of loss on the first trial during the first few seconds of the retention interval.

Keppel & Underwood (1962) suggested that interference from prior trials is responsible for forgetting in most short-term memory environments. Researchers commonly use the same set of items across trials, or sample from a small set (e.g., consonants), so one could easily become confused about the exact position occupied by an item on a given trial. Nairne et al. (1999) attempted to minimize interference by using different items on every trial and reconstruction of order as the retention measure; in a reconstruction test the just-presented items are given back in random order and the task is to place the items into their original presentation order. Under these conditions very little evidence of forgetting was found across retention intervals ranging from 2 to 96 s—e.g., reconstruction performance averaged 78% correct after 2 s of distraction and 73% correct after 96 s.

Historically, proponents of the standard model have explained data like these by appealing to long-term memory (e.g., Atkinson & Shiffrin 1968, Cowan 1995). In the absence of interference, such as on the first trial of an experimental session, subjects can recover information from long-term memory, which essentially masks the contribution of decay in short-term memory. Thus, little short-term forgetting is expected under these circumstances, even though the short-term memory trace presumably continues to decay. This kind of account can easily handle the recent findings of Nairne et al. (1997), showing that the word length effect, like distractor effects, emerges only after several trials in an experimental session. Again, on early trials both item types can be recovered from long-term memory, where decay and activities designed to prevent decay (rehearsal), play less of a role. As long-term memory becomes cluttered from prior trials, however, the subject shifts strategically to short-term memory, for which time-based factors are important.

On reflection, it is difficult to see why this account has such wide appeal. First, there is no direct evidence confirming that subjects actually do shift from long- to short-term retrieval within an experimental session. Moreover, Nairne & Kelley (1999) found that the phonological similarity effect remains robust when interference from prior trials is minimized. If subjects tend to rely on recovery from long-term memory in an "uncluttered" environment, then phonological similarity should play a reduced role (i.e., because long-term retrieval tends to be influenced more by semantic factors). Second, and more troubling, short-term memory is assigned a kind of "back-up" status in this account—that is, something to be used only when recovery from long-term memory is problematic. It suggests that a subject's first line of defense is to recover information from long-term memory; recovery from short-term memory occurs only when long-term traces are difficult to access. Yet, of course, in most natural environments people have not been exposed to multiple "lists" of to-be-remembered information. Thus, in the prototypical case—remembering a telephone number—it is not clear why short-term memory would even be needed.

A better explanation is given by general distinctiveness accounts, which assume that forgetting is caused by a failure to discriminate current trial information from information presented on previous trials (e.g., Baddeley 1976, Brown et al. 2001). There is limited forgetting on the first trial in a Brown-Peterson task because there are no prior trials to produce interference; also, when interference is reduced by using different items on every trial, the amount of forgetting should be sharply reduced (Nairne et al. 1999). The crucial role played by prior trials is illustrated nicely in a study by Turvey et al. (1970) that uses a version of the Brown-Peterson task. Different groups of subjects were asked to count backward as a distractor activity for group-specific intervals (e.g., one group counted for 10 s, another for 15 s, and another for 20 s). Foreshadowing Greene (1996), equivalent amounts of forgetting were found across groups in this between-subject design (0.33, 0.30, and 0.30, respectively). On the critical trial, however, all groups were switched to the same 15 s distractor period. Retention performance dropped in the 10-s group (from 0.33 to 0.20), stayed roughly constant in the 15-s group (0.30 to 0.28) and improved in the 20-s group (0.30 to 0.38). Note that the passage of time—and therefore the opportunity for decay—was equated across the groups on the critical 15-s trial, yet performance depended on the timing of prior trials.

One can parsimoniously explain these data by assuming that the relative, rather than absolute, durations of distractor activity affect the discriminability of target information at test (Baddeley 1976). Memory improves whenever target items can be easily discriminated from the memories created by prior trials. According to distinctiveness accounts, it is the ratio of the interpresentation interval to the length of the current retention interval that matters. In the Turvey et al. (1970) case, the interpresentation interval corresponds to the period separating the presentation of the to-be-remembered information on trial N-1 from the presentation on trial N. Note that on the critical trial containing 15 s of distraction, prior trial information is relatively "closer" in time for the 10-s group ($10/15 = 0.67$) than for the 15-s group ($15/15 = 1.0$) or the 20-s group ($20/15 = 1.33$). The actual data are predicted well by these ratios. The Turvey et al. (1970) data are particularly important in the present context because they show that short-term memory performance can decrease, stay the same, or even increase depending on timing variables. Comparable results have been found in other contexts: Neath & Knoedler (1994), for example, showed that memory for early items in a list sometimes improves as the length of the retention interval increases (see also Bjork 2001, Wright et al. 1985). Once again, such a finding is inconsistent with the notion of decay, although it can be handled relatively easily by simple distinctiveness accounts (Neath 1998).

The Role of Intervening Activity

The preceding studies establish quite clearly that the passage of time is not necessarily a predictor of memory loss. Time is often correlated with forgetting, but the correlation is far from perfect. The exceptions are important, though, because they help to falsify simpleminded notions of decay. Decay theories also have trouble explaining why forgetting can depend on the specific activities that occur during the retention interval. In his original work on the distractor paradigm, Brown (1958) found that the similarity between the distractor activity and the to-be-remembered material played only a minor role. However, subsequent researchers have established that the nature of the distractor activity clearly does matter (see Greene 1992).

For example, significantly more forgetting is found in a short-term memory environment when the modalities of presentation and distraction match. More forgetting is found when target items are presented aloud and the distractor activity is also auditory (Elliott & Strawhorn 1976, Proctor & Fagnani 1978). Semantic (Dale & Gregory 1966) and phonological (Wickelgren 1965) similarity between the presented items and the content of distraction also increase the amount of short-term forgetting. Interference from prior study trials may be the major determinant of short-term forgetting (i.e., proactive interference), but it is now generally agreed that similarity-based retroactive interference also plays a role.

