Journal of Experimental Psychology: General
1989, Vol. 118, No. 4, 346-373

0096-3445/89/$00.75

Serial Position and Set Size in Short-Term Memory: The Time Course of Recognition

Brian McElree and Barbara Anne Dosher
Columbia University

Subjects viewed sequentially presented lists of 3–6 words, which were followed by a recognition probe. Memory retrieval speed (dynamics) and strength were measured in an interruption speed–accuracy trade-off (SAT) procedure and a collateral reaction time (RT) procedure. In SAT, item strengths depended on serial position, but only two retrieval speeds were observed: a fast rate for the last item in the study list (a case of immediate repetition between study and test) and a slow rate for all other items that was independent of serial position and set size. Serial-position-dependent strengths and set-size-dependent criterion shifts accounted for standard RT patterns that have been taken as evidence for serial scanning in short-term memory.

Summary of Experiments

We examined retrieval in immediate memory for short lists, using speed–accuracy trade-off (SAT) and comparable reaction time (RT) paradigms. We replicated Monsell's (1978) demonstration of strong recency or serial position effects on RT, which average to produce linear, parallel set size functions for both positive and negative trials. In SAT, memory strength or probability (indexed by asymptotic accuracy) and retrieval speed (indexed by rate and intercept) were separately estimated for full retrieval functions. An SAT experiment showed that the serial position effects in RT are paralleled by analogous effects on asymptotic memory strength. The dynamics of retrieval (rate and intercept) were equal for all serial positions except the most recent and were independent of set size. There was a large speeding in retrieval when the test item was the same as the last list member (immediate repetition), which is a replication of a similar finding by Wickelgren, Corbett, and Dosher (1980) in long lists. Thus observed SAT rate differences between set sizes when serial position data were pooled reflected the relative proportion of immediate repetitions in different set sizes. Set size per se had no effect on retrieval speed, although different serial position mixes resulted in different asymptotic accuracy. The immediate repetition effect did not depend on a physical match between the last list element and the test. All results were replicated in a second SAT experiment, in which we presented list items in lowercase letters and test items in uppercase letters. In that experiment we also examined the effect of recency on negative trials. Recent negatives (items presented in the immediately previous memory list) yielded higher false alarms than did distant negatives, especially early in retrieval. Serial-scanning models, which predict slower rate and/or longer intercepts for items in larger memory sets, could be rejected. Random-walk models of memory such as Ratcliff's (1978) can accommodate our findings when modified to decouple the search set from the memory set. However, we found that Wickelgren and Norman's (1966) direct-access strength model gave a direct and principled account of the pattern of asymptotic differences. When coupled with a simple retrieval mechanism and immediate repetition, it adequately accounted for retrieval from short-term lists in both positive and negative trials.

Introduction

A classic issue in memory research concerns the nature of retrieval in immediate or short-term recognition memory. Although important in its own right, this issue has implications for our overall understanding of human memory. For example, a distinction between short- and long-term memory can be motivated partly by the degree to which retrieval processes differ for immediately and less recently presented material. Short-term retrieval is also an important component process in a number of more complex cognitive behaviors such as language perception and production, reasoning, and problem solving.

Much of our current understanding of short-term retrieval has come from studies involving the probe recognition task, which was introduced by Sternberg (1966). In this task, subjects study a list of usually one to six items, which is followed immediately by a test item. In Sternberg's original work, subjects judged the list status of the probe, either new or old, as rapidly as possible while minimizing errors. Typically, the measure of primary interest is reaction time (RT) as a function of various experimental factors that serve to highlight important features of the retrieval process.

In this set of experiments, we used the interruption or cued-response speed–accuracy trade-off (SAT) methodology to examine response speed and accuracy across the complete time course of short-term item recognition. This method enables one to control total processing time by requiring subjects to respond immediately after a response cue presented at various times during retrieval. Measuring accuracy under a range of

This research was partly supported by National Science Foundation grant BNS-83-12148 to Barbara Anne Dosher.

We thank Sam Glucksberg, John Kounios, Ben Murdock, Roger Ratcliff, Saul Sternberg, and Ted Wright for helpful comments or discussion.

Correspondence concerning this article should be addressed to Barbara Anne Dosher, Box 28, Schermerhorn Hall, Department of Psychology, Columbia University, New York, New York 10027.

controlled processing times allows for the independent assessment of retrieval speed and ultimate accuracy. We examined the effects of two fundamental variables on retrieval, namely, item recency and the size of the memory set. *Recency* refers to the delay between test and the last occurrence of the tested item. Average recency covaries with the size of the memory set, and so we separately examined items in each list position. Our primary objective was to determine the relative impact of recency and set size on retrieval speed and overall accuracy. Our findings place strong constraints on the proper model for short-term item retrieval.

Extant Reaction Time Theories and Data

In a series of pioneering studies, Sternberg (1966, 1969) first documented the effect of set size on recognition RT and its implications for the nature of immediate memory retrieval. RTs for positive and negative responses were both linearly related to memory set size, with approximately equal slopes. In contrast to the findings in other recognition paradigms (e.g., study–test), the serial position of the positive test probes did not appear to affect response latency. From these data, Sternberg drew two notable inferences. The linear set size function implied a serial comparison of the probe with elements of the memory set, whereby each new item in the memory set added an additional comparison process and hence a constant increment in response time. Equal slopes of positive and negative tests suggested that the comparison stage was exhaustive rather than self-terminating on finding a match. In a self-terminating search, comparisons involving a positive probe should terminate on average in half the number of comparisons needed for negative probes, resulting in a 2:1 slope relation. Moreover, if the memory set is scanned in the approximate order of study, then a self-terminating scan model predicts marked effects of the serial position of the probe.

We first briefly outline some empirical results that have subsequently emerged and proved problematic for the exhaustive serial scan model. Following this is a brief discussion of some well-known alternative proposals. Next, the interruption speed–accuracy trade-off (SAT) variant of Sternberg's paradigm is introduced, and predictions from broad classes of models concerning set size and recency functions are contrasted and subsequently tested.

Empirical Issues

In research after Sternberg's (1966, 1969) original reports, investigators have questioned the generality and robustness of linear parallel set size functions (for a review, see Corballis, 1975, or Sternberg, 1975). Briggs (1974), for example, reviewed a number of studies and found that about 62% were fit better by a logarithmic than by a linear set size function. Other researchers have found different slopes for negative and positive responses (e.g., Kristofferson, 1972). Yet, as Sternberg (1975) pointed out, many of these discrepancies might be attributed to potentially significant procedural variations, and on the whole, linear parallel functions remain an accurate generalization of the effects of set size on RT.

The plausibility of a serial exhaustive scan, however, has been substantially weakened by examinations of the properties of RT distributions. Schneider and Shiffrin (1977) reported that variance increases more for positive than for negative responses with larger set sizes. An exhaustive scan model predicts that RT variances, like mean RTs, should increase at an equal rate for both responses. An analysis of RT distributions for various set sizes was reported by Hockley and Corballis (1982; see also Hockley, 1984). They used the ex-Gaussian distribution advocated by Ratcliff and Murdock (1976) to describe the shape of the RT distribution. Hockley and Corballis found that the increase in variance with larger set sizes primarily reflects increases in the positive skew of the distributions, coupled with a relatively minor effect on the leading edge or minimum reaction times; that is, whereas all times may be slightly slowed by an increase in the memory set, only a proportion of the longer response times are strongly affected (Ratcliff & Murdock, 1976). A serial exhaustive scan model predicts, to the contrary, that increasing set size should shift the entire distribution in the direction of longer times.

Further problems for the exhaustive scan model have emerged from a number of studies in which effects of recency on both positive and negative responses were found. When the test probe is drawn from the study list and rehearsal is either minimal or constrained to preserve study order, then recency may simply be related to the list position of the item during study (Baddeley & Ecob, 1973; Clifton & Birenbaum, 1970; Juola & Atkinson, 1971). Contrary to early reports, an impressive number of researchers have now reported typical positive serial position functions (Aube & Murdock, 1974; Clifton & Birenbaum, 1970; Corballis, 1967; Corballis, Kirby, & Miller, 1972; Forrin & Cunningham, 1973; Monsell, 1978; Morin, De Rosa, & Shultz, 1967; Murdock & Franklin, 1984; Ratcliff, 1978). Except for a small primacy effect, these studies generally showed a monotonic increase in RT with the decreasing recency of the positive probe. Moreover, in two particularly well-controlled experiments, Monsell (1978) demonstrated that the serial position functions for set sizes of one to five superimpose in all positions save the primary one, once serial position is defined in terms of recency or the number of items intervening between study and test.

On its own, the demonstration of serial position effects in positive RT would not be sufficient to reject the notion of a serial exhaustive scan because such effects may be localized to stages other than scanning. In fact, Sternberg (1975) suggested that recency may simply serve to speed the encoding of the test probe. Such a stance, however, is inconsistent with the fact that recency of negative probes affects RT in the opposite direction: Atkinson, Herrmann, and Westcourt (1974) and Monsell (1978), among others, report that response time is longer and accuracy is lower the more recently a negative probe was presented as a member of a former study set. If recency were a simple encoding phenomenon, then negative responses, like positive ones, would be facilitated by recency rather than inhibited. In fact, these results show that the mechanism underlying recognition judgments is responsive to properties of items outside the experimenter-defined memory set—a finding that is not easily handled by simple serial scan models (Ratcliff, 1978).

Alternative Models

Analysis of RT distributions and strong recency effects challenge the validity of a serial exhaustive scan model. These results are more consistent with self-terminating scan or direct-access comparison processes. Several examples are briefly outlined in the following discussion, not with the intention of presenting a complete survey of extant models, but rather to illustrate broad classes of retrieval mechanisms.

With one exception, serial self-terminating mechanisms have proved difficult to reconcile with the basic demonstration of linear parallel functions. The exception is, of course, Theios and colleagues' (Theios, 1973; Theios, Smith, Haviland, Traupman, & Moy, 1973) pushdown stack model of short-term memory. The stack contains stimulus–response records for each member of the stimulus ensemble. In addition to generating linear and parallel set size functions, the model predicts an impressive range of effects, including those of serial position, stimulus probability, and sequential response dependencies. However, the model may be too paradigm specific to qualify as a general model of retrieval from immediate memory. Specifically, there appears to be no principled and parsimonious way in which the pushdown storage structure can be extended to cases in which typical Sternberg-like results reliably occur when the stimulus ensemble is large or unknown. Furthermore, self-terminating models predict invariance of minimum RT (McNicol & Stewart, 1980; Sternberg, 1975), whereas minimum RT shows small shifts with set size (e.g., Hockley & Corballis, 1982).

Direct- or parallel-access mechanisms may be less paradigm specific. Parallel-access mechanisms account for set size and serial position effects in one of two ways. One approach, exemplified by Murdock (1971; but see also Townsend & Ashby, 1983), follows the general strategy adopted by Sternberg (1966): The size of the memory set is assumed to directly affect the overall speed of comparison processes. However, in Murdock's (1971) model, the scanning process operates in parallel across all members of the memory set, but individual processing rates vary with the serial position of the test probe. For positive probes, RT is determined by the processing rate for the test probe's serial position. When the test probe is negative, RT is determined by the total time taken for the slowest serial position to be processed to a criterion. By positing a suitable relation between serial position and processing rate, the model can generate both appropriate serial position curves and linear parallel set size functions (see Murdock, 1971). Presumably, the skewing of the RT distribution noted by Hockley and Corballis (1982) reflects the inclusion of items with slower processing rates for larger set sizes.

Another class of direct-access models extends the principles of trace-strength theory (Norman & Wickelgren, 1969; Wickelgren & Norman, 1966) to the probe-recognition paradigm. The size of the memory set is not assumed to affect processing rate directly. Rather, the strength of an item's representation in memory depends on set size and serial position. It is the strength value that, in turn, determines reaction time (Anderson, 1973; Baddeley & Ecob, 1973; Corballis et al., 1972; Nickerson, 1972; Ratcliff, 1978). Such an approach is generally consistent with a number of recent strength-based memory models, including the adaptive resonance model (Grossberg & Stone, 1986), MINERVA (Hintzman, 1984), SAM (Gillund & Shiffrin, 1984), and TODAM (Hockley & Murdock, 1987; Murdock, 1982). Murdock (1985) demonstrated that fitting and deriving explicit predictions from strength-based models entails detailed assumptions about both strength distributions and the decision mechanism that maps strength onto latency.

One direct-access model that can produce linear parallel set size functions and serial position effects is Ratcliff's (1978) random-walk theory of memory retrieval. Ratcliff assumed that comparison processes operate in parallel across all members of the memory set, terminating when the relatedness value of the probe and an element in the memory set exceeds a matching threshold or when all comparisons terminate at a nonmatch threshold. He explicitly adopted a variable match approach to modeling positive set size effects, assuming that larger set sizes lower the degree of relatedness or resonance between the positive test probe and items in the memory set. (Although in principle a match process might reflect interactive processing of the test item and a memory representation, in practice it can be considered as a strength assumption in the short-term memory domain.) Negative responses depend on set size because they must wait for all comparisons to terminate in a nonmatch and all nonmatches have equal resonance statistics. Like Murdock's (1971) account, this assumption ensures that RT will increase with set size because the expected duration of the slowest comparison process increases with the number of comparisons. Taken together, match or strength variations and set-size-dependent decision mechanisms can produce linear set size functions and the pattern of RT distributions noted by Hockley and Corballis (1982). By allowing the relatedness parameter to vary for each serial position, Ratcliff (1978) accommodated serial position mean and distributional patterns.

As an alternative to either a pure serial or parallel model, we note that existing data may also be consistent with hybrid models that postulate a mixture of judgments that are based on a direct-access mechanism and a serial-scan mechanism. Atkinson and Juola (1974) suggested that recognition judgments are first mediated by a direct-access familiarity (strength) mechanism. Fast positive responses or fast negative responses result from probes with familiarity values that either exceed an adjustable positive criterion or fall below an adjustable negative criterion. Intermediate familiarity values initiate a second mechanism to search or scan the memory set. Serial position effects reflect the sensitivity of the direct-access mechanism to trace strength. As Hockley and Corballis (1982) noted, the skewing of the RT distribution follows from the fact that memory set size affects the duration of the second, serial-scan mechanism on just the proportion of trials with intermediate familiarity values.