Additional support for modality-specific interference comes from the extensive literature on the suffix effect (for reviews see Crowder 1976, Greene 1992). The suffix effect refers to the reduction in recency recall that occurs when auditory lists end with a redundant (nonrecalled) verbal item. The amount of interference obtained tends to vary directly with the acoustic similarity between the suffix and final list item. Silently presented suffixes do not generally lead to recency reduction, nor do suffixes that simply share a meaningful relation to the final list item. Successful models of the suffix effect generally assume that the auditory suffix interferes with, or "overwrites," the sensory features of the last list item, leading to poorer immediate retention (e.g., Nairne 1990a). Importantly, this modality-specific interference operates over very short time windows. If the suffix is delayed for a second or two following the last list item, the suffix effect is sharply reduced. It therefore seems clear that similarity-based interference can occur over intervals encompassed by activation-based accounts of immediate retention.

Several investigators have attempted to reduce the similarity between the target material and the distractor task to such an extent that interference between the two seems unlikely (e.g., Cowan et al. 1997a, Reitman 1974). For example, Cowan et al. (1997a) presented pairs of short tones that differed slightly in frequency and asked the subject to decide whether the second tone was higher or lower in pitch than the first. The delay between the tones ranged up to 12 s, during which the subject performed a silent distracting task (tracking the movement of a visual icon on a computer screen). The tracking task should have produced little or no interference with the tone comparison task, because the two are quite dissimilar, yet correct performance on the tone task declined significantly with delay.

This time-based performance loss seems consistent with decay theory. Note that a comparable result would not be found with verbal material because the subject could presumably rehearse the material during the retention interval. The fact that significant loss occurred, with essentially no interfering material present (and rehearsal possible, in principle), suggests that the forgetting was due uniquely to the passage of time. However, in a subsequent reanalysis of their data, Cowan et al. (2001) discovered that the relative distinctiveness of the tones on the current comparison trial influenced the amount of measured forgetting. Just as in the Turvey et al. (1970) data, performance depended on the ratio of the interpresentation interval to the current retention interval. Cowan et al. (2001) concluded that the relative distinctiveness of the tones predicted memory loss as well or better than the pure passage of time.

Summary

In his famous attack on Thorndike's "law of disuse," McGeoch (1932) used the fact that memory sometimes fails to decline, or even improves, over time as powerful evidence against the idea that information spontaneously decays. He also pointed to the fact that long-term memory for a given set of materials often depends on the specific activities that occur during the retention interval (i.e., retroactive interference). Historically, most memory theorists have accepted these arguments and rejected decay as an explanatory concept for long-term remembering. However, for short-term memory the situation is quite different—decay plays a prominent role in the standard model and is widely accepted as one of the main determinants of short-term forgetting.

As the preceding review indicates, both of the main arguments used by McGeoch (1932) to attack decay in long-term remembering apply to short-term remembering as well. Short-term memory can decrease, stay constant, or even improve over time depending on the situation. The specific activities that occur during the retention interval also matter—e.g., the more similar the distractor material is to the target material, the more forgetting one typically sees. Of course, the fact that short-term retention is influenced by interference does not rule out decay—both interference and decay could be operating in short-term memory environments. In fact, Baddeley (1990) has argued that both factors are important, with decay exerting its main influence in only the first few seconds of the retention interval. As discussed above, Baddeley & Scott (1971) attempted to replicate the findings of Keppel & Underwood (1962)—using longer lists to avoid ceiling effects—and found small amounts of forgetting on the first trial. The loss was virtually complete by 5 s, leading to the conclusion that decay occurs very rapidly and plays only a subsidiary role in the Brown-Peterson task. As Baddeley (1990) put it, ". . . something like trace decay occurs in the Peterson task, but is complete within five seconds, and is certainly not sufficient to explain the substantial forgetting that occurs in the standard paradigm" (Baddeley 1990, p. 48). The idea that decay occurs quickly is important because it fits the assumption of the standard model that activation is lost in a second or two and must be refreshed through rehearsal.

However, it is difficult to see how proponents of decay can explain all the situations in which time exerts no influence. For instance, as reviewed in the previous section, current research indicates that it is complexity, rather than actual duration, that mediates most word length effects in immediate recall. Again, long durations can actually lead to better immediate memory in some situations. Moreover, several researchers have reported that output duration is a better predictor of memory span than presentation duration, and the typical output period far exceeds the 2-s decay window (e.g., Dosher 1999, Hulme et al. 1999). Finally, as I discuss in the next section, it is possible to construct relatively successful models of immediate memory that reject decay and rely entirely on various forms of interference to explain short-term forgetting (e.g., Brown & Hulme 1995, Brown et al. 2000, Nairne 1990a, Neath & Nairne 1995). For example, Brown et al. (2001) have shown how the data of Baddeley & Scott (1971)—limited forgetting over the first few seconds of distraction—can be explained by appealing to intrasequence interference; that is, interference produced among the items in a current to-be-remembered list (see also Melton 1963). As I discuss below, one of the major advantages that models of this type have over the standard model is their assumption that remembering is cue driven.

RECONCEPTUALIZING SHORT-TERM MEMORY

In this section I discuss alternatives to the standard model. As noted throughout, one of the most important conceptual hurdles faced by activation accounts is the proposal that recovery from short-term memory is essentially independent of cueing. Kintsch and colleagues recently noted that questions about retrieval touch on "the essence of working memory because of the common assumption that information 'in' working memory is directly and effortlessly retrievable" (Kintsch et al. 1999). This notion does not fit the commonly accepted mnemonic principle that all remembering is cue dependent. As Endel Tulving put it, "Every phenomenon of episodic memory depends on both storage and retrieval conditions" (Tulving 1983)—every phenomenon, apparently, except activated items in short-term memory.