Summary

According to one view of the short-term list recognition data, the primacy/recency effects seen in serial position functions are primary; approximately linear set size functions are a secondary consequence of averaging over different sets of serial positions. Longer list lengths shift the recency mix of both positive and negative tests. Serial position functions are generally consistent with variance and distributional results. Murdock (1985) perhaps most aggressively pursued this logic, although Monsell's (1978) results also support such a conclusion. Under this view, those few researchers who obtained no serial position effects (e.g., Sternberg, 1966) used longer retention intervals (≥ 1 s), which allowed partial rehearsal to alter subjective recency. This view is supported by data from controlled-rehearsal studies (Seamon & Wright, 1976). The serial position mechanism that conspires to yield linear set size functions is somewhat mysterious. However, adopting the alternative view—that set size effects are primary and serial position effects secondary (e.g., Sternberg, 1975)—does not lessen the mystery. In either case, analyses of serial position data can be critical in evaluating models, and we focus on this analysis here. Serial position effects on RT may reflect a recency-based retrieval mechanism, such as a backwards self-terminating scan (e.g., Murdock & Anderson, 1975). Serial position effects are also consistent with strength effects coupled with a direct-access retrieval mechanism. A number of these positions are nearly indistinguishable from RT data but may be contrasted through speed–accuracy methods.

Speed–Accuracy Trade-Off (SAT) Methodology

In a typical RT version of a probe-recognition task, study time and retention interval are usually set in a fashion that minimizes error rates (less than 10%). Potential differences in accuracy between conditions are difficult to observe with typical sample sizes and, consequently, are rarely considered. Models of short-term memory retrieval are usually contrasted only in terms of reaction time properties. Unfortunately, many of the theoretically distinct classes of models outlined earlier yield, or can be made to yield, equivalent RT predictions. However, as Reed (1976), Pachella (1974), and others showed, models with similar RT predictions nevertheless may differ substantially when the full time course of retrieval is examined. The signal–response speed–accuracy trade-off method (Corbett, 1977; Corbett & Wickelgren, 1978; Dosher, 1976, 1979, 1981, 1982; Reed, 1973, 1976; Wickelgren & Corbett, 1977) provides one means of studying this time course. The method consists of interrupting the retrieval process at various points after the onset of the recognition probe by presenting a tone as a cue to respond. The result is an SAT function in which retrieval time is controlled and response accuracy (usually d') is the dependent measure. By varying the time of interruption between 0.1 and 2.5 s, one can measure the accrual of accuracy over retrieval time.

The full retrieval functions for recognition memory show a period of chance performance, followed by a period of rapid increases in accuracy and finally by accuracy's reaching asymptote as retrieval time is further increased. (Some short-term memory retrieval functions—e.g., in Reed's 1973 study—show late decline in accuracy as information is rapidly forgotten during the retrieval interval.) In general, three parameters suffice to describe these functions: an asymptotic accuracy parameter that reflects memory information limitations, an intercept, and a rate of rise from chance to asymptote. The latter two parameters jointly summarize the dynamics of retrieval. The rising portion of the SAT function may reflect either continuous accrual of information or the distribution of finishing times of a quantal process (Dosher, 1976, 1979, 1981, 1982; Meyer, Irwin, Osman, & Kounios, 1988; Ratcliff, 1988).

Empirically, when RT and SAT paradigms have been performed in comparable experiments, there are strong correlations between parameters of the SAT function and RT. In keeping with early strength models of memory (Murdock & Dufty, 1972; Norman & Wickelgren, 1969), RT can covary with asymptotic strength even in the absence of dynamics differences (Dosher, 1982, 1984a, 1984b; Ratcliff, 1978; Wickelgren, 1977). RT may also covary with dynamics parameters (Dosher, 1981). Hence RT differences may reflect an unknown mix of dynamic and strength effects.

The predicted relation between standard RT and SAT functions is strong under virtually all models. The degree of direct time comparability is model dependent. (We found a strong relation between RT and SAT results in the present studies; consequently, we assume that subjects do not elect to use very different strategies in the two paradigms.) In the case of quantal models, the SAT functions and the distribution of RTs are both assumed to depend on the distribution of finishing times for the measured process. In the case of continuous models, the SAT depends rather directly on the accrual of information, whereas the distribution of RTs depends on accrual of information *and* on the setting of response criteria; in this case, even minimum RTs may be quite delayed in relation to the rise of the SAT function. In addition, interruption cues in SAT may require additional processing, thus shifting SAT functions in relation to RT (see also Ratcliff, 1988).

In subsequent sections, the various accounts of retrieval are contrasted in terms of their respective SAT predictions concerning the effects of set size and recency.

SAT Set Size Effects

Asymptotic Accuracy

With the exception of strength-based models, most scanning accounts, such as Sternberg's (1966) and Murdock's (1971), assume that each study item is perfectly represented in the memory set. However, data from an SAT study reported by Reed (1976), reanalyzed here in Figure 1, remind us that asymptotic accuracy levels differ reliably across set sizes of one, two, and four items. Error rates in comparable RT data were below 9% and generally would be ignored. What, in fact, these data suggest is that items from different set sizes are not equal in their overall strength values.

Reed's (1976) data and our data here require the modification of scanning or other models that assume equal strength values. For scan models, each item may be probabilistically

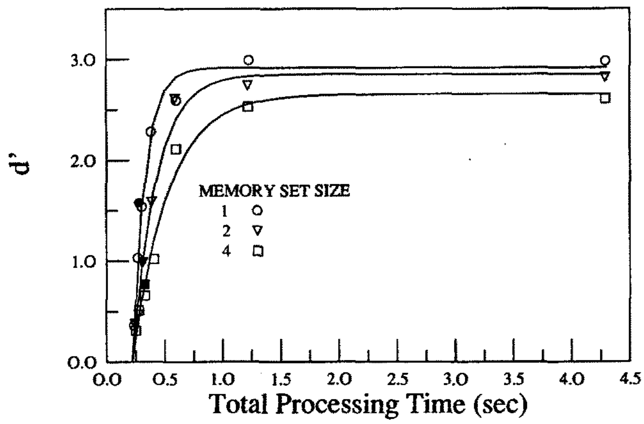

Figure 1. d' as a function of lag latency for set sizes of one, two, and four consonants reported by Reed (1976). (Open symbols are interruption speed–accuracy trade-off data. Solid symbols represent comparable reaction time data. Smooth curves are fits of the exponential model [Equation 1] with different rates [β] and asymptotes [λ] for each set size but a common intercept [δ].)

represented in the scan process with set size determining accuracy. Or, in accordance with Treisman and Doctor's (1987) approach, each comparison process may have some probability of error that depends on set size. Direct-access models require equivalent strength assumptions. Reed (1976) tested the predictions of asymptotic accuracy levels for a number of models with a strength component. He considered three capacity-sharing strength-allocation models, which differed in their allocation policy: specifically, (a) that strength is inversely related to the set size (Baddeley & Ecob, 1973), (b) that strength is the inverse square root of set size (Anderson, 1973), and (c) that strength is inversely proportional to the $\frac{2}{3}$ power of set size (Reed, 1976). These capacity models all vastly overestimated the observed asymptotic differences. However, strength need not and probably should not be construed strictly as a capacity. A viable alternative, consistent with Monsell's (1978) and Murdock's (1985) analysis, simply assumes that strength varies as a function of serial position. This implies that overall strength will decrease as set size increases, although not necessarily as much as capacity models suggest.

Retrieval Dynamics

Granting that modified models accommodate asymptotic accuracy differences, the critical contrasts concern the dynamic portions of the SAT functions. Dynamic differences are summarized by intercept and rate parameters.

Under most assumptions, exhaustive scanning models predict substantial differences between set sizes in the dynamics of the SAT curve. The exact form and magnitude of these differences, of course, depend on further assumptions concerning the nature of the scan process, its relation to the cued-response paradigm, and distributions of scan times. We interpret Sternberg's (1966) model as a quantal or uninterruptible scan process, yielding no information until the scan is complete. Thus the SAT reflects the distribution of finishing times, coupled with either guessing or scan-controlled accuracy. (For a further discussion, see also Meyer et al., 1988.) The SAT intercept and rate are determined by the leading edge and variability of the finishing-time distribution. The SAT asymptote reflects limits in the accuracy of the scan process. When two conditions differ in asymptotic accuracy but reflect the same distribution of finishing times, the SAT functions rise from intercept to asymptote (proportionately) in the same time. Thus SAT curves with the same proportional relation to asymptote reflect identical dynamics. (Deviations from strict proportionality due to guessing are minor.) Whether set size primarily affects rate or intercept or both depends on its relative impact on the leading edge and variability of scan times.

Hypothetical results of a modified exhaustive scan model are shown in Figure 2. It is assumed that the test item is compared exhaustively with a list representation of length n but that there is some chance of an error (either miss or false alarm) in each comparison. Each list comparison had an independent probability of producing a false alarm (.01 in Figure 2), so that overall false alarm rate increased with list length. Target trials had a miss rate that was also assumed to increase with list length. Thus even when the scan is complete, the comparison operations will have yielded an error some fraction of the time. (This is only one possible extension of Sternberg's 1966 model to account for response errors.) The distribution of scan completion times was assumed to be exponential overall, with an increase in mean finishing time

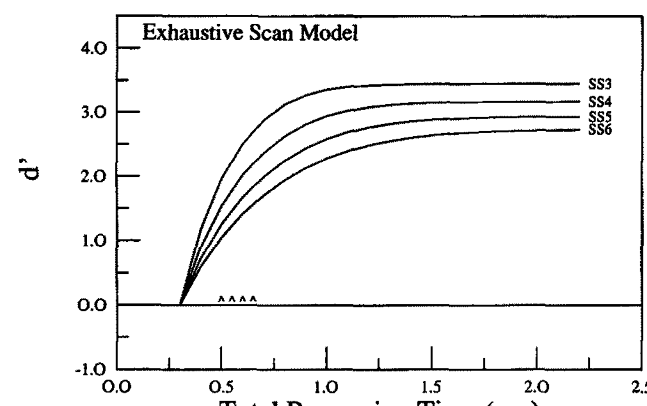

Figure 2. Predicted speed–accuracy trade-off (SAT) functions for set sizes three through six from the modified exhaustive scan model. (The finishing times for the scan process were assumed to be exponentially distributed, whereby each additional item in a set increases the mean finishing time by 50 ms. Set size was assumed to control asymptotic accuracy levels; the probability of a miss increases by .02 and the probability of a false alarm by approximately .01 with each additional item in a set. Responses before the completion of a scan are random guesses with probability correct of .5. The ^ marks above the 0 d' level on the x axis correspond, from left to right, to the time at which set sizes three, four, five, and six reach 2/3 of asymptote. This illustrates a pure rate effect of set size on the SAT functions; other distributional assumptions might predict intercept as well as rate differences.)

of 50 ms for each additional list member. Responses occurring before scan completion are random guesses. The hit and false alarm rates resulting from either completed scans or guessing at any retrieval time t were transformed into the equivalent d'. The overall exponential finishing time assumptions of Figure 2 illustrate a pure rate effect of set size on the SAT functions. Distributional assumptions more typical of multistage models (e.g., gamma functions; see Townsend & Ashby, 1983) exhibit apparent intercept, as well as rate, shifts with increasing memory set size.

Treisman and Doctor (1987) modified Sternberg's (1966) model by assuming that multiple fast scans are strategically performed throughout the course of retrieval. Accuracy is jointly determined by the size of the memory set and the number of scans performed before a response. Because the size of the memory set determines the time needed for a single scan, set size consequently imposes a limit on the number of scans performed in a given interval of time. Accuracy will grow at a rate determined by the time taken for a single scan, producing set size functions that, like those illustrated in Figure 2, rise disproportionally to asymptote. However, accuracy accumulation is not two-state quantal. As the number of gradations increases, it becomes more similar to a continuous accumulation model.

In the parallel self-terminating model of Murdock (1971), processing starts from a common time intercept; information accrual is governed by rate constants associated with the serial positions of the test elements. If each serial position is tested equally often, the overall processing rates for various set sizes will necessarily differ because they are determined simply by the average of their respective rate constants. The magnitude of the differences depends on the parameters governing the assignment of rates to serial position.

Ratcliff's (1978) extension of random-walk models to memory data embodies a continuous-accrual and direct-access metaphor. The treatment of set size and serial position is very similar to that of Murdock (1971). Predicted set size effects for Reed's (1976) data combined a match or resonance value for each item (assumed to be invariant over serial position and set size, insofar as Reed used a long retention interval) with a decision mechanism that was based on parallel processing of a (parallel) search set defined by the memory set. Elsewhere in Ratcliff's article, however, serial positions for different set sizes were each assigned a different strength in order to account for RT differences. In either case, the decision mechanism produces slower rates and lower asymptotes for larger set sizes. (The quantitative details vary with assumed bias parameters, as discussed later.)

On the basis of comparisons of SAT functions for set sizes of one, two, and four items, Reed (1976) concluded that there were differences in the SAT dynamics as a function of set size, although among models that he considered, only Theios et al.'s (1973) pushdown stack model fit the functions particularly well. However, we view Reed's conclusions with caution. Because the explicit strength models considered by Reed vastly overpredicted the observed asymptotic differences in set size, he ignored those differences and fit the functions with a common asymptote. Ignoring genuine asymptotic differences distorts estimates of dynamic properties. In addition, Reed did not examine serial position functions within set size. We suggest that the proper interpretation of differences in the dynamics of set size SAT functions first requires a careful examination of the dynamics of the serial positions.

SAT Recency Effects

Asymptotic Accuracy

Wickelgren et al. (1980) reported that SAT asymptotic accuracy depends directly on recency in the same way as untimed forgetting functions. Thus asymptotic differences for different set sizes could reflect the average statistics of the serial positions of the set. Unfortunately, for our purposes, Wickelgren et al. (1980) used superspan constant-length lists of 18 items. Therefore, Wickelgren et al. did not measure the dependency of asymptotic pattern on list length.

Retrieval Dynamics

Once again, the critical issue is whether serial position or set size differences extend to the dynamic properties of SAT functions. Differences in the SAT dynamics for various serial positions could result directly from either serial or parallel self-terminating scans. In the former case, if the list is scanned in the approximate order of recency, then the expected duration of a scan will decrease as a direct function of the recency of a probe. Accuracy under these conditions will rise at a rate that is faster the more recently a test probe was studied.

Hypothetical functions for a backwards serial scan model are shown in Figure 3 for serial positions in a set size of 5. These functions were generated under the assumption that serial position affects asymptotic accuracy. Accuracy monotonically decreases for probes drawn from less recent serial positions with a small primacy advantage for the first position in the list. The overall distribution of scan times was assumed to be exponential; each additional comparison process added 50 ms to the overall mean finishing time. In Figure 3, we illustrate the fact that if the memory list is scanned from the most recent item to the least recent item, terminating on a match, then the rate of rise in accuracy over time covaries with serial position. Again, other distributional assumptions (e.g., gamma functions) may introduce intercept shifts, as well as rate shifts.