The proposal that retrieval from activated memory and retrieval from long-term memory "differ in important ways" (Cowan 1999) receives some support from microscopic analyses of reaction time patterns in short- and long-term retrieval environments. For example, several researchers have found that the retrieval dynamics for the last few items in a memory list are different than for earlier items (e.g., McElree & Dosher 1989, McElree 1998, Wickelgren et al. 1980). There are also classic studies showing that the recency effect in free recall, which is believed by many to reflect the dumping of activated information, is sensitive to different variables than the primacy and asymptotic portions of the serial position curve (e.g., Glanzer & Cunitz 1966). However, these studies do not establish how retrieval over the immediate term is actually accomplished. Remembering over the short term may still be cue driven, as in long-term memory, but the cues controlling performance may simply change across retention environments (Brown et al. 2000, Burgess & Hitch 1999). Indeed, many other studies have found that remembering over the short and long term show similarities and can often be fit with the same type of model (e.g., Brown et al. 2000, Nairne 1990b, 1991).

Cue-Driven Immediate Retention

There are many factors that have influenced the movement toward cue-driven accounts of immediate retention. One factor is the sensitivity of immediate recall to item-specific variables such as lexicality, word frequency, and concreteness. As noted above, in the standard model there is no obvious reason why lexicality or concreteness should affect the availability of an item, once activated, yet each produces large and consistent effects on immediate recall. Researchers generally refer to these effects as long-term memory contributions to immediate memory, and they are increasingly assumed to arise from a redintegration process wherein the decayed or degraded immediate memory trace is used as a cue to sample an appropriate candidate from long-term memory (e.g., Hulme et al. 1999, Nairne 1990a, Schweickert 1993).

More direct evidence for cueing effects comes from work on "release" from proactive interference (Wickens 1970). In the prototypical paradigm, subjects receive immediate memory trials in which successive lists are drawn from the same conceptual class (e.g., pieces of furniture, animals, or rhyme categories). Performance typically declines over trials, presumably because it becomes increasingly difficult to discriminate items on the current trial from conceptually similar items that occurred on previous trials. On trial N + 1 list items are drawn from a new conceptual class (e.g., moving from furniture to animals) and performance dramatically improves, a finding known as release from proactive interference. A standard interpretation of this effect is that people use the conceptual class as a "cue" to guide short-term recall; the effectiveness of the cue, in turn, is determined by the degree of cue overload (Watkins & Watkins 1975). On build-up trials, the cue becomes overloaded—that is, it predicts many items—and performance suffers; on release trials, the conceptual cue uniquely specifies the items on the current trial and performance gains are recorded.

Strong support for the cueing interpretation of proactive interference comes from studies manipulating the nature of the cues at test. For example, one can obtain release from proactive interference at the point of test, after the list has been presented, if discriminating cues are provided (Dillon & Bittner 1975, Gardiner et al. 1972). Recently, Tehan & Humphreys (1996, 1998) demonstrated the power of a distinctive cue in a paradigm requiring subjects to remember the second of two four-item trial blocks (on some trials people are asked to recall after the first block to guarantee attention to the first block). The critical manipulation varied whether items in the first block shared conceptual properties with items in the second, to-be-recalled, block—e.g., the first block might contain the items "jail-silk-orange-peach," whereas the second block contains "page-leap-carrot-witch." Note that orange and carrot share certain conceptual properties (color, edible, etc.). Tehan & Humphreys (1996) found that if subjects were asked to recall the "vegetable" from the second block, no interference was found (orange is not a vegetable). However, if the cue was "type of juice," recall of carrot was impaired if orange occurred in the first block. Thus, susceptibility to proactive interference depended entirely on the type of cue presented at test.

Evidence for cue-driven immediate retention also comes from the analysis of errors in immediate recall. Errors are usually not random, but rather follow certain rules, suggesting that position of occurrence may be an important retrieval cue. For example, when an item is recalled in the wrong serial position in immediate recall, it tends to be placed in a nearby position. One typically finds regular error gradients that drop off with distance from the original position of occurrence (e.g., Healy 1974). If lists are grouped and subjects misplace an item from one group into another group, the item tends to be placed in an identical relative serial position (e.g., Henson 1999). Finally, when people intrude an item from a previous list, it is likely to have occurred at the same serial position in the previous list (Estes 1991). These data suggest that people are not simply outputting activated items directly from short-term memory, but rather are using position of occurrence as a retrieval cue to decide what happened moments before.

Although there may be disagreements about which cues predominate, collectively these data have led many short-term memory theorists to conclude that remembering over the short term is definitely cue driven. As I discuss below, recent formal models of immediate retention tend to be cue driven, although some are hybrid models that assume that some form of "direct retrieval" from short-term memory is possible. Even accounts that closely mimic the assumptions of the standard model, such as Baddeley's working memory model (Baddeley 1986), recognize that direct retrieval cannot explain all the particulars of immediate retention. For example, it is difficult to derive the phonological similarity effect—poorer memory for lists composed of similar sounding items—from the assumptions of the standard model. Working memory proponents generally assume that memory is impaired because recall requires "... discrimination among the memory traces ... similar traces will be harder to discriminate, leading to a lower level of recall" (Baddeley 1990). The discrimination process is left unspecified, but it is in all likelihood cue driven.

Hybrid Models

The term hybrid model refers to a class of current models that retain important elements of the standard model—e.g., activation-based remembering, rehearsal, and/or decay—but assume that retrieval cues play an important role in short-term remembering as well. A detailed review of these models is beyond the scope of this article, so I provide only a few brief descriptions here. Schweickert (1993) has proposed a multinomial processing tree model that closely mimics the standard model except for the proposal of an additional redintegration stage. List presentation leads to the formation of active traces in short-term memory, which over time become degraded. Schweickert is noncommittal about the actual process controlling degradation, assuming that either decay or interference may contribute. During recall, the subject's first line of attack is a "direct readout" of the active trace, as in the standard model, which occurs successfully with probability I. If direct readout fails (with probability 1-I), an attempt is made to interpret the degraded trace through redintegration.