Similar dynamic differences are predicted by parallel self-terminating models such as Murdock's (1971), which explicitly assume that processing rate varies as a function of serial position. Under Murdock's model, the processing rate parameters directly affect dynamics. Because Ratcliff (1978) assumed identical resonance or strength distributions for all nonmatch comparisons and identical dynamics for any one match comparison, serial positions within a set size may show approximately identical dynamics, although they differ in asymptotic accuracy. However, the decision rule produces set size differences in dynamics by requiring exhaustive processing of the memory set on all nonmatch comparisons.

In Wickelgren et al.'s (1980) constant-length lists, the test probe's serial position affected terminal accuracy levels only.

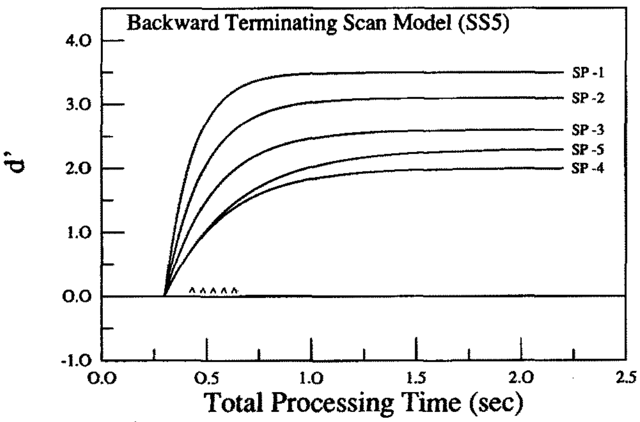

Figure 3. Predicted serial position speed–accuracy trade-off (SAT) functions (in a set size of five items [SS5]) from a backwards serial self-terminating scan model. (Asymptotic accuracy monotonically decreases as the probe is drawn from less recent serial positions, with the exception of a small increment for the primary list position. The finishing time distribution was assumed to be exponential, whereby mean increases by 50 ms for each serial position from the end of the list. In keeping with Monsell [1978], serial position is labeled in terms of recency, counting backwards from the test item, denoted in position 0, to the study position of the probe. For example, the most recent serial position is labeled −1, the next most recent position −2, and so on. The ^ marks above the 0 d' level on the x-axis correspond, from left to right, to the time at which the retrieval functions for serial positions −1, −2, . . ., −5 reach 2/3 of asymptote. This illustrates a pure rate effect of serial position on the SAT functions.)

The dynamics were identical, with the exception of the most recently presented item. Wickelgren et al. found (see also Dosher, 1981) that the retrieval dynamics for the most recently presented item (immediate repetition) were significantly faster than for any other position. Speeded dynamics for immediate repetition was related to a distinction between an item in active and passive memory states. When little or no mental activity has intervened between the study and testing of an item, then the item remains active in primary memory. This state is contrasted with the more passive state of other items displaced from active processing by intervening mental activity. However, it is unclear whether this result will hold in a subspan domain. In addition, Wickelgren et al. did not examine whether the length of the list affected the dynamics of retrieval in a constant way for all serial positions.

Summary

Various models of short-term memory retrieval differ in their accounts of serial position and/or set size effects. Because most of these models are generally consistent with extant RT data (either as they stand or with minor modifications), it is difficult to discriminate between them strictly within the context of RT paradigms. However, many of these models of item recognition yield quite different predictions concerning the respective effects of serial position and set size on the full time course of retrieval: most critically, in predictions of the dynamic of retrieval measured independently of asymptotic strength, as through the interruption SAT method.

We report two interruption SAT studies (Experiments 1 and 3) and a complementary RT study (Experiment 2) designed to discriminate between these broad classes of retrieval models. In the SAT studies we examined full time course serial position functions within different set sizes, independently estimating the dynamics and asymptotic strength of retrieval both within and across the memory sets. In order to establish a strong basis on which to generalize from SAT to standard RT studies, we report in Experiment 2 the pattern of RT data that emerge from the experimental parameters used in the SAT studies. In addition, in Experiments 2 and 3, we examined and report the effect of the recency of a lure on the RT and the time course of lure rejection.

Experiment 1

In Experiment 1 we used the interruption SAT paradigm to examine retrieval functions for all serial positions of memory set sizes of three and five words. Memory list lengths of three and five are within the estimated "subspan" range for words (see Burrows & Okada, 1975; Okada & Burrows, 1978) and yet are predicted to differ in terminal accuracy and retrieval dynamics (see Figure 2).

We used a relatively fast presentation rate (500 ms) and a short retention interval (300 ms). These are typical of conditions under which subjective and objective recency are strongly coupled (Monsell, 1978). A pilot RT study with set sizes of three, four, and five words and a slight longer retention interval (400 ms) was run to confirm that standard serial position effects (Monsell, 1978) occurred in our experimental situation. (In the Method section, we describe the materials, design, display, and timing; only the mode of response differed. These data were from 10 subjects, after 30 practice trials.) In Table 1 we present the mean correct RT and error rates for each serial position, for the average of positives, and for negative probes for each set size. In Figure 4 we show the serial position RT functions graphically. The serial position RTs display both a strong recency and a weaker primacy effect. In comparisons across set sizes, the serial positions are approximately superimposed, as reported by Monsell (1978). These RT patterns demonstrate that minor implementation differences between our experiments and Monsell's are unimportant. We now consider the SAT version of the serial position and set size experiment.

Method

Subjects

Four subjects each completed eight 1-hr sessions. Subjects BM, GR, and GM were affiliated with the laboratory and volunteered their services. Subject GD was paid. All subjects either were experienced with the cued-response SAT methodology or performed at least a 1-hr initial training session.

Table 1
Average Reaction Times (RT) and Proportion Errors (PE) in Pilot Study

	Serial position													
	−5		−4		−3		−2		−1		Y		N	
Set size	RT	PE	RT	PE	RT	PE	RT	PE	RT	PE	RT	PE	RT	PE
5	732	.11	785	.15	741	.07	699	.08	599	.04	711	.09	716	.06
4			730	.09	748	.14	728	.06	604	.02	702	.08	715	.06
3					649	.06	692	.05	609	.04	650	.05	692	.07

Note. Y = positive probes; N = negative probes.

Design and Stimuli

We used a varied set procedure. Two lists of 50 two-syllable nouns were constructed from Paivio, Yuille, and Madigan's (1968) norms to serve as stimuli. We constructed the study list for a particular trial by drawing randomly without replacement from one of the two lists. The lists were alternately sampled on successive trials.

Memory list lengths of three or five words were randomly mixed within a session. Within a session and within a set size, there were equal numbers of positive and negative trials. The recognition probe on positive trials was randomly drawn from one of the list positions so that each position was equally sampled across a session. The recognition probe on negative trials was randomly drawn from the same 50-word stimulus list as the memory set. Thus all negative recognition probes were relatively nonrecent, inasmuch as they could not have been seen on the previous trial. This sampling procedure also ensures that negative probes are not repeated on adjacent trials. Each condition was tested at each of eight interruption points, or lags, ranging from 0.1 to 1.8 s (see the following section).

Procedure

Stimulus presentation and response collection were controlled by an Apple IIe computer with which we used a Sanyo VM4209 monitor along with a Superclock II timing card. The words of the study list were presented in uppercase letters with an approximate character size of 6 × 4 cm, viewed at a distance of approximately 40 cm.

Each of the eight sessions consisted of a total of 512 trials, divided into four blocks. Across the eight sessions, this yielded a total of 32 trials per lag for each positive serial position probe within each set size, 96 negative probes per lag for set size of three, and 160 negative probes per lag for set size of five. Subjects were free to rest between blocks, but once a block of trials started, it proceeded automatically. The sequence of events (see Figure 5) was as follows:

1. The word "READY" appeared for 500 ms at the top in the center of the screen.

2. The screen cleared and a series of angular brackets were presented for 500 ms. These brackets appeared in the center of the screen and enclosed a region in which the words from the study list were presented. The initial number of brackets on each side denoted the set size of the trial (either three or five words).

3. The words of the study list were presented in the enclosed region in succession for 500 ms each. The presentation of each successive word reduced the number of brackets on each side by one. Thus the brackets provided a running countdown of the number of study words remaining. The appropriate number of brackets remained on the screen at all times, spanning a 50-ms pause between successive presentations of the study words.

4. After the presentation of the last study word, the screen cleared and a high-contrast mask was displayed over the region used for presentation of the study words. It remained on until the probe word was presented.

5. Three hundred milliseconds after the presentation of the last study word, the recognition probe was presented in uppercase letters, approximately 8 cm below the region used for presentation of the study list. The probe item appeared for a variable length of time, depending on the cue to respond.

6. At 0.10, 0.20, 0.30, 0.40, 0.55, 0.9, 1.3, or 1.8 s after the onset of the probe item, the subject was cued by a brief tone to respond.

7. The subject executed a yes–no recognition response as quickly as possible after the onset of the cue-to-respond tone.

8. Latency feedback appeared on the screen for 300 ms to enable the subjects to monitor their performance. Subjects were instructed to respond within 270 ms of the tone, regardless of their accuracy performance. They were told that responses that took longer that 270 ms were too long and that responses that took less than 150 ms were anticipations.

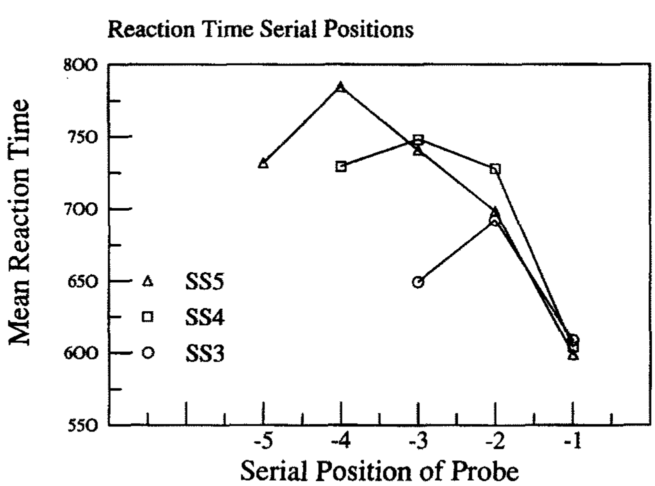

Figure 4. Average correct positive reaction time as a function of serial position for set sizes (SS) three through five. (Serial position is labeled in terms of recency, counting backwards from the test item, position 0, to the study position of the probe, −1 for the most recent serial position, −2 for the next, and so on.)

Results and Discussion

The latency and proportion correct for positive and negative trials and corresponding *d*'s are presented in Table 2, averaged over subjects. (Individual subject data is available from us.) These measures were presented for the two set size conditions and for the serial positions within each set size.

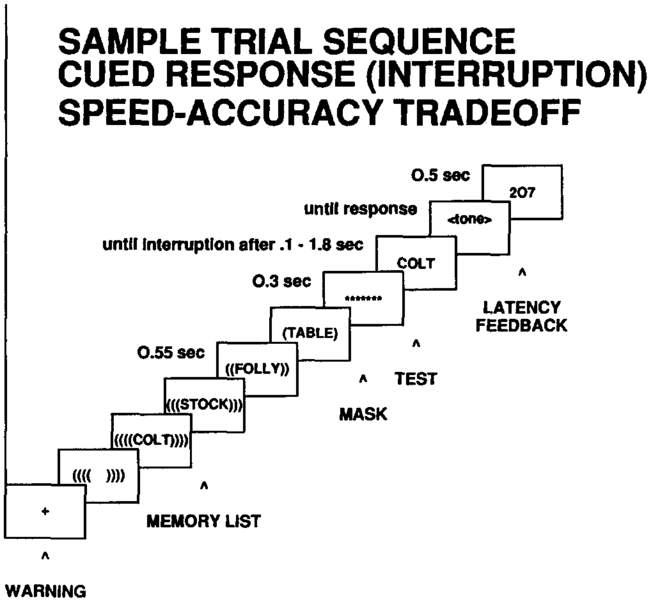

Figure 5. A sample trial sequence and timing.

Interruption (Lag) Latency Analysis

Latency to respond depended on interruption point in both the set size, $F(7, 21) = 9.742$, $p = .0000$, and the serial position functions, $F(7, 21) = 9.904$, $p = .0000$ (for the comparison of serial positions within set size three), and $F(7, 21) = 7.640$, $p = .0001$ (for the comparison of serial positions within set size five). Latency did not, however, directly vary as a function of any of the stimulus conditions: Neither the set size, $F(1, 3) < 1$, *ns*, nor the serial position of the probe—$F(2, 6) < 1$, *ns*, for serial positions within set size three and $F(4, 12) < 1$, *ns*, for serial positions within set size five—significantly affected latency. Although latency was slightly faster for set size three across the 0.10- to 0.55-s lags, the interaction between lag and set size proved nonsignificant, $F(7, 21) < 1$. Within set size, however, the interaction of serial position and lag was significant for set size five, $F(28, 84) = 1.989$, $p = .0086$. This interaction reflects the faster latency for the most recent serial position throughout the early interruption points (0.10 to 0.55 s). The same interaction in set size three data was nonsignificant, $F(14, 42) = 1.399$.

In general, latency varies with time of interruption but not with experimental condition (Dosher, 1981, 1982, 1984a, 1984b). In that case, including latency in total retrieval time affects absolute time estimates but does not contribute to condition differences in SATs. However, latency does occasionally vary with experimental condition, usually in response to some obvious stimulus property (Dosher, 1981). Here, latency was somewhat faster when the probe matched the last list item (the case of immediate repetition). This speeding is also seen in other properties of the SAT function for this serial position (to be discussed). This effect on latency resulted in an apparent difference in set sizes that is based in differential contributions of immediate recency to the average. Reed (1976) also found small set size differences in latency. In the analyses to be presented, we followed the standard convention of plotting SAT functions as accuracy against total processing time (lag time plus latency). Excluding the latency difference would underestimate the effect of immediate recency in SATs.

Speed–Accuracy Trade-Off (SAT) Functions

Empirical set size and serial position SAT functions (total processing time and d' at each lag) are listed in Table 2. (The d' statistic for some individual subjects included a minimum-error correction for those few conditions in which hit probability was 1 or false alarm probability was 0. Technical considerations involving this correction process and potential interactions with subsequent model fits will be extensively treated in a separate discussion section.) As with previous SAT studies (e.g., Dosher, 1976, 1981, 1982, 1984a; Reed, 1976; Wickelgren et al., 1980), these functions can be closely approximated by an exponential approach to a limit:

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}),\ t > \delta, \text{ else } 0. \quad (1)$$

Alternatively, these empirical functions can be fit by a retrieval function that explicitly assumes a time-bounded diffusion (continuous random-walk) process (Ratcliff, 1978):

$$d'(t) = \frac{\lambda}{\sqrt{1 + \nu^2/(t - \delta)}},\ t > \delta, \text{ else } 0. \quad (2)$$

In both functions, $d'(t)$ is the predicted d' at time t, λ controls the asymptotic accuracy level, and δ is the intercept or time before which accuracy is at chance level. In Equation 1, β is an exponential rate parameter that indexes how rapidly accuracy rises from chance to asymptotic level. In Equation 2, ν^2 is a combined diffusion variance term that functions in the same way as β. Both β and ν^2 have the approximately proportional character required by quantal retrieval models and by many continuous retrieval models. (See the preceding discussion of Figure 2.)

We fit models with both functional forms to all our data. Both equations are used here in a purely descriptive fashion. Whereas the diffusion model has internal consistency constraints on parameters when fitting certain kinds of RT designs, there are no such checks in the SAT domain. Also, Equation 2 is for a uniprocess diffusion model. All statements about rate differences reflect a translation into that uniprocess form. Usually, Equations 1 and 2 are virtually interchangeable empirically. Prior analyses of associative recognition data, which had slower dynamics than the single-item recognition data studied here, have consistently shown small but negligible preference for the exponential form. However, in the item-recognition domain, the diffusion function with slower ν^2 parameters generally reaches asymptote more slowly than do the data. This is illustrated in Ratcliff's (1978) fits of Reed's (1976) data in which the diffusion model shows a delayed intercept estimate that underestimates the first lag data in order to allow a faster rate parameter. This technical difficulty introduces a bias to capture condition differences in intercept rather than in rate. It is possible that introducing noise in the

Table 2
Experiment 1: Average Latency (in s), Proportion Correct, and d′

Item	Lag 1	2	3	4	5	6	7	8
Set size 3								
Positive trials								
Latency	0.266	0.223	0.217	0.211	0.201	0.189	0.193	0.204
Proportion correct	0.573	0.733	0.835	0.945	0.960	0.990	0.983	0.995
Negative trials								
Latency	0.276	0.233	0.217	0.202	0.201	0.197	0.197	0.191
Proportion correct	0.660	0.842	0.895	0.922	0.982	0.990	0.995	0.995
Positive vs. negative trials: d′	0.639	1.709	2.268	3.065	3.830	4.210	4.338	4.517
Serial position −3								
Latency	0.272	0.233	0.214	0.208	0.200	0.194	0.198	0.192
Proportion correct	0.660	0.630	0.808	0.915	0.970	0.990	1.000	1.000
d′	0.694	1.410	2.168	2.872	3.974	4.203	4.586	4.586
Serial position −2								
Latency	0.273	0.234	0.209	0.203	0.202	0.197	0.194	0.189
Proportion correct	0.428	0.658	0.775	0.938	0.923	0.990	0.962	0.990
d′	0.366	1.488	2.060	2.998	3.507	4.203	4.085	4.442
Serial position −1								
Latency	0.267	0.219	0.215	0.207	0.203	0.194	0.196	0.211
Proportion correct	0.770	0.910	0.925	0.990	1.000	1.000	1.000	1.000
d′	1.393	2.517	2.864	3.616	4.406	4.347	4.586	4.586
Set size 5								
Positive trials								
Latency	0.281	0.237	0.224	0.217	0.208	0.193	0.194	0.195
Proportion correct	0.483	0.550	0.725	0.858	0.880	0.920	0.925	0.910
Negative trials								
Latency	0.275	0.237	0.219	0.209	0.202	0.195	0.195	0.190
Proportion correct	0.650	0.782	0.847	0.907	0.972	0.972	0.977	0.980
Positive vs. negative trials: d′	0.457	0.943	1.670	2.424	3.201	3.438	3.514	3.460
Serial position −5								
Latency	0.277	0.238	0.222	0.213	0.203	0.192	0.194	0.192
Proportion correct	0.387	0.407	0.660	0.793	0.855	0.880	0.915	0.895
d′	0.353	0.549	1.468	2.174	3.123	3.350	3.456	3.448
Serial position −4								
Latency	0.279	0.232	0.226	0.215	0.205	0.193	0.200	0.188
Proportion correct	0.373	0.395	0.560	0.702	0.720	0.860	0.842	0.853
d′	0.369	0.527	1.203	1.880	2.616	3.176	3.113	3.191
Serial position −3								
Latency	0.275	0.246	0.221	0.210	0.200	0.194	0.194	0.194
Proportion correct	0.385	0.545	0.695	0.892	0.913	0.925	0.952	0.880
d′	0.336	0.928	1.629	2.643	3.437	3.541	3.729	3.406
Serial position −2								
Latency	0.278	0.233	0.222	0.212	0.202	0.196	0.195	0.195
Proportion correct	0.522	0.525	0.798	0.935	0.915	0.940	0.915	0.938
d′	0.568	0.889	1.935	3.072	3.401	3.703	3.435	3.619
Serial position −1								
Latency	0.274	0.224	0.214	0.210	0.211	0.194	0.195	0.191
Proportion correct	0.748	0.868	0.897	0.967	0.998	0.998	0.998	0.988
d′	1.296	2.132	2.486	3.266	4.349	4.325	4.374	4.237

intercept of the diffusion model would eliminate this difficulty (R. Ratcliff, personal communication, November 1987). However, no analytic form is available for this case. We show fitted functions for the exponential because they involve fewer systematic misfits to the data and more stable parameter estimates. However, because the diffusion model is of greater immediate theoretical interest, we also report the comparable diffusion fits.

An iterative hill-climbing algorithm (Reed, 1976), similar to Stepit, was used to fit systems of retrieval equations to the data. The quality of the fit was assessed by the statistic

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (d_i - \hat{d}_i)^2/(n-k)}{\sum_{i=1}^{n} (d_i - \bar{d})^2/(n-1)}, \quad (3)$$

wherein d_i are the observed d' values, the $\hat{d}_i$ *are the predicted values,* $\bar{d}$ is the mean, n is the number of data points, and k is the number of free parameters. This statistic represents the goodness of fit adjusted for the number of free parameters. All conclusions to be reported are based on analyses of individual subject data. The effects were sufficiently consistent over subjects that they can be summarized by analyses performed on the average data.

Serial Position Functions

In Figure 6 we plot d', averaged over subjects and the last four interruption points, for each serial position. This is an empirical measure of asymptotic differences in the recognition performance. Within both set sizes, the serial position of the probe significantly affected asymptotic accuracy—$F(2, 6) = 9.42$, $p = .014$, for serial positions within set size three and $F(4, 12) = 12.06$, $p = .0004$, for serial positions within set size five—producing typical bowed functions. However, accuracy levels differed for nominally identical serial positions from different set sizes (see Figure 6). The overlapping RT serial position functions of Monsell (1978) represented performance only on positive trials. Our d' measure scaled each positive trial type against negative trials for the appropriate set size. Although the hit rates (for positives alone) did not overlap perfectly in our data, set-size-dependent changes in criterion cannot be ruled out in either the RT or the SAT paradigm, especially as the set size is saliently cued at the beginning of each list presentation.

In fitting both retrieval equations, we allotted a separate asymptotic parameter (λ) for each serial position within set size. When asymptotic differences were ignored, the resulting fits showed systematic deviations for all subjects and relatively poor fit, with an average R^2 value of .811 for the exponential model and an average R^2 of .778 for the diffusion model. (R^2 values in the .800 range often result from capturing the generally increasing character of the SAT functions. Serious models must capture the additional condition-specific differences.) Allowing λ to vary with serial position improved the quality of the fit for all subjects, on average increasing the R^2 values to .933 for the exponential model and .899 for the diffusion model.

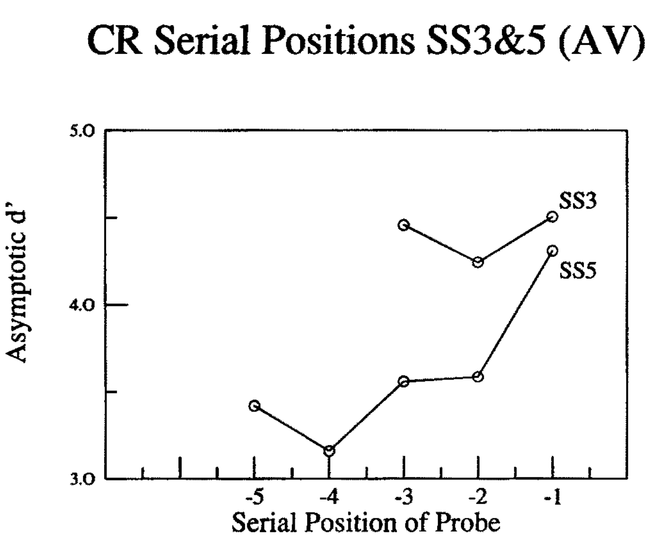

Figure 6. Observed average asymptotic d' values for serial positions within set sizes (SS) three and five. (Serial position is labeled in terms of recency, counting backwards from the test item, position 0, to the study position of the probe, −1 for the most recent serial position, −2 for the next, and so on.)

Of critical interest are dynamics differences (reflected in intercept and rate) for serial position and set size. In total, 12 different models were fit to the data for each of the retrieval equations. There was a minor discrepancy in how the exponential and diffusion equations best treated the dynamic properties of the SAT functions. The exponential equation produced marginally better fits when the rate (β) parameter, as opposed to the intercept (δ) parameter, was varied. Conversely, the diffusion equation produced better fits when intercept (δ) rather than rate (ν^2) was allowed to vary. This is related to the technical difficulties of the diffusion model described earlier. The patterns of condition effects were identical; it was simply a question of which dynamics parameter captured the differences. Forcing an effect into the rate parameter of the diffusion equation distorted asymptotic estimates. Rather than directly constraining the two equations to fit the data with the same type of dynamic parameter, and thereby introducing artificial parameter trade-offs, we allowed each to fit dynamic differences with its preferred dynamic parameter. Consequently, when using the exponential equation, we first systematically varied the rate parameter as a function of serial position, found the best fit and then examined whether additional intercept variation improved this fit. When using the diffusion equation, we first varied the intercept and then examined whether variation in rate further improved the fit. From this point, the results are fully reported for the less problematic exponential form. However, the analogous diffusion fits are included in all the tables. The few cases with small discrepancies from the exponential are explicitly mentioned.

Allowing separate rate parameters for the two set sizes did not reliably improve the goodness of fit over the model that assumed common dynamics (with different asymptotic strength values for both models). The average R^2 was .933 for an exponential model with separate βs for each set size, which is comparable with an average R^2 of .933 for the common dynamics model. (The average R^2 was .894 for a diffusion model with separate δs for each set size, in comparison with .899 for the equal dynamics model.) The major shortcoming of this and related models was the tendency to underestimate d' at the early interruption points for the most recent serial positions. Simply allotting one rate to probes drawn from the most recent study position, irrespective of set size, and another rate to probes drawn from every other serial position, an 8λ–2β–1δ exponential model, improved the goodness of fit for all subjects and the average, yielding an average R^2 of .956. (Equivalently, an 8λ–$1\nu^2$–2δ diffusion model improved the fit for all subjects, raising the average R^2 to .942.) The parameter values and the goodness of fit resulting from the best exponential and diffusion models are shown in Table 3 for the average data and for each individual subject. In Figure 7 we

Table 3
Experiment 1: Serial Position Parameter Estimates and Goodness of Fit

Parameters	Average	Subjects: BM	GD	GM	GR
Exponential equation					
Set size 3 λs					
SP −1	4.54	4.45	4.74	4.56	4.49
SP −2	4.38	4.18	4.59	4.03	4.73
SP −3	4.63	4.60	4.77	4.62	4.50
Set size 5 λs					
SP −1	4.26	3.98	4.64	4.36	4.21
SP −2	3.86	3.60	4.22	4.00	3.66
SP −3	3.74	3.69	3.81	3.42	4.04
SP −4	3.12	3.03	3.49	2.66	3.29
SP −5	3.46	3.04	3.66	3.03	4.15
Recency β	6.69	10.0	4.64	5.49	7.84
Other β	3.72	3.93	3.28	3.99	3.79
Common δ	.333	.303	.321	.330	.379
R^2	.956	.927	.919	.912	.925
Diffusion equation					
Set size 3 λs					
SP −1	5.27	5.08	5.04	4.79	5.49
SP −2	4.86	4.43	4.91	4.29	5.39
SP −3	5.12	4.88	5.14	4.92	5.12
Set size 5 λs					
SP −1	4.94	4.53	4.91	4.56	5.21
SP −2	4.28	3.84	4.55	4.28	4.14
SP −3	4.14	3.94	4.16	3.67	4.58
SP −4	3.47	3.22	3.79	2.86	3.72
SP −5	3.85	3.26	4.00	3.25	4.67
Common ν^2	.319	.216	.228	.170	.412
Recency δ	.345	.303	.372	.374	.374
Other δ	.419	.403	.477	.467	.435
R^2	.942	.930	.888	.897	.898

Note. SP = serial position.

present the average observed d' values for serial positions within each set size, along with the descriptive functions generated by the exponential model in Table 3.

The retrieval advantage for probes of the most recent serial position, the case of immediate repetition, translates on average into a 44% faster rate parameter in the exponential model (see Table 3). This retrieval advantage was consistent and substantial for all 4 subjects, ranging from 30% to 62% faster rates. (The same pattern is observed in the δ parameter of the diffusion model: The average intercept for the most recent serial position was estimated at 345 ms, in comparison with 419 ms for all other serial positions. The intercept advantage was consistent over the 4 subjects and ranged from 61 to 105 ms.)

We report the results of two additional fits in order to reinforce the adequacy of the model presented in Table 3. First, we examined a model in which each serial position (within set size) was allowed a separate rate. This is a way to test whether there is a more general advantage for recency that extends past the most recent item. For the exponential, the β estimates were, on average, 6.57, 3.54, and 3.49 for set size three and 6.68, 4.94, 4.16, 3.10, and 3.44 for set size five, in order of the most recent to most distant serial positions. Despite some small tendency to decline with serial position, the differences in β as a function of serial position beyond the most recent position were relatively minor when compared with the sharp discontinuity between the most recent position and all others, and the added parameters do not improve R^2s. (The diffusion δ estimates showed an exactly comparable pattern.) Moreover, the slight tendency beyond the most recent item was clearly conditioned on trade-offs with estimates of asymptote, in some cases producing systematic asymptotic misfits.