This second stage of Schweickert's model—the interpretation or redintegration stage—provides a vehicle for explaining many of the findings discussed above that have proven troubling for the standard model. The degraded trace becomes a cue of sorts that is interpreted by accessing long-term knowledge, particularly about language processing (see also Hulme et al. 1997). It is in the redintegration stage that the effects of lexicality, word frequency, and concreteness are presumed to occur. For example, the degraded traces for words are presumably easier to interpret than those for nonwords, leading to the lexicality effect in immediate recall. By placing the locus in the redintegration stage, the model is able to dissociate so-called long-term memory influences, such as lexicality and frequency, from the effects of articulation rate, which are presumed to primarily affect the direct readout stage.

One of the advantages of the multinomial tree model is its clear predictions about how different variables should interact. Word length and lexicality are assumed to selectively influence different stages of the recall process: Word length affects the probability of trace degradation (for the same reasons described by the standard model), and lexicality affects the probability of trace interpretation. As a result, these factors, when combined factorially, are expected to produce an underadditive interaction in correct recall. Just this pattern has been obtained in relevant studies—i.e., the size of the word length effect is smaller for words than nonwords (Besner & Davelaar 1982). Thus, by adding the cue-drive, redintegration stage, the model can handle findings that seem to falsify the standard model. At the same time, Schweickert's model is silent about the mechanisms or processes that actually underlie the act of recall. For example, how is "direct readout" accomplished? Is it cue driven, or is the information available by virtue of its activation alone? It may turn out that the model's assumptions about the need for separate and independent routes to recall are well justified, but that its implied assumptions about decay and direct readout are not.

The "start-end" model proposed recently by Henson (1998) is another example of a hybrid model. Henson (1998) assumes that items are coded relative to the beginning and end of a sequence. Essentially, items are associated with position codes, and these position codes are then used as cues to drive recall. In this sense, the deep structure of the model differs significantly from the standard model—items are not activated and immediately accessible, but rather are selected for recall on the basis of a cued-driven response competition process. Through simulations, Henson (1998) has shown how this machinery can be used to explain much of the phenomena of immediate and delayed retention of serial order, such as serial position curves, error distributions, grouping effects, and modality effects. However, the model is unable to explain word length effects and phonological similarity effects without additional assumptions.

To handle time- and item-based effects, Henson (1998) appeals to the main assumptions of the standard model. In particular, he assumes that "each presentation and rehearsal of an item activates its phonological representation to a fixed amount that subsequently undergoes exponential decay" (Henson 1998, p. 106). The activation process increases the probability of recall directly, by bringing an item closer to its recall threshold, although it can increase the chances of phonological confusions as well (thus explaining the phonological similarity effect). Word length effects are explained by appealing to the dynamics of rehearsal: Long words tend to receive less activation, because they cannot be rehearsed as efficiently, and recall suffers as a result. Thus, as in the standard model, activation plays a role as a mnemonic property that is independent of any cueing mechanism. For the reasons described throughout this article, these assumptions of the start-end model are not well supported by the data.

Unitary Models

The two models just described are examples of current hybrid models. There have been other efforts to combine elements of the standard model with some kind of cue-based mechanism (e.g., Burgess & Hitch 1999), but I turn my attention now to models that contain virtually no assumptions in common with the standard model. These models assume no direct connection between activation level and memory success, propose little or no role for rehearsal, and reject the concept of decay in favor of item-based interference. I refer to these models as unitary models because they also assume similar processes for short- and long-term remembering (what differs is the retrieval cues in effect). Although not reviewed here, there is considerable evidence suggesting that short- and long-term memory often follow similar rules (e.g., Brown et al. 2000, Melton 1963, Nairne 1991, Crowder & Neath 1991).

The feature model (Nairne 1988, 1990a) is an example of a unitary model. All remembering over the short term is cue driven in this model, based on an analysis of residual processing records. What sits in short-term memory is not an activated "item" that can be directly retrieved, but rather a constellation of activated cues that the subject uses to reconstruct what happened moments before. In most cases, these "cues" consist of remnants of prior processing records—e.g., records of just-presented list items that have become degraded through interference. A recall candidate is selected from long-term memory, based on a similarity-driven sampling rule (a choice rule), in a manner resembling that employed by context models of categorization (e.g., Nosofsky 1986). Importantly, it is not the match between the to-be-interpreted cues and long-term memory candidates that turns out to be important; instead, what determines performance is how well cues uniquely specify one or more of the target items (see Nairne 2001).

Short-term forgetting in this model occurs because the available cues become poor predictors of the target items. Processing records are overwritten by subsequently occurring material (as a function of similarity), making it more difficult to interpret the records correctly. Basing interference on similarity enables the model to explain why performance can depend on the specific activities that occur during the retention interval, and it helps explain benchmark phenomena such as the phonological similarity effect as well. Overall, increasing the similarity among list items tends to reduce the predictive value of common features; any given residual cue tends to be predictive of several target items (it becomes overloaded), which lowers the chances of remembering a given target item in its correct position. Performance declines with increasing list length for a very similar reason. Cues in short-term memory are effective only to the extent that they are distinctive—that is, they uniquely predict target items.

Simulations of the feature model have been applied to most of the phenomena of immediate memory with success (see Nairne 1990a; Neath & Nairne 1995; Neath 1998, 1999). For example, Neath & Nairne (1995) demonstrated how the feature model can handle the nearly linear relation between articulation rate and span, as well as the interactions found between word length, modality of presentation, and articulatory suppression. Neath (2000) has recently shown how the model can account for the rather complex effects of irrelevant speech on short-term memory performance (see also Surprenant et al. 2000). For the present purposes, the important point is that these effects are simulated without any recourse to either rehearsal or decay. No special mnemonic properties are assigned to the activated contents of short-term memory—they merely act as retrieval cues that may or may not lead to correct retrieval. People forget with an increasing delay because retrieval cues change with time, not because of spontaneous decay. Such a cue-based model can easily handle the fact that memory may remain constant, or even improve, as time passes. Provided that distinctive cues are not interfered with by subsequent material, or are reinstated in some way, immediate memory should remain at high levels.