We also examined whether there were any consistent residual differences in rate for the two set sizes once the most recent serial position was factored out. To do so, we allotted independent rates to the most recent serial position, irrespec-

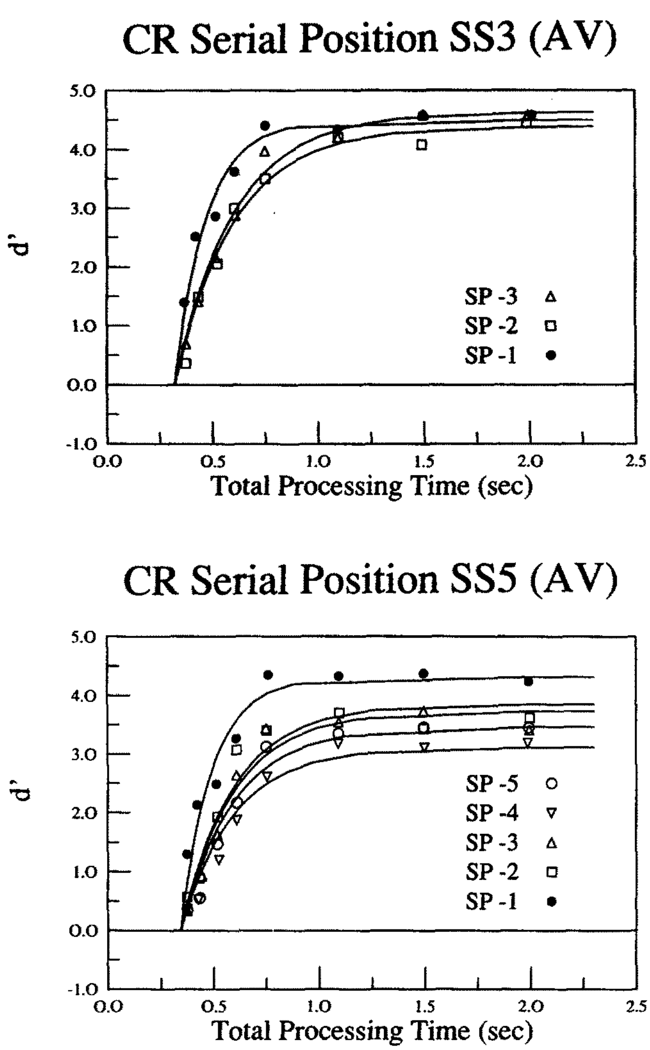

Figure 7. Observed average d' values as a function of total processing time for serial positions with set sizes three (SS3; top panel) and five (SS5; bottom panel). (Smooth functions are derived from the estimated parameters of the exponential model fits presented Table 3.)

tive of set size, and to the remaining serial positions within each set size separately, namely, with an 8λ–3β–1δ exponential model (or an 8λ–1ν²–3δ diffusion model), wherein the second and third βs (or δs) represented separate dynamics for set sizes three and five minus the most recent serial position. These fits did not improve on those shown in Table 3, and estimated rate differences between set sizes three and five were small and inconsistent across subjects.

The serial position data are therefore relatively easy to summarize. The serial position of a probe appears primarily to affect the asymptotic accuracy of recognition performance. Serial position may affect the strength of an item's memorial representation in a manner that follows rather directly from strength theories of recognition memory (Norman & Wickelgren, 1969; Wickelgren & Norman, 1966). Differences in retrieval dynamics as a function of serial position appear to be restricted to probes drawn from the most recent serial position—the case of immediate repetition between study and test. This later finding replicated, in a subspan list domain, the results with superspan lists reported by Wickelgren et al. (1980). It thereby reinforces the distinction drawn between an active memory for the last item or thought and a passive memory for preceding items or thoughts.

One might assume that the immediate-repetition retrieval advantage is solely a consequence of a perceptual or physical match mediating responses to probes drawn from the most recent serial position. Such a view is plausible, given that the retention interval in this study (300 ms) was well within the range that could be extrapolated from studies such as that of Posner, Boies, Eichelman, and Taylor (1969). In Experiments 2 and 3 we directly examined this possibility by altering the type case between study and test.

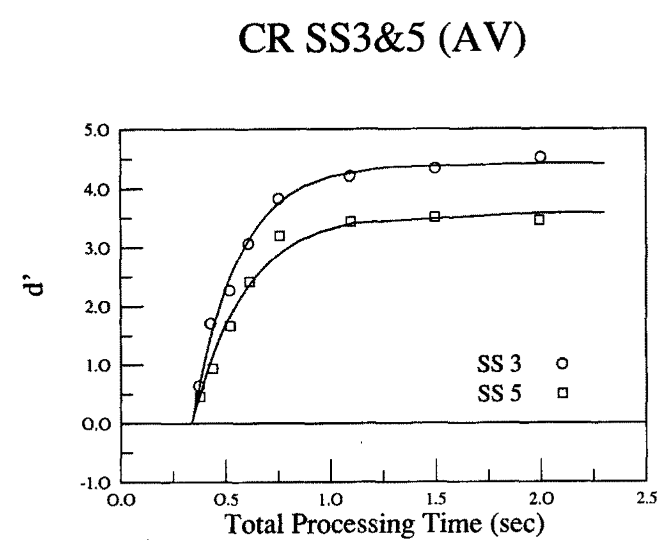

Figure 8. Observed average *d'* values as a function of total processing time for set sizes (SS) three and five. (Smooth functions are derived from the estimated parameters of the exponential model fits presented in Table 4.)

Set Size Functions

Asymptotic performance varied as a function of set size (collapsed across serial position; see Figure 8). This difference in asymptotic accuracy (measured as the average of the longest four lags as before) was significant, $F(1, 3) = 136.9, p = .0013$. A 1λ–1β–1δ exponential model fit, which does not account for this significant difference, was poor for all subjects, with average $R^2 = .895$ ($R^2 = .854$ for the diffusion model). The parameter estimates and the resulting R^2 values for a model that allows different asymptotes for the two set sizes but assumes identical dynamics (2λ–1β–1δ exponential model or 2λ–1ν²–1δ diffusion model) are shown in Table 4. In Figure 8 we present the average observed *d'* values for the two set sizes, along with the descriptive functions for the exponential model in Table 4.

However, on the basis of the results of the serial position analysis, we expected a small dynamic difference favoring set size three. This is because the speeded case of immediate repetition represents one third of the average data for set size three, but only one fifth of the average data for set size five; that is, set size differences were expected on the basis of an averaging artifact. Whether the difference is large enough to detect in this experiment depends on the size of the immediate-repetition benefit. When the two set sizes were allowed independent rates, there was a modest but consistent difference, although R^2 improved for only 2 of the 4 subjects (BM and GR). For the exponential model, the rate for set size three was on average 13% faster than for set size five, and 3 of the 4 subjects showed a difference. (According to the diffusion equation, all 4 subjects showed faster intercept for set size three than for set size five, with the average difference of 50 ms.) An analysis of set size data from which we excluded the most recent serial position generally supported the averaging explanation, in that there were essentially no residual set size effects.

Summary

Set size and serial position have strong and systematic effects on asymptotic levels of accuracy. In both serial position and set size analyses, there was essentially no evidence to support the assertion that the size of the memory set affects retrieval dynamics. Retrieval does appear to be speeded for the most recent position, a finding that extends the related earlier observation of Wickelgren et al. (1980). There is no evidence in our data, nor was there evidence in the prior data of Wickelgren et al., to suggest an effect of recency on items in serial position −2 and earlier. Strictly speaking, we can conclude only that any such effect is quite small in comparison with the immediate repetition advantage. We argue that any effect of set size on retrieval dynamics is best attributed to differential weighting of the most recent serial position in the overall set size function. These results have obvious negative consequences for a number of models of memory retrieval. We postpone detailed descriptions of the consequences for memory models until after we describe results for different font and recent-negative manipulations contained in Experiments 2 and 3.

Table 4
Experiment 1: Set Size Parameter Estimates and Goodness of Fit

Parameters	Average	Subjects BM	GD	GM	GR
		Exponential equation			
Set size 3 λ	4.47	4.37	4.68	4.29	4.57
Set size 5 λ	3.52	3.32	3.79	3.27	3.72
Common β	4.28	4.76	3.66	5.04	4.06
Common δ	.339	.315	.328	.360	.361
R^2	.984	.980	.945	.977	.939
		Diffusion equation			
Set size 3 λ	5.40	4.84	5.49	4.86	5.33
Set size 5 λ	4.24	3.69	4.46	3.69	4.32
Common ν^2	.540	.264	.492	.317	.429
Common δ	.367	.381	.389	.399	.421
R^2	.954	.923	.911	.959	.946

Possible *d'* Correction Artifacts

However, before proceeding to those experiments, we raise and argue against various concerns about nearly perfect performance levels and the interaction with model estimation. For some subjects, accuracy levels, especially for serial position −1 in set size three at asymptote, required a correction factor to yield measurable *d'*s. In order to estimate *d'*s, a very small proportion of error was assumed when observed performance was in fact perfect. The small error probability was approximately half of what would have resulted from a single error. Functionally, this set a maximum on calculated *d'*s. If a true *d'* (had it been measurable in an extremely large sample size) were actually higher than the maximum allowed, then the maximum would have artifactually lowered asymptotic estimates for that condition and hence would have artifactually inflated dynamics estimates. It was precisely because of these concerns that we used set sizes of three and larger in our studies. There are two facts to notice about this possible artifact. First, the possible artifact would, if anything, have artificially introduced a dynamic preference for set size three, with its generally higher asymptotic levels; that is, the artifact would have worked against our finding.

Second, it is quite possible that this artifactual effect did inflate our estimates of the size of dynamic benefit for the most recent serial position. If we assume that the entire benefit for the most recent serial position was in asymptote and that the dynamics were identical to other set members, then the asymptotic levels for the most recent serial position would have had to be *d'* of 6 or more in comparison with the *d'*s for other serial positions of set size five in the 3.4–3.8 range. Our best guess is that the dynamic benefit for immediate repetition is perhaps only slightly overestimated by our model fits. This belief is based in Reed's (1976) set size data and Wickelgren et al.'s (1980) serial position data. Reed's memory list experiments differed from ours primarily in selection of shorter list lengths, longer retention intervals, and the use of a small confusable stimulus set (similar consonants). Because of the small confusable stimulus set, he rarely observed immeasurable *d'*s. We assert that the dependence of dynamics on set size in Reed's data is, as in our data, the consequence of averaging, which includes the (subjectively) most recent item (the case of his set size of one, of course, included only the most recent item, assuming that subjects rehearsed only members of the current list). Hence his data probably reflect true dynamics benefits for the most recent item with measurable asymptotic levels. The same pattern emerged in Wickelgren et al.'s (1980) data when more complex *d'* calculations were used to estimate high *d'*s. In that study, the most recent item differed in dynamics when *d'*s at asymptote were not artificially truncated.

We specifically did not use a consonant set like Reed's (1976) because it does not allow adequate control over the recency of negatives. In a small set, all possible items are relatively recent. We did try to lower the asymptotic levels for one subject (BM) by drawing each word set from the same category list. Although the words were in some sense similar, this, in fact, did little to affect the asymptotic levels. However, this experiment did replicate the findings of Experiment 1.

Experiment 2

The notable lack of a clear set size effect on retrieval dynamics and the coexistence of significant serial position effects argue against a serial exhaustive scan, as well as against some parallel scan models. A number of studies have demonstrated the coexistence of serial position effects and parallel set size functions that are normally associated with exhaustive serial scanning (e.g., Burrows & Okada, 1971; Monsell, 1978). However, it is sometimes suggested (Sternberg, 1975) that conditions that maximize serial position effects (fast presentation rates and short retention intervals) minimize the use of that exhaustive scanning. In the following RT experiment, we verified the coexistence of strong serial position effects with average data that are superficially compatible with an exhaustive scan explanation, in the same circumstances as our SAT experiment, in which the results excluded an exhaustive scan mechanism.

In addition, in this experiment and its complementary SAT version (Experiment 3), we examined two issues concerning the effects of recency. First, they eliminate the possibility that the retrieval advantage for probes drawn from from the most recent serial position is based on a direct perceptual or physical match between study and test form. In both experiments the case was altered between study list and test. Second, the recency of negatives was manipulated. A number of researchers report that negative probes that are either repeated negatives or members of a recently presented study set take longer and are less accurately rejected than other relatively nonrecent negatives (Atkinson et al., 1974; Monsell, 1978). Thus the retrieval mechanism must be responsive to items other than those in the immediate memory set. This suggests that item strength or familiarity is a salient factor in the decision process. In this experiment we replicated the "recent negative" effect in the RT domain, whereas in the subsequent SAT study, we examined the full time course of such effects.

Method

Subjects

Eighteen subjects participated in the study. All subjects participated in order to fulfill a introductory psychology course requirement.

Design and Stimuli

The stimuli were words drawn from the same two lists of 50 words as in the previous experiment. Set sizes of 3, 4, 5, and 6 words were presented in a varied set procedure. Positive probes tested each serial position within a set size equally often. In order to equate sampling from the respective serial positions, the number of trials with each set size varied directly with the number of serial positions. The number of negative probes equaled the number of positives for each set size. Negative probes were either recent or distant. Recent negatives (RNs) were drawn from the most recent three serial positions of the immediately previous trial but not the previous probe. Distant negatives (DNs) had last occurred at least three trials back.

In order to disallow a straightforward perceptual match between a positive probe and the most recently presented item, all study words were presented in lowercase letters. The test probes were presented in the uppercase format used in Experiment 1.

Procedure

With the exception of changing the case between study and test, all aspects of stimulus presentation were the same as described in Experiment 1. However, in this experiment, subjects were not interrupted with a cue to respond but rather were instructed to respond as quickly and as accurately as possible.

For each subject there was a total of 576 experimental trials, presented in four blocks of 144 trials each. This yielded a total of 16 positive trials for each serial position within the four set sizes. All subjects were given an initial practice session, consisting of 10 trials, to become acquainted with the procedure.

Results and Discussion

A trial was excluded if latency fell outside ±2.5 standard deviations of the subject's average positive or negative response time. Fewer than 2.3% of trials were excluded according to this criterion.

Serial Position Effects

The average correct RT and proportion of error for each serial position are presented in Table 5; RTs are plotted in Figure 9. Within each set size separately, the serial position of the probes significantly affected RT—$F(2, 34) = 14.53$, $p = .0000$, for set size three; $F(3, 51) = 26.9$, $p = .0000$, for set size four; $F(4, 68) = 21.4$, $p = .0000$, for set size five; and $F(5, 85) = 23.3$, $p = .0000$, for set size six—and the accuracy of responding, $F(2, 34) = 1.8$, $p = .18$, for set size three; $F(3, 51) = 4.86$, $p = .0048$, for set size four; $F(4, 68) = 11.32$, $p = .0000$, for set size five; and $F(5, 85) = 13.9$, $p = .0000$, for set size six. The serial position functions in Figure 9 were almost superimposed, although not as cleanly as those reported by Monsell (1978). All set sizes showed a dramatic recency component; mean RTs for serial position −1 differed at most by 7 ms. The primacy effect is somewhat less consistent.

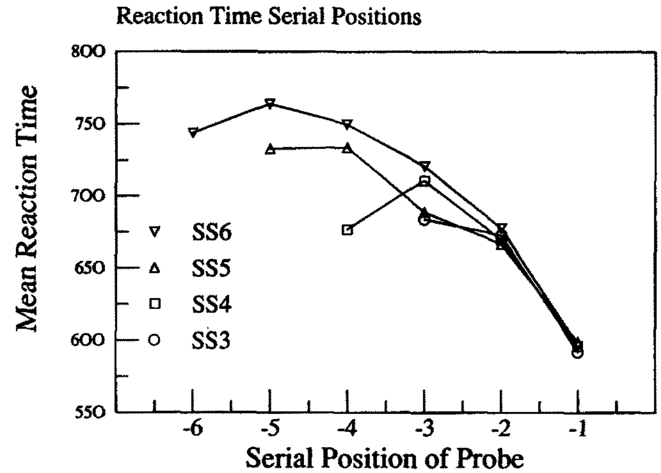

Figure 9. Average correct positive reaction time as a function of serial position for set sizes (SS) three through six. (Serial position is labeled in terms of recency, counting backwards from the test item, position 0, to the study position of the probe, −1 for the most recent serial position, −2 for the next, and so on.)

Set Size Functions

In Table 6 we present the average correct RT and proportion of error for each set size by probe type. In Figure 10 we present these RTs along with the fitted regression lines. Not surprisingly, RT significantly increased with larger set sizes, $F(3, 51) = 17.2$, $p = .0000$. Trial type also significantly affected RT, $F(2, 34) = 31.7$, $p = .0000$; positive responses were faster than distant and recent negatives. With a regression model, the estimated intercept for positive responses was 599 ms, in comparison with 671 ms for distant negatives and 689 ms for

Table 5
Experiment 2: Average Reaction Times (RT) and Proportion Errors (PE)

	Serial position											
	−6		−5		−4		−3		−2		−1	
Set size	RT	PE	RT	PE	RT	PE	RT	PE	RT	PE	RT	PE
6	744	.35	764	.30	750	.28	721	.13	678	.10	596	.05
5			733	.23	734	.26	689	.14	667	.08	599	.06
4					677	.13	711	.15	670	.08	596	.06
3							684	.09	673	.06	592	.05

Table 6
Experiment 2: Average Reaction Times (RT) and Proportion Errors (PE)

	Set size							
	3		4		5		6	
Probe type	RT	PE	RT	PE	RT	PE	RT	PE
Positive	649	.06	662	.10	680	.15	699	.21
Distant negative	703	.05	739	.06	734	.06	750	.02
Recent negative	750	.09	777	.15	808	.26	814	.26

recent negatives. However, the regression lines in Figure 10 illustrate that the size of the memory set had equal impacts on all three types of trials, $F(6, 102) = 1.53$, *ns*, resulting in approximately equal slope estimates: specifically, 16 ms/item for positive responses, 13 ms/item for distant negatives, and 22 ms/item for recent negatives. This similarity of slopes further corroborates the reports of Burrows and Okada (1971) and Monsell (1978) that strong serial position effects are not incompatible with linear parallel set size functions.

The recency of the negative probe significantly increased rejection time, $F(1, 17) = 41.4$, $p = .0000$, and false alarms, $F(1, 17) = 43.1$, $p = .0000$. On average, recent negatives (members of the previous memory set) increased latencies by 50 ms and false alarm rates by 13% in comparison with distant negatives (studied at least three trials back). This less detailed replication of Monsell's (1978) finding demonstrates linear parallel RT as a function of set size coexistent with the influence of factors not easily explained within a strict serial exhaustive scan model.

Experiment 3

In this experiment, the time course of responses to recent and distant negatives were compared. Monsell (1978), and others, demonstrated that recent exposure to a negative probe

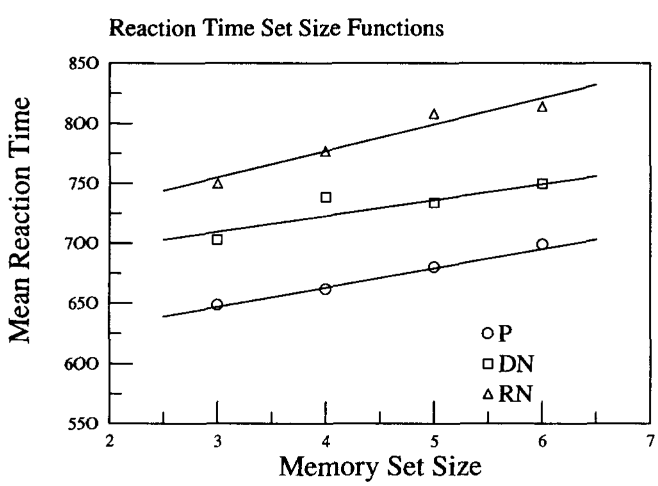

Figure 10. Average correct positive and negative reaction time as a function of set size (three through six). (Smooth functions represent the best fitting linear regression lines.)

substantially impairs the speed and accuracy of rejection. Rejection of recent negatives may be hindered by residual strength or familiarity stemming from prior study. Do recent negatives function in the same way as low-strength positives throughout the full time course of retrieval, or are false positives corrected by more specific list information? Investigating the time course of processing for recent negatives was the primary focus of this SAT experiment. In addition, the SAT set size and serial position functions in this experiment were directly compatible with those of the previous RT experiment. Hence this is a replication of Experiment 1 but with recent and distant negatives and for set sizes four and six.

Method

Subjects

Four subjects each completed eight 1½-hr sessions. Of these subjects, 3 (BM, GM, and GR) had participated in Experiment 1. The 4th (LB), a new subject, was given an initial 1-hr training session in the SAT methodology.

Design and Stimuli

With two exceptions, the design and stimuli were the same as in Experiment 1. First, set sizes of four and six words were tested. Second, negative probes within each set size were either recent (RN) or distant (DN) negatives. Recent negatives were drawn from the most recent four serial positions of the previous trial but were not the previous recognition probe. On average, recent negatives occurred 7.5 items back for set size four and 9.5 items back for set size six. Distant negatives last occurred at least three trials back.

Procedure

The procedure was the same as in Experiment 1 with two exceptions. First, as in Experiment 2, all study words were presented in lowercase letters, whereas test probes were presented in an uppercase format. Second, the lag of the cue to respond was adjusted slightly: Subjects were cued to respond at 0.1, 0.25, 0.35, 0.45, 0.60, 0.90, 1.3, or 1.8 s after the onset of the test probe.

Each of the eight sessions consisted of 640 trials divided into four blocks. Across the eight sessions, this yielded a total of 32 trials per lag for each serial position within set size. In total, there were 64 trials per lag for each of the two types of negative trials associated with set size four and 96 trials per lag for those associated with set size six.

Results and Discussion

The latency and proportion correct for positive and negative trials and the associated d's are presented in Table 7 for the average over subjects. These measures are presented for the two set size conditions (four and six) and for each serial position within set size.

Interruption (Lag) Latency Analysis

As with Experiment 1, latency was longer for earlier interruption points for both the set size, $F(7, 21) = 26.1$, $p = .0000$, and serial position functions, $F(7, 21) = 31.35$, $p = .0000$, for

Table 7
Experiment 3: Average Latency (in s), Proportion Correct, and d′

Item	Lag 1	2	3	4	5	6	7	8
			Set size 4					
Positive								
Latency	0.260	0.211	0.191	0.187	0.184	0.179	0.173	0.177
Proportion correct	0.598	0.772	0.840	0.887	0.933	0.962	0.970	0.960
Distant Negative (DN)								
Latency	0.269	0.231	0.205	0.186	0.182	0.181	0.180	0.166
Proportion correct	0.575	0.762	0.895	0.965	0.978	0.970	0.967	0.985
Recent negative (RN)								
Latency	0.268	0.233	0.212	0.197	0.187	0.184	0.177	0.168
Proportion correct	0.477	0.735	0.760	0.865	0.892	0.920	0.940	0.942
Positive vs. DN: *d′*	0.469	1.510	2.311	3.100	3.541	3.729	3.788	4.007
Positive vs. RN: *d′*	0.204	1.401	1.732	2.388	2.793	3.349	3.455	3.522
DN vs. RN: *d′*	0.266	0.108	0.579	0.711	0.748	0.380	0.333	0.484
Serial position −4								
Latency	0.268	0.230	0.199	0.188	0.183	0.180	0.175	0.168
Proportion correct	0.502	0.707	0.803	0.832	0.918	0.940	0.962	0.923
d′	0.213	1.299	2.361	2.845	3.570	3.601	3.728	3.615
Serial position −3								
Latency	0.267	0.225	0.201	0.189	0.183	0.180	0.177	0.172
Proportion correct	0.490	0.630	0.702	0.860	0.860	0.940	0.940	0.940
d′	0.220	1.099	1.830	3.026	3.109	3.565	3.572	3.793
Serial position −2								
Latency	0.267	0.213	0.199	0.187	0.183	0.184	0.174	0.182
Proportion correct	0.555	0.798	0.900	0.870	0.957	0.980	0.983	0.973
d′	0.361	1.625	2.589	3.033	3.873	3.934	4.016	4.093
Serial position −1								
Latency	0.260	0.211	0.193	0.183	0.182	0.178	0.177	0.175
Proportion correct	0.853	0.955	0.950	1.000	1.000	0.990	1.000	1.000
d′	1.323	2.527	3.147	4.202	4.347	4.078	4.229	4.449
			Set size 6					
Positive								
Latency	0.263	0.220	0.200	0.191	0.189	0.184	0.179	0.173
Proportion correct	0.567	0.640	0.735	0.800	0.835	0.842	0.863	0.882
Distant negative (DN)								
Latency	0.266	0.231	0.205	0.191	0.188	0.184	0.178	0.180
Proportion correct	0.502	0.760	0.873	0.923	0.930	0.945	0.960	0.950
Recent negative (RN)								
Latency	0.266	0.234	0.215	0.201	0.192	0.186	0.178	0.176
Proportion correct	0.460	0.635	0.728	0.803	0.808	0.842	0.840	0.850
Positive vs. DN: *d′*	0.183	1.082	1.819	2.301	2.498	2.670	2.981	2.977
Positive vs. RN: *d′*	0.083	0.720	1.278	1.743	1.914	2.052	2.168	2.352
DN vs RN: *d′*	0.166	0.248	0.539	0.614	0.645	0.611	0.714	0.543
Serial position −6								
Latency	0.265	0.227	0.209	0.192	0.187	0.185	0.175	0.175
Proportion correct	0.490	0.472	0.618	0.723	0.750	0.760	0.790	0.805
d′	0.097	0.646	1.476	2.037	2.199	2.405	2.699	2.724
Serial position −5								
Latency	0.264	0.233	0.210	0.198	0.193	0.184	0.183	0.173
Proportion correct	0.472	0.517	0.533	0.635	0.715	0.683	0.712	0.695
d′	0.231	0.762	1.294	1.867	2.088	2.130	2.455	2.162
Serial position −4								
Latency	0.267	0.227	0.207	0.194	0.190	0.185	0.173	0.178
Proportion correct	0.472	0.600	0.623	0.705	0.743	0.762	0.817	0.865
d′	0.221	0.982	1.489	1.987	2.211	2.388	2.870	2.980
Serial position −3								
Latency	0.266	0.228	0.204	0.190	0.189	0.186	0.179	0.178
Proportion correct	0.582	0.570	0.752	0.817	0.848	0.930	0.910	0.935
d′	0.217	0.893	1.906	2.352	2.551	3.165	3.372	3.456
Serial position −2								
Latency	0.263	0.227	0.199	0.187	0.186	0.183	0.178	0.177
Proportion correct	0.595	0.712	0.892	0.942	0.970	0.923	0.957	1.000
d′	0.254	1.311	2.452	3.193	3.378	3.096	3.668	4.031
Serial position −1								
Latency	0.261	0.214	0.191	0.186	0.184	0.180	0.176	0.182
Proportion correct	0.785	0.945	1.000	1.000	1.000	0.990	0.990	1.000
d′	0.835	2.328	3.491	3.755	3.810	3.832	3.997	4.031

set size four and $F(7, 21) = 23.2$, $p = .0000$, for set size six. No other latency differences were apparent in the set size data. Within each set size, however, latency varied with serial position, $F(3, 9) = 3.81$, $p = .052$, for set size four and $F(5, 15) = 7.62$, $p = .001$, for set size six. The effect of serial position interacted with interruption point, $F(21, 63) = 2.29$, $p = .006$, for set size four and $F(35, 105) = 1.74$, $p = .017$, for set size six. This interaction is primarily due to the most recent serial positions' producing shorter latencies at the early interruption points. Once again, these small differences are included in dynamics by the indexing of accuracy as a function of total processing time.

Serial Position Functions

In Figure 11 we plot d', averaged across subjects and the last four interruption points, as an empirical measure of asymptotic accuracy for each serial position. Serial position of the probe significantly affected asymptotic recognition accuracy for both set sizes, $F(3, 9) = 13.1$, $p = .0012$, for set size four and $F(5, 15) = 24.4$, $p = .0000$, for set size six. Asymptotic levels for identical serial positions, shown in Figure 11, did not overlap for the two set size functions. As in Experiment 1, each serial position within the two set sizes required a separate asymptote estimate (λ; average $R^2 = .929$ for the exponential model, and average $R^2 = .915$ for the diffusion model). Fits that did not account for asymptotic differences were poor (average $R^2 = .690$ for the exponential model, and average $R^2 = .673$ for the diffusion model).

Models assuming various dynamic differences were tested. As in Experiment 1, allotting a separate dynamic parameter to probes drawn from the most recent serial position, irrespective of set size, and a single dynamic parameter to all of the remaining serial positions (a 10λ–2β–1δ exponential model or, alternatively, a 10λ–$1\nu^2$–2δ diffusion model) produced the best overall fit. The parameter values and R^2s for the best exponential and diffusion models are shown in Table 8 for the average data and each individual subject. In Figure 12 we show the average d' values for each serial position within set size. Smooth functions are generated by the exponential model in Table 8.

In the exponential model, the average rate parameter for the most recent serial position (immediate repetition) was 52% faster than for other positions. All 4 subjects showed this pattern; differences ranged from 45% to 55%. (In the diffusion model this pattern appeared in the fitted intercept. The intercept of the most recent serial positions was 103 ms earlier than for other probes in the average data. All subjects showed this pattern; differences ranged from 102 to 116 ms.)