Cue-dependent forgetting is also central to the OSCAR model proposed recently by Brown et al. (2000) and to the phonological loop model of Burgess & Hitch (1999). Both models assume that associations are formed between items, represented as feature vectors, and dynamic context signals. Context signals are represented as sets of oscillators that change systematically, at idiosyncratic rates, over time. At recall, the context signal is reset—the oscillators are "rewound"—and each successive state is used as a cue to recall associated items. The details of the storage assumptions are beyond the scope of this article, but the product of the cueing process tends to be a blurry record, requiring a cleanup or redintegration stage. It is during the redintegration stage, which in OSCAR is driven by a similarity-based comparison process, that many important recall phenomena, such as the phonological similarity effect, primarily arise (see also Lewandowsky 1999).

The major difference between OSCAR and the loop model (Burgess & Hitch 1999) is the failure of OSCAR to incorporate the assumptions of the standard model. The phonological loop model is actually a hybrid model—recall is cue driven, but rehearsal and decay are used to account for a number of recall effects. OSCAR essentially rejects the concept of decay and assumes that the same principles apply over a wide range of time scales. As a result, OSCAR has some trouble accounting for rehearsal-based effects, such as the word length effect, which can be easily handled by the loop account. However, as we have seen above, the evidence for rehearsal and decay is mixed at best, and it is possible to explain time-based effects in alternative ways (e.g., through interference; see Neath & Nairne 1995, Brown et al. 2000). Moreover, unitary models such as OSCAR are potentially capable of explaining a wider range of data, over both short- and long-term time scales, because they do not rely on the mnemonic properties of activation. Because retrieval in the standard model is cue independent, relying solely on activation, it must differ in a fundamental way from the retrieval processes governing long-term retention.

SUMMARY AND CONCLUSIONS

The preceding two sections provide a brief and selective review of some recent formal models covering immediate retention. In each of these models remembering over the short term is assumed to be primarily cue driven, although rehearsal and decay—the two main assumptions of the standard model—contribute to performance in some cases. Focusing on cue-driven processes, as opposed to the direct retrieval of activated "items," offers a number of advantages. First, and perhaps most importantly, it lays the groundwork for a truly unified account of remembering. Virtually all researchers recognize that long-term remembering is cue driven; acknowledging that short-term remembering is cue driven as well helps explain why short- and long-term retention often show similarities, and it releases the theorist from the unreasonable assumption that activated traces have special properties outside of particular retrieval environments.

More concretely, cue-driven accounts easily handle the item-specific long-term memory influences that characterize remembering over the short term. As we have seen, immediate retention is influenced by a number of variables, such as lexicality, word frequency, and concreteness, that are likely to affect one's ability to sample an appropriate recall candidate from long-term memory. In addition, it is easy to see how memory performance could decrease, stay constant, or even improve over time depending on the available constellation of retrieval cues. In unitary models, such as the feature model, short-term memory is simply conceived as a repository for cues that are used for reconstructing the immediate past. No items are stored—only feature-based cues that are, by themselves, not recallable. Such a conceptualization is vastly different from intuitive notions about activated items "sitting" in short-term memory awaiting direct recall.

This kind of view also has no trouble handling the possibility that inhibitory effects will occur in immediate retention. It is almost certainly the case that item accessibility can be lowered over the short term—that is, you become less likely to remember an item—and there is no easy way to represent inhibition in the standard model. Rather than assuming an item is in some special state of inaccessibility, one can assume that there are simply cue constellations that reduce the likelihood of recovering an item as a possible response (response suppression mechanisms may also be at work in some instances). Note that inhibition conceived in this way is not a special state of the item; it is simply a byproduct of the particular cue constellation that happens to be driving memory.

What role then should the standard model play in our efforts to understand how we remember over the short term? The juggler metaphor certainly has heuristic value, and it provides a nice organizational rubric for a variety of immediate memory phenomena. However, even as a heuristic, the standard model is misleading. It leads one to the conclusion that forgetting rates are fixed, like gravity, rather than variable, as much of the data suggest. It also suggests that the main vehicle for short-term storage is rehearsal when, in fact, much of the variability in immediate retention turns out to be independent of rehearsal. Finally, it leads one to the conclusion that remembering is a direct byproduct of activation. Whereas it may be reasonable to propose activation in the brain, it is not activation per se that predicts performance. It is the interpretation of that activation, through a cue-driven retrieval process, that explains how we remember over the short term.

ACKNOWLEDGMENTS

I thank Gordon Brown, Ian Neath, and Gerry Tehan for many thoughtful discussions about the ideas presented in this review. Thanks are also due to Nelson Cowan and Margaret Wilson for commenting on an earlier version of the manuscript.

LITERATURE CITED

Atkinson RC, Shiffrin RM. 1968. Human memory: a proposed system and its control processes. In *The Psychology of Learning and Motivation*, ed. KW Spence, JT Spence, 2:89–105. New York: Academic

Avons SE, Wright KL, Pammer K. 1994. The word-length effect in probed and serial recall. *Q. J. Exp. Psychol.* 20:249–64

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S. 1996. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci.* 7:25–31

Baddeley AD. 1976. *The Psychology of Memory*. New York: Basic Books

Baddeley AD. 1986. *Working Memory*. Oxford: Oxford Univ. Press

Baddeley AD. 1990. *Human Memory: Theory and Practice*. Oxford: Oxford Univ. Press

Baddeley AD. 1992. Working memory. *Science* 255:556–59

Baddeley AD. 2000. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4:417–23

Baddeley AD, Hitch G. 1974. Working memory. In *The Psychology of Learning and Motivation*, ed. GH Bower, 8:647–67. New York: Academic

Baddeley AD, Lewis VJ, Vallar G. 1984. Exploring the articulatory loop. *Q. J. Exp. Psychol.* 36:233–52

Baddeley AD, Scott D. 1971. Short term forgetting in the absence of proactive interference *Q. J. Exp. Psychol.* 23:275–83