Retrieval dynamics did not appear to differ for serial positions beyond the most recent. When each serial position was allowed an independent rate parameter, we observed a marked discontinuity between the most recent serial position and all other positions and little difference between other serial positions. This was a replication of the pattern of Experiment 1. The average exponential rate parameters (βs) were estimated at 7.95, 4.50, 3.57, and 4.25 for set size four and 8.95, 4.75, 3.27, 3.33, 4.01, and 3.44 for set size six, in order of most recent to most distant. (For the diffusion model, the intercept parameters showed the equivalent pattern.) As in Experiment 1, the dynamics for serial positions from different set sizes (excluding the most recent) differed little if at all when fit with either model.

The most parsimonious model is the one applied to the data in Experiment 1 with estimates for this experiment shown in Table 8. Altering case between study and test did little to attenuate either the differences in asymptotic accuracy levels for different serial positions or the faster rate (or intercept) for probes from the most recent serial position. A physical or perceptual match apparently contributes little to the effect of immediate recency.

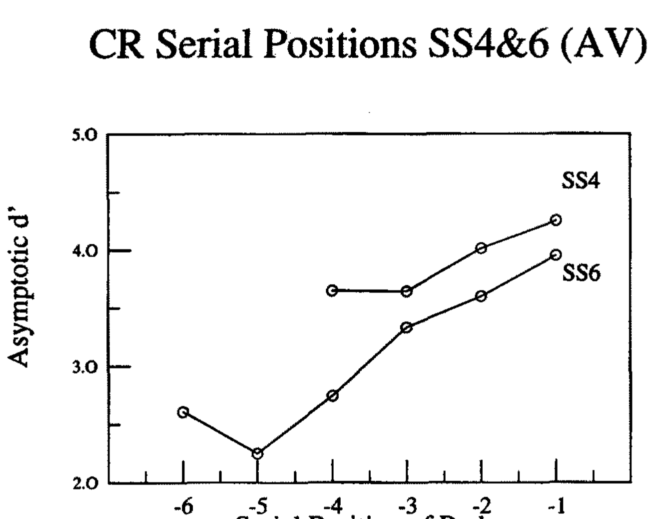

Figure 11. Observed average asymptotic d' values for serial positions within set sizes (SS) four and six. (Serial position is labeled in terms of recency, counting backwards from the test item, position 0, to the study position of the probe, −1 for the most recent serial position, −2 for the next, and so on.)

Set Size Functions

In Figure 13 we plot the average d' versus total processing time pooled over serial position for each set size. These functions base d' on distant negatives only. Asymptotic accuracy of set sizes four and six were significantly different, $F(1, 3) = 47.59$, $p = .006$, when performance on the last four lags was used as the measure of asymptote. When asymptote differences were ignored, model fits were relatively poor (average $R^2 = .843$ for the exponential model, and average $R^2 = .840$ for the diffusion model). The exponential model that allows asymptotic differences (2λ–1β–1δ) and the related diffusion model (2λ–$1\nu^2$–1δ) are shown in Table 9. These models fit the data of all the subjects quite well (average data $R^2 = .983$ for the exponential model, and average $R^2 = .978$ for the diffusion model).

In Experiment 1, there was an effect of set size on retrieval dynamics that we attributed to averaging the retrieval advantage for the most recent serial position into the overall set size functions. In this experiment, allowing separate rates (or intercepts) for each set size neither improved R^2 nor yielded

Table 8
Experiment 3: Serial Position Parameter Estimates and Goodness of Fit

Parameters	Average	Subjects BM	GM	GR	LB
		BM	GM	GR	LB
Exponential equation					
Set size 4 λs					
SP −1	4.29	4.39	4.63	4.28	3.97
SP −2	4.20	4.21	4.52	4.42	3.80
SP −3	3.66	3.81	3.61	4.18	3.21
SP −4	3.85	4.28	3.86	4.02	3.18
Set size 6 λs					
SP −1	3.98	3.56	4.21	4.30	4.02
SP −2	3.86	3.69	3.58	4.41	3.74
SP −3	3.29	3.26	2.81	3.63	3.47
SP −4	2.75	3.08	2.42	3.04	2.41
SP −5	2.34	2.72	2.07	2.45	2.09
SP −6	2.63	2.89	2.39	3.05	2.18
Recency β	8.34	10.0	10.0	7.75	6.54
Other β	3.96	4.46	5.52	4.05	3.13
Common δ	.339	.303	.377	.378	.333
R^2	.961	.889	.937	.869	.926
Diffusion equation					
Set size 4 λs					
SP −1	4.82	4.75	5.02	4.73	4.64
SP −2	4.55	4.35	4.86	4.72	4.16
SP −3	3.97	3.95	3.90	4.51	3.52
SP −4	4.12	4.44	4.16	4.35	3.47
Set size 6 λs					
SP −1	4.47	3.85	4.57	4.73	4.68
SP −2	4.13	3.85	3.87	4.76	4.09
SP −3	3.52	3.40	3.03	3.90	3.79
SP −4	2.94	3.21	2.61	3.26	2.63
SP −5	2.49	2.83	2.23	2.61	2.28
SP −6	2.82	3.01	2.58	3.28	2.38
Common ν^2	.207	.128	.173	.203	.317
Recency δ	.348	.314	.344	.387	.346
Other δ	.452	.431	.446	.493	.456
R^2	.964	.895	.950	.852	.930

Note. SP = serial position.

a reliable pattern of parameter values. This minor discrepancy between SAT studies is plausibly related to the expected size of the averaging difference. With the larger set sizes in this experiment (four and six), the differential contribution of the most recent serial position was expected to be smaller and apparently was not sufficient to yield as clear a dynamic difference as did the smaller set sizes (three and five) of Experiment 1. In this experiment, there was no support for overall set size effects on retrieval dynamics. Only terminal accuracy levels of the two set sizes appeared to differ.

Recent and Distant Negatives

We indexed the differences between recent and distant negatives by two methods of both computing d' measures and fitting the empirical SAT functions. We computed two sets of conventional d's, one that entailed using false alarms from recent negatives and one that entailed using false alarms from distant negatives (those quoted earlier) within set size. The hit rates in both cases pooled over the serial position of positive probes within set size. These yield the positive versus recent negative (P vs. RN) and positive versus distant negative (P vs. DN) SAT functions shown in Figure 14. In Figure 14a we show set size four, and in 14b we show set size six. The d's based on recent negatives yielded significantly lower asymptotic performance, $F(1, 3) = 29.6$, $p = .012$, for set size four and $F(1, 3) = 205$, $p = .0007$, for set size six, when the last four lags were used as a measure of asymptote. The best simple exponential model of the two upper curves of both panels varied asymptote for all four conditions and rate with the kind of negative (4λ, 2β, and 1δ). In addition to lower asymptotes, recent negative rate parameters were on average 20% slower than for distant negatives (average $R^2 = .986$). Recent negatives slowed rate for all subjects, with estimated decrements of 15%–25%. (Fits for which we used the diffusion equation yielded a similar, though slightly larger, decrement

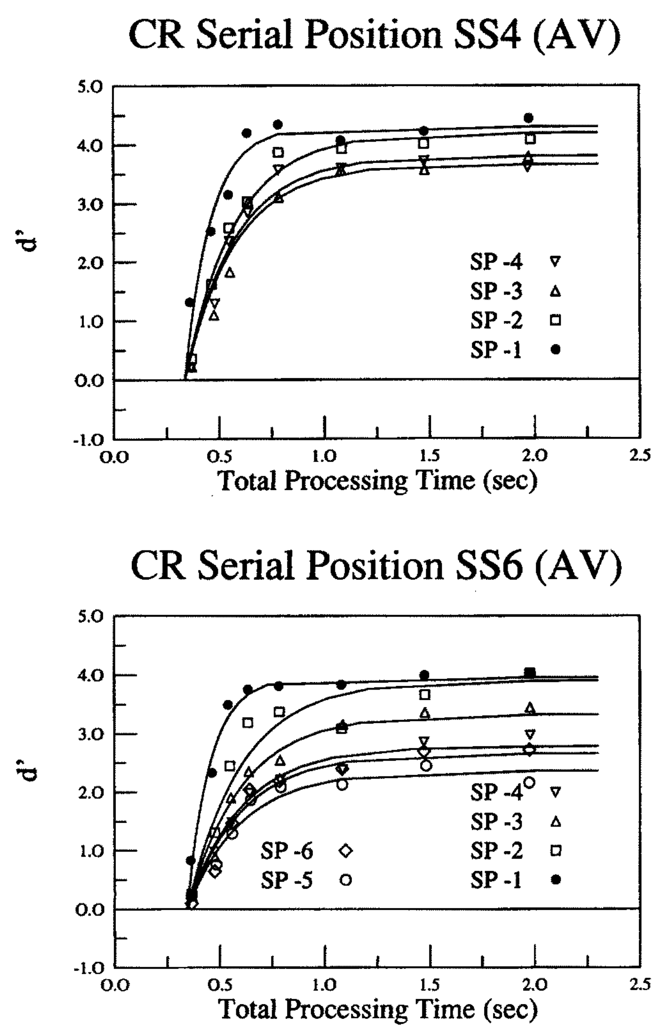

Figure 12. Observed average d' values as a function of total processing time for serial positions within set size four (SS4; top panel) and set size six (SS6; bottom panel). (Smooth functions are derived from the estimated parameters of the exponential model fits presented in Table 8.)

in the rate associated with recent negatives, averaging 38% and ranging from 25% to 47% across subjects.)

The asymptotic difference between recent and distant negatives reflected higher false alarm rates for recent negatives. This was the SAT equivalent to the RT results of Experiment 2. The difference in the rate parameters reflected differences in false alarms to recent and distant negatives that are not constant over the time course of retrieval. To directly examine the differences in false alarms for recent and distant negatives as a function of lag, a pseudo-d' was computed from $z_{FA(RN)} - z_{FA(DN)}$. Increased "yes" responses to recent negatives, in comparison with distant negatives, yielded a positive d'. Larger pseudo-d' values represented poorer performance; however, this scaling is consistent with increased strength of recent negatives. The RN versus DN d's form the lower curves in the two panels of Figure 14.

The RN versus DN functions both tended to be nonmonotonic, although this trend was more apparent for set size four than six. The functions showed a rise to an early, relatively high level of false alarms (300–900 ms), followed by a decrease and/or eventual leveling off of the false alarm rate (900–2,000 ms). The rate differences in the P versus DN and P versus RN functions, which we formed by scaling positives against either recent or distant negatives, were a consequence of the early tendency for false alarm followed by later correction. These results suggest that information accrual in recent negative probes shifted in midretrieval from a fairly high strength value to a corrected strength value. The terminal accuracy difference between recent and distant negatives represents an imperfect correction process.

In the diffusion model, a two-phase or two-process notion assumes that the resonance driving information accrual is altered at some point $t^* > \delta$. The shift represents the availability of new information. Before t^*, the simple diffusion Equation 4a holds. After t^*, Equation 4b holds (Ratcliff, 1980):

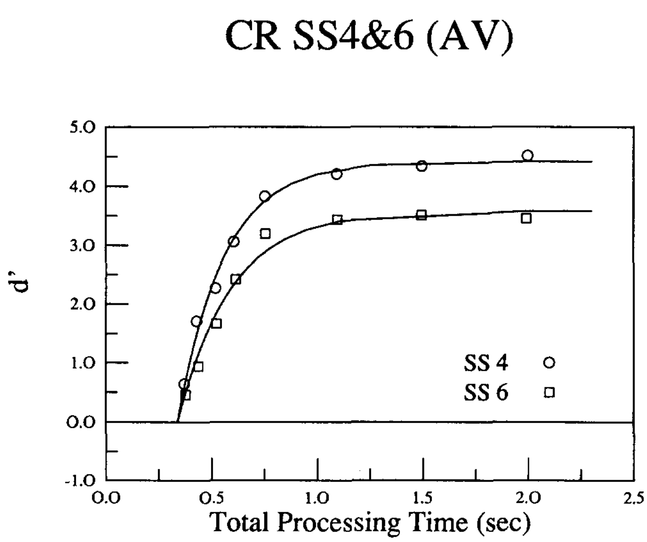

Figure 13. Observed average d' as a function of total processing time for set sizes (SS) four and six. (Smooth functions are derived from the estimated parameters of the exponential model fits presented Table 9.)

Table 9
Experiment 3: Set Size Parameter Estimates and Goodness of Fit

Parameters	Average	BM	GM	GR	LB
		Subjects			
	Exponential equation				
Set size 4 λ	3.95	4.15	4.08	4.22	3.35
Set size 6 λ	2.93	3.13	2.73	3.22	2.64
Common β	4.49	4.93	4.69	4.44	4.05
Common δ	.349	.311	.348	.387	.351
R^2	.983	.955	.969	.907	.976
	Diffusion equation				
Set size 4 λ	4.24	4.46	4.32	4.54	3.65
Set size 6 λ	3.15	3.37	2.89	3.47	2.89
Common ν^2	.188	.175	.161	.196	.240
Common δ	.452	.412	.448	.489	.456
R^2	.977	.931	.986	.904	.969

$$d'(t) = \frac{\lambda_1}{\sqrt{1 + \nu^2/(t - \delta)}}, \text{ for } \delta < t < t^*; \quad (4a)$$

$$d'(t) = \frac{\lambda_2 + (\lambda_1 - \lambda_2)(t^* - \delta)/(t - \delta)}{\sqrt{1 + \nu^2/(t - \delta)}}, \text{ for } t \geq t^*. \quad (4b)$$

Here, λ_1 is the asymptotic level being approached by the first phase or process, and λ_2 is the asymptotic level of the second process. Exponential analogues to these equations are shown in Equations 5a and 5b:

$$d'(t) = \lambda_1(1 - e^{-\beta(t-\delta)}), \text{ for } \delta < t < t^*; \quad (5a)$$

$$d'(t) = [\lambda_2 + (\lambda_1 - \lambda_2)(t^* - \delta)/(t - \delta)] \times (1 - e^{-\beta(t-\delta)}), \ t > \delta \text{ for } t \geq t^*. \quad (5b)$$

Although they yielded reasonable R^2s for average and individual subject data, single-process models systematically underestimated performance at the early lags (500–800 ms) and, to a lesser extent, overestimated performance at subsequent lags. Two-process models yielded slightly better goodness of fit for the average data and for each subject individually (average $R^2 = .988$ for the exponential form and average $R^2 = .984$ for the diffusion form, in comparison with .983 and .976 for single-process forms). The average λ_1s were 0.90 and 0.83 for set sizes four and six, and the respective average λ_2s were 0.30 and 0.59. For all subjects, estimated λ_1 exceeded λ_2 by 0.50 to 1.66 d' units for set size four and by 0.13 to 0.93 d' units for set size six. (For the diffusion model, the average λ_1s were 1.05 and 0.99 for set sizes four and six and the respective average λ_2s were 0.22 and 0.59; $\lambda_1 > \lambda_2$ by 0.672 to 1.62 d' units for set size four and by 0.22 to 1.35 d' units for set size six.) The different λ_1 and λ_2 values accommodated the temporal pattern of the data seen in Figure 14. Set sizes four and six may have differed in the degree of correction, but more data would be required to substantiate this claim.