Baddeley AD, Thomson N, Buchanan M. 1975. Word length and the structure of short-term memory. *J. Verbal Learn. Verbal Behav.* 14:575–89

Baddeley AD, Wilson B. 1985. Phonological coding and short-term memory in patients without speech. *J. Mem. Lang.* 24:490–502

Besner D, Davelaar E. 1982. Basic processes in reading: two phonological codes. *Can. J. Psychol.* 36:701–11

Bishop DVM, Robson J. 1989. Unimpaired short-term memory and rhyme judgment in congenitally speechless individuals: implications for the notion of 'articulatory coding'. *Q. J. Exp. Psychol.* 41A:123–41

Bjork RA. 1975. Short-term storage: the output of a central processor. In *Cognitive Theory*, ed. F Restle, RM Shiffrin, NJ Castellan, HR Lindman, DB Pisoni, 1:151–72. Hillsdale, NJ: Erlbaum

Bjork RA. 2001. Recency and recovery in human memory. See Roediger et al. 2001, pp. 211–32

Bourassa D, Besner D. 1994. Beyond the articulatory loop: a semantic contribution to serial order recall of subspan lists. *Psychol. Bull. Rev.* 1:122–25

Brown GDA, Hulme C. 1995. Modeling item length effects in memory span: no rehearsal needed? *J. Mem. Lang.* 34:594–21

Brown GDA, Neath I, Chater N. 2001. SIMPLE: a local distinctiveness model of scale—invariant memory and perceptual identification. Unpublished manuscript

Brown GDA, Preece T, Hulme C. 2000. Oscillator-based memory for serial order. *Psychol. Rev.* 107:127–81

Brown J. 1958. Some tests of the decay theory of immediate memory. *Q. J. Exp. Psychol.* 10:12–21

Burgess N, Hitch GJ. 1999. Memory for serial order: a network model of the phonological loop and its timing. *Psychol. Rev.* 106:551–81

Caplan D, Rochon E, Waters GS. 1992. Articulatory and phonological determinants of word-length effects in span tasks. *Q. J. Exp. Psychol.* 45A:177–92

Caza N, Belleville S. 1999. Semantic contribution to immediate serial recall using an unlimited set of items: evidence for a multilevel capacity view of short-term memory. *Int. J. Psychol.* 34:334–38

Coltheart V, Langdon R. 1998. Recall of short word lists presented visually at fast rates: effects of phonological similarity and word length. *Mem. Cogn.* 26:330–42

Cowan N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychol. Bull.* 104:163–91

Cowan N. 1995. *Attention and Memory: An Integrated Framework*. New York: Oxford Univ. Press

Cowan N. 1999. The differential maturation of two processing rates related to digit span. *J. Exp. Child Psychol.* 72:193–209

Cowan N, Day L, Saults JS, Keller TA, Johnson Y, Flores L. 1992. The role of verbal output time in the effects of word length on immediate memory. *J. Mem. Lang.* 31:1–17

Cowan N, Nugent LD, Elliot EM, Geer T. 2000. Is there a temporal basis of the word length effect? A response to Service (1998). *Q. J. Exp. Psychol.* 53A:647–60

Cowan N, Saults JS, Nugent LD. 1997a. The role of absolute and relative amounts of time in forgetting within immediate memory: the case of tone pitch comparisons. *Psychol. Bull. Rev.* 4:393–97

Cowan N, Saults JS, Nugent LD. 2001. The ravages of absolute and relative amounts of time on memory. See Roediger et al. 2001, pp. 315–30

Cowan N, Wood NL, Nugent LD, Treisman M. 1997b. There are two word length effects in verbal short-term memory: opposed effects of duration and complexity. *Psychol. Sci.* 8:290–95

Cowan N, Wood NL, Wood PK, Keller TA, Nugent LD, Keller CV. 1998. Two separate verbal processing rates contributing to short-term memory span. *J. Exp. Psychol.: Gen.* 127:141–60

Crowder RG. 1976. *Principles of Learning and Memory*. Hillsdale, NJ: Erlbaum

Crowder RG, Neath I. 1991. The microscope metaphor in human memory. See Hockley & Lewandowsky 1991, pp. 111–25

Dale HC, Gregory M. 1966. Evidence of semantic coding in short-term memory. *Psychol. Sci.* 5:153–54

Dillon RF, Bittner LA. 1975. Analysis of retrieval cues and release from proactive inhibition. *J. Verbal Learn. Verbal Behav.* 14:616–22

Dosher BA. 1999. Item interference and time delays in working memory: immediate serial recall. *Int. J. Psychol.* 34:276–84

Elliott LA, Strawhorn RJ. 1976. Interference in short-term memory from vocalization: aural versus visual modality differences. *J. Exp. Psychol.: Hum. Learn. Mem.* 2:705–11

Ellis NC, Hennelly RA. 1980. A bilingual word-length effect: implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *Br. J. Psychol.* 71:43–52

Engle RW, Kane MJ, Tuholski SW. 1999. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. See Miyake & Shaw 1999, pp. 102–34

Estes WK. 1991. On types of item coding and source of recall in short-term memory. See Hockley & Lewandowsky 1991, pp. 155–74

Gardiner JM, Craik FIM, Birtwistle J. 1972. Retrieval cues and release from proactive inhibition. *J. Verbal Learn. Verbal Behav.* 11:778–83

Gathercole SE, Hitch GJ. 1993. Developmental changes in short-term memory: a revised working memory perspective. In *Theories of Memory*, ed. AF Collins, SE Gathercole, MA Conway, PE Morris, pp. 189–210. Hove, UK: Erlbaum

Glanzer M, Cunitz AR. 1966. Two storage mechanisms in free recall. *J. Verbal Learn. Verbal Behav.* 5:351–60

Gorfein DS. 1987. Explaining context effects on short-term memory. In *Memory and Learning: The Ebbinghaus Centennial Conference*, ed. DS Gorfein, RR Hoffman. Hillsdale, NJ: Erlbaum

Greene RL. 1992. *Human Memory: Paradigms and Paradoxes*. Hillsdale, NJ: Erlbaum

Greene RL. 1996. The influence of experimental design: the example of the Brown-Peterson paradigm. *Can. J. Exp. Psychol.* 50:240–42

Healy AF. 1974. Separating item from order information in short-term memory. *J. Verbal Learn. Verbal Behav.* 13:644–55

Henry LA. 1991. Development of auditory memory span: the role of rehearsal. *Br. J. Dev. Psychol.* 9:493–511

Henson RNA. 1998. Short-term memory for serial order: the Start-End Model. *Cogn. Psychol.* 36:73–37

Henson RNA. 1999. Positional information in short-term memory: relative or absolute? *Mem. Cogn.* 27:915–27

Hitch GJ, Halliday MS. 1983. Working memory in children. *Philos. Trans. R. Soc. London Ser. B* 302:324–40

Hockley WE, Lewandowsky S, eds. 1991. *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock.* Hillsdale, NJ: Erlbaum

Hulme C, Maughan S, Brown GDA. 1991. Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to short-term memory span. *J. Mem. Lang.* 30:685–701

Hulme C, Newton P, Cowan N, Stuart G, Brown G. 1999. Think before you speak: pauses, memory search, and trace redintegration processes in verbal memory span. *J. Exp. Psychol.: Learn. Mem. Cogn.* 25:447–63

Hulme C, Roodenrys S, Schweickert R, Brown GDA, Martin S, Stuart G. 1997. Word frequency effects on short-term memory tasks: evidence for a multinomial processing tree model of immediate serial recall. *J. Exp. Psychol.: Learn. Mem. Cogn.* 23:1217–32

Hulme C, Thomson N, Muir C, Lawrence A. 1984. Speech rate and the development of short-term memory span. *J. Exp. Child Psychol.* 38:241–53

Hulme C, Tordoff V. 1989. Working memory development: the effects of speech rate word length and acoustic similarity on serial recall. *J. Exp. Child Psychol.* 47:72–87

Keppel G, Underwood BJ. 1962. Proactive inhibition in short-term retention of single items. *J. Verbal Learn. Verbal Behav.* 1:153–61

Kintsch W, Healy AF, Hegarty M, Pennington BF, Salthouse TA. 1999. Models of working memory: eight questions and some general answers. See Miyake & Shah 1999, pp. 412–41

Kintsch W, VanDijk TA. 1978. Toward a model of text comprehension and production. *Psychol. Rev.* 85:363–94

Landauer TK. 1962. Rate of implicit speech. *Percept. Mot. Skills* 15:646

LaPointe LB, Engle RW. 1990. Simple and

complex spans as measures of working memory capacity. *J. Exp. Psychol.: Learn. Mem. Cogn.* 16:1118–33

Lewandowsky S. 1999. Redintegration and response suppression in serial recall: a dynamic network model. *Int. J. Psychol.* 34:434–46

Lovatt P, Avons SE, Masterson J. 2000. The word-length effect and disyllabic words. *Q. J. Exp. Psychol.* 53A:1–22

Lovett MC, Reder LM, Lebiere C. 1999. Modeling working memory in a unified architecture. See Miyake & Shah 1999, pp. 135–82

McElree B. 1998. Attended and non-attended states in working memory: accessing categorized structures. *J. Mem. Lang.* 38:225–52

McElree B, Dosher BA. 1989. Serial position and set size in short-term memory: the time course of recognition. *J. Exp. Psychol.: Gen.* 118:346–73

McGeoch JA. 1932. Forgetting and the law of disuse. *Psychol. Rev.* 39:352–70

Melton AW. 1963. Implications of short-term memory for a general theory of memory. *J. Verbal Learn. Verbal Behav.* 2:1–21

Miller GA. 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63:81–97

Miyake A, Shah P. 1999. *Models of Working Memory*. Cambridge: Cambridge Univ. Press

Nairne JS. 1988. A framework for interpreting recency effects in immediate serial recall. *Mem. Cogn.* 16:343–52

Nairne JS. 1990a. A feature model of immediate memory. *Mem. Cogn.* 18:251–69

Nairne JS. 1990b. Similarity and long-term memory for order. *J. Mem. Lang.* 29:733–46

Nairne JS. 1991. Positional uncertainty in long-term memory. *Mem. Cogn.* 19:332–40

Nairne JS. 1996. Short-term/working memory. In *Memory*, ed. EL Bjork, RA Bjork, pp. 101–26. New York: Academic

Nairne JS. 2001. A functional analysis of primary memory. See Roediger et al. 2001, pp. 283–96

Nairne JS, Kelley MR. 1999. Reversing the phonological similarity effect. *Mem. Cogn.* 27:45–53

Nairne JS, Neath I, Serra M. 1997. Proactive interference plays a role in the word-length effect. *Psychol. Bull. Rev.* 4:541–45

Nairne JS, Whiteman HL, Kelley MR. 1999. Short-term forgetting of order under conditions of reduced interference. *Q. J. Exp. Psychol.* 52A:241–51

Naveh-Benjamin M, Ayres TJ. 1986. Digit span, reading rate, and linguistic relativity. *Q. J. Exp. Psychol.* 38A:739–51

Neath I. 1998. *Human Memory: An Introduction to Research Data and Theory*. Pacific Grove, CA: Brooks/Cole

Neath I. 1999. Modelling the disruptive effects of irrelevant speech on order information. *Int. J. Psychol.* 34:416–24

Neath I. 2000. Modeling the effects of irrelevant speech on memory. *Psychol. Bull. Rev.* 7:403–23

Neath I, Knoedler AJ. 1994. Distinctiveness and serial position effects in recognition and sentence processing. *J. Mem. Lang.* 33:776–95

Neath I, Nairne JS. 1995. Word-length effects in immediate memory: overwriting trace-decay theory. *Psychol. Bull. Rev.* 2:429–41

Nicolson R. 1981. The relationship between memory span and processing speed. In *Intelligence and Learning*, ed. M Friedman, JP Das, N O'Conner. New York: Plenum

Nosofsky RM. 1986. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol.: Gen.* 115:39–57

Osgood CE. 1953. *Method and Theory in Experimental Psychology*. New York: Oxford Univ. Press

Page MPA, Norris D. 1998. The primacy model: a new model of immediate serial recall. *Psychol. Rev.* 105:761–81

Peterson LR, Peterson MJ. 1959. Short-term retention of individual verbal items. *J. Exp. Psychol.* 58:193–98

Posner MI, Konick AW. 1966. On the role of interference in short-term retention. *J. Exp. Psychol.* 72:221–31

Proctor RW, Fagnani CA. 1978. Effects of distractor-stimulus modality in the Brown-Peterson distractor task. *J. Exp. Psychol.: Hum. Learn. Mem.* 4:676–84

Reitman JS. 1974. Without surreptitious rehearsal information in short-term memory decays. *J. Verbal Learn. Verbal Behav.* 13:365–77

Roediger HL III, Nairne JS, Neath I, Suprenant AM, eds. 2001. *The Nature of Remembering: Essays in Honor of Robert G. Crowder*. Washington, DC: Am. Psychol. Assoc.

Roodenrys S, Hulme C, Alban J, Ellis AW, Brown GD. 1994. Effects of word frequency and age of acquisition on short-term memory span. *Mem. Cogn.* 22:695–701

Schweickert R. 1993. A multinomial processing tree model for degradation and redintegration in immediate recall. *Mem. Cogn.* 21:167–75

Schweickert R, Boruff B. 1986. Short-term memory capacity: magic number or magic spell? *J. Exp. Psychol.: Learn. Mem. Cogn.* 12:419–25

Schweickert R, Guentert L, Hersberger L. 1990. Phonological similarity, pronunciation rate, and memory span. *Psychol. Sci.* 1:74–77

Service E. 1998. The effect of word length on immediate serial recall depends on phonological complexity not articulatory duration. *Q. J. Exp. Psychol.* 51A:283–304

Shiffrin RM. 1999. 30 years of memory. In *On Human Memory: Evolution Progress and Reflections of the 30th Anniversary of the Atkinson-Shiffrin Model*, ed. C Izawa, pp. 17–33. Mahwah, NJ: Erlbaum

Smith EE, Jonides J. 1997. Working memory: a view from neuroimaging. *Cogn. Psychol.* 33:5–42

Surprenant AM, LeCompte DC, Neath I. 2000. Manipulations of irrelevant information: suffix effects with articulatory suppression and irrelevant speech. *Q. J. Exp. Psychol.* 53A:325–48

Tan L, Ward G. 2000. A recency-based account of the primacy effect in free recall. *J. Exp. Psychol.: Learn. Mem. Cogn.* 26:1589–625

Tehan G, Fallon AB. 1999. A connectionist model of short-term cued recall. In *Prospectives in Cognitive Science*, ed. J Wiles, T Dartnell, pp. 221–37. Stamford, CT: Ablex

Tehan G, Humphreys MS. 1988. Articulatory loop explanations of memory span and pronunciation rate correspondence: A cautionary note. *Bull. Psychon. Soc.* 26:293–96

Tehan G, Humphreys MS. 1996. Cuing effects in short-term recall. *Mem. Cogn.* 24:719–32

Tehan G, Humphreys MS. 1998. Creating proactive interference in immediate recall: building a dog from a dart, a mop and a fig. *Mem. Cogn.* 26:477–89

Tehan G, Lalor DM. 2000. Individual differences in memory span: the contribution of rehearsal access to lexical memory and output speed. *Q. J. Exp. Psychol.* 53A:1012–38

Tulving E. 1983. *Elements of Episodic Memory*. New York: Oxford Univ. Press

Turvey MT, Brick P, Osborn J. 1970. Proactive interference in short-term memory as a function of prior-item retention interval. *Q. J. Exp. Psychol.* 22:142–47

Vallar G, Cappa SF. 1987. Articulation and verbal short-term memory: evidence from Anarthria. *Cogn. Neurosci.* 4:55–78

Walker I, Hulme C. 1999. Concrete words are easier to recall than abstract: evidence for a semantic contribution to short-term serial recall. *J. Exp. Psychol.: Learn. Mem. Cogn.* 25:1256–71

Watkins OC, Watkins MJ. 1975. Buildup of proactive inhibition as a cue-overload effect. *J. Exp. Psychol.: Human Learn. Mem.* 104:442–52

Wickelgren WA. 1965. Acoustic similarity and retroactive interference in short-term memory. *J. Verbal Learn. Verbal Behav.* 4:53–61

Wickelgren WA, Corbett AT, Dosher BA. 1980. Priming and retrieval from short-term memory: a speed-accuracy tradeoff analysis. *J. Verbal Learn. Verbal Behav.* 19:387–404

Wickens DD. 1970. Encoding categories of words: an empirical approach to meaning. *Psychol. Rev.* 77:771–15

Wilson M. 2001. The case for sensorimotor coding in working memory. *Psychol. Bull. Rev.* 8:44–57

Wilson M, Emmorey K. 1998. A “word length effect” for sign language: further evidence on the role of language in structuring working memory. *Mem. Cogn.* 26:584–90

Wright AA, Santiago HC, Sands SF, Kendrick DF, Cook RG. 1985. Memory processing of serial lists by pigeons, monkeys and people. *Science* 229:287–89

Annual Review of Psychology
Volume 53, 2002

CONTENTS

Analysis of Latent Variables

Indexes

Errata

Online log of corrections to the Annual Review of Psychology corrections:
Steven Regeser López and Peter J. Guarnaccia
Cultural Psychopathology: Uncovering the Social World of Mental Illness
Annu. Rev. Psychol. 2000, Vol. 51: 571–598.
http://psych.annualreviews.org/errata.shtml