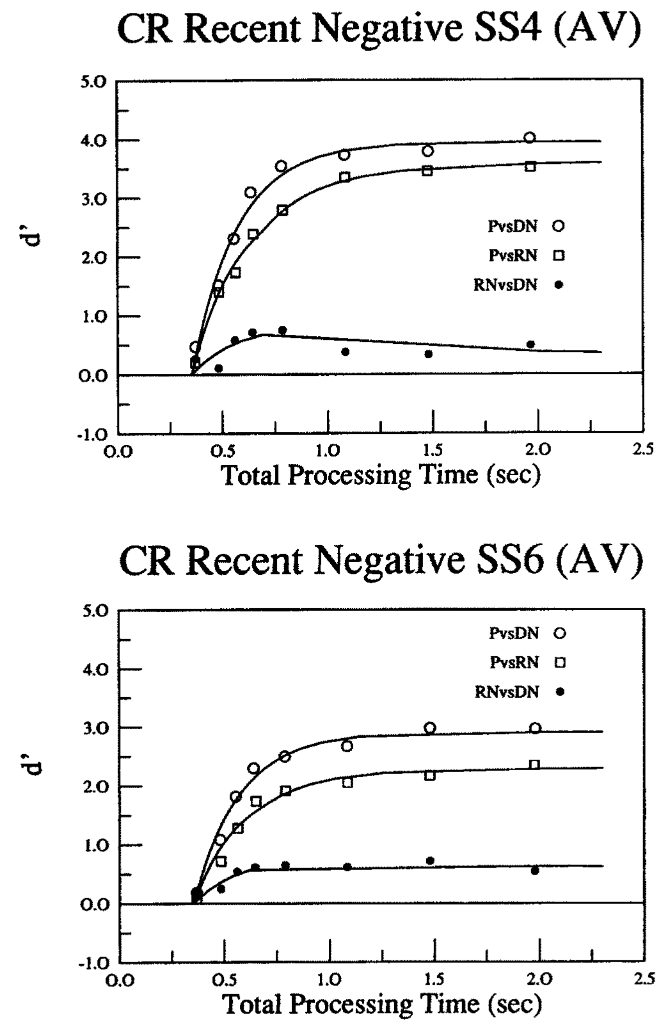

Figure 14. Observed average *d'* values for positives versus recent negatives (PvsRN), positives versus distant negatives (PvsDN), and recent versus distant negatives (RNvsDN) for set sizes (SS) four (top panel) and six (bottom panel). (Smooth functions are derived from the estimated parameters of the dual-process exponential model fits.)

Summary

In this experiment we replicated (with set sizes four and six and font mismatch) the findings of Experiment 1 (with set sizes three and five and font match) concerning retrieval dynamics of set size and serial position. Again, we found that retrieval dynamics for the most recent serial position, or immediate repetition between last list element and test, were speeded in relation to other serial positions. No additional set size effect was found. In addition, recent negatives were more difficult to reject throughout retrieval and especially so early in retrieval. This pattern was captured by a two-process model for recent negatives with relatively high initial strength followed by incomplete correction. The two-process equations assume that the correction occurs quantally at t^*, but a continuous correction variant would likely fit our data equally well. Correction must be a consequence of more specific features or information accruing relatively late in retrieval. Two-process models do not place strong constraints on what information may serve to correct a residual strength impression. The form is consistent with any plausible candidate, such as the accrual of contextual information, temporal tags, or list tags.

General Discussion

Asymptotic Accuracy

Overall error rates at asymptote were less than 10%, the level of performance often assumed to minimize differences in retention. Nonetheless, asymptotic accuracy of SAT functions varied with serial position, producing typical bowed functions. Asymptotic accuracy levels were also lower for larger set sizes. The coexistence of serial position and set size effects suggests that the latter are perhaps derived from the former (Monsell, 1978; Murdock, 1971, 1985).

There is an apparent difference between our asymptotic accuracy results, in which serial positions were not superimposed, and the RT data, in which they were approximately superimposed (save the primacy position). In keeping with early strength models of memory (Murdock & Dufty, 1972; Norman & Wickelgren, 1969), RT may reflect asymptotic strength differences alone. In Dosher's (1982) study, RT was strongly correlated (within each subject) with differences in asymptotic strength of SAT functions when dynamics differences were small. Some studies in the RT domain, like ours, have shown patterns of serial position overlap; others have not (i.e., Murdock & Franklin, 1984). However, small deviations from superimposition in our RT data were in the same direction as the asymptotic accuracy difference in SAT. Moreover, the RT data represent a subset of data contributing to SAT accuracy. RT serial position functions represent performance on correct positive trials in isolation. The asymptotic accuracy measures represent performance on all positive and negative trials. In the SAT data, performance on comparable positives is very similar but favors small set sizes by 1%–3% in hits; negative trials for small set sizes also show 1%–5% fewer false alarms in relation to larger set sizes. This may entirely reflect differences between either positives or negatives alone, partly compensated by set-size-dependent criterion shifts. In either event, the large differences in RT or asymptote between set sizes reflect differences in noncomparable serial positions (primacy positions and positions not represented in shorter set sizes). The major effects in both paradigms reflect primacy and recency, for which the analogous serial position patterns bear striking similarity.

Murdock (1985) accounted for RT data entirely within a trace-strength framework. Serial position functions were modeled by two strength parameters, one representing a recency component and the other a primacy component. By judicious selection, recency and primacy in turn generated overall set size effects. Although the comparable serial positions in our data contributed slightly to the overall set size difference in SAT asymptote (i.e., were slightly nonoverlapping), the difference largely reflects the performance in noncomparable serial positions. Nonetheless, a simple, two-parameter primacy and recency model is unlikely to handle the entire pattern of asymptotic accuracies. The size of the memory set was correlated with small residual differences in the asymptotic accuracy levels of comparable serial positions. We outline, as follows, an alternative strength model proposed by Wickelgren and Norman (1966) that adequately accommodates our observed asymptotic patterns.

Retrieval Dynamics

Though not predicted, differences in asymptotic performance certainly can be accommodated by pure serial exhaustive scan models such as Sternberg's (1966, 1969) or parallel self-terminating models such as Murdock's (1971). Each comparison may have some probability of error that is contingent on set size and/or serial position (see Treisman & Doctor, 1987), or items may be probabilistically represented in the scan. More critical for the assessment of these models is the evidence that retrieval dynamics do not vary with either set size or serial position, except for the most recent position. Rather dramatic dynamic differences in the empirical SAT functions, over and above any asymptotic differences, are the predicted signatures of exhaustive or self-terminating scan models. In fits with both the diffusion and the exponential retrieval equations, we consistently found essentially equal dynamics for different overall set size functions and all serial positions save the most recent. This dynamic invariance strongly argues against scanning explanations of retrieval from short-term list memory in our SAT experiments and, by extension, in similar RT experiments.

Invariance of retrieval dynamics coupled with the observed systematic asymptotic effects strongly support a simple strength (or probability) mechanism coupled with a simple retrieval mechanism. We describe a modified variant of a diffusion model (Ratcliff, 1978) as follows. However, the results are compatible with any direct-access or comparison retrieval model in which the explicit memory set does not strongly control performance.

However, neither Ratcliff's (1978) account nor any other current account of Sternberg's (1966) paradigm predicts the rather dramatic retrieval advantage for probes drawn from the most recent serial position that emerged in both SAT experiments. This finding replicated, in a subspan domain, Wickelgren et al.'s (1980) result in superspan lists. Probes drawn from the most recent serial position are in effect cases of immediate repetition between study and test. Wickelgren et al. suggested that the last study item remains in an active or primed state unless interfering mental activity intervenes between study and test. Recognition of an item in an active state may be tantamount to a matching procedure if normal retrieval operations are circumvented. (It may be possible to alter models to accommodate this finding: For example, in Ratcliff's 1978 diffusion model, one might assume a different variance for cases of immediate repetition.)

The results of Experiment 3 suggest that the immediate match benefit can be mediated by a relatively abstract level of representation, at least to the degree that it is not case sensitive (cf. Posner et al., 1969). The immediate match benefit in our experiments survived the font-mismatch manipulation. Our best guess is that the speeding of retrieval for the last list element is of approximately the same magnitude with and without font match. An examination of the data for the 3 subjects who served in both Experiments 1 and 3 showed remarkable similarity in estimated dynamic parameters. When both sets of data were fit at the same time for those subjects, the immediate-repetition parameters were essentially the same. Without doubt, it would be possible to construct a display situation in which font-based matches were important. In our experiments, however, the lexical identity of the last list element appears to be the critical factor.

The retention interval in our experiments was quite short (300 ms). However, an immediate repetition retrieval advantage appeared to persist for a substantially longer period of time. Wickelgren et al. (1980) observed it with retention intervals of 700 ms. Moreover, Dosher (1981) observed a similar advantage in paired-associate recognition that survived 3 s of counting backwards but not 3 s of interpolated learning. This latter result, if applicable to item recognition, suggests that the critical determinant of this advantage is not so much time or decay of activation per se but rather the nature and similarity of the interpolated mental activity. Sternberg (1966, 1969) reported no serial position effects with a retention interval of more than 1 s and explicit instructions to rehearse the members of the study list. Hence we must suppose that covert rehearsal behaves in somewhat the same way as overt presentation with respect to serial position effects, decoupling external and internal recency.

Last, we attribute observed dynamics differences in set size data that include the most recent serial position to the different proportions of immediate repetition cases included in the pooling. Reed's (1976) larger pooled differences among set sizes one, two, and four, in comparison with our contrasts of three with five and four with six, are expected under the proportional pooling argument. One in one versus one in two trials yields a larger difference than one in three versus one in five or one in four versus one in six. The estimated benefit for the last list element in Reed's confusable consonant lists is not inconsistent with the estimated benefit in our word lists.

Recent Negatives

Ratcliff (1978) and Monsell (1978), among others, pointed out that recent negative effects are compelling evidence against a pure scan or search account of immediate item recognition. The presence of such effects indicates that recognition procedures are responsive to items outside the experimenter-defined memory set. Our experiments provide further corroborating evidence for this claim, showing that subjects reject significantly more slowly (Experiment 2) and are more likely to give false alarm responses to (Experiment 3) a recently presented negative probe.

Subsequent analysis of the time course for these responses demonstrated a biphasic false alarm rate. During an initial retrieval period (100–900 ms), recent negatives elicited a high false alarm rate that was compatible with susceptibility to residual trace strength. Later in retrieval (more than 900 ms), the false alarm rate was modestly attenuated, which indicates the late impact of more list-specific information. We treated this pattern as a shift from uncorrected to corrected strength late in retrieval. Two-phase information accrual may be regarded by some as evidence for two distinct types of retrieval operations—for example, a mixture of judgments based on direct access and a slow serial scan of the list (Atkinson & Juola, 1974). However, if serial scanning was the source of late correction in recent negatives, then the same mechanism should produce concomitant effects in positive responses. We found no evidence that retrieval dynamics varied according to the pattern predicted by standard versions of serial scans. Consequently, a theoretically consistent treatment of both negative and positive response patterns requires a more neutral view of the late correction mechanism.

Random Walk With Search Set

Our findings render most scanning models of short-term memory retrieval extremely unlikely and place strong constraints on other models. In this section, we consider the direct-access accumulation model of Ratcliff (1978) in some detail, as one possible short-term memory mechanism. According to the diffusion model, memory retrieval is a set of independent, direct-access comparison operations. A positive response follows when any comparison (in the search set) yields a match, and a negative response follows when all comparisons (in the search set) yield a mismatch. Reaction time is affected by the resonance (strength) of the dominating comparisons and by their number. In SAT, asymptote is determined by the mean and variance of the resonance (strength) distributions, and rate is controlled by item variance and intrinsic drift noise. Intercept is a free parameter in accounting for either RT or SAT data. This model is compatible with pure strength control of RT or with a more complex set of factors controlling RT.

As outlined previously, Ratcliff (1978) demonstrated that simple changes in a resonance parameter (the diffusion model's equivalent of a strength parameter), coupled with implicit set size variations, are sufficient to accommodate either set size or serial position RT effects. (Ratcliff, 1978, did not model both effects simultaneously as suggested by Murdock, 1971, 1985. Instead, he estimated patterns of the relatedness parameter for each effect separately, without attempting to derive positive set size functions from serial position effects.) Consider Ratcliff's (1978) treatment of Reed's (1976) SAT data. In this treatment he estimated an intercept, a rate, and an asymptote from set size one data, which yielded predicted d' data at various processing times. The search set was assumed to equal the explicit memory set. Some set of hit and false alarm rates that might have produced the set size one d's were selected. Ratcliff assumed that all match comparisons for all serial positions and all set sizes yield identical hit rates (have identical resonance processes). All mismatch comparisons of either the negative probe or the positive probes to an inappropriate serial position were assumed to yield identical false alarm rates (have identical resonance properties). (Because hit and false alarm data were not available for Reed's data, and because Reed used a long retention interval, this is a reasonable set of simplifying assumptions.) Given these assumptions, the probability of hit and false alarm rates for larger set sizes could simply be multiplied through. (Probability of a hit for set size n is the probability that at least one of the match and $n - 1$ mismatch comparisons yield a positive response. Probability of a false alarm for set size n is the probability that at least one of the n mismatch comparisons yields a positive response.)

Ratcliff's (1978) computation, for most selections of hit and false alarm rates, predicted slower rates for larger set sizes. All of the dynamic differences arise as a consequence of the size of the search set, and the mismatch processes. The existence and size of predicted dynamic set size effects depends on the bias factor assumed in the initial selection of hit and false alarm rates for set size one. It would be possible, in principle, to select a pattern of hit and false alarm rates that would eliminate the prediction of unequal dynamics for different set sizes. (The assumption of a constant false alarm rate across retrieval time within a set size, and very unequal false alarms between set sizes, is one such pattern.) However, for the pattern of false alarms that we observed (changing over time and similar between set sizes), a substantial set size effect on rate is predicted.

Ratcliff's (1978) account of set size effects in SAT data was based on at least two important assumptions: (a) that all positive probes both within and between memory set size are equal and (b) that the search set equals the memory set. These assumptions were necessary, given the unavailability of serial position information and in order to make model estimation tractable. In our studies, however, serial position effects at asymptote were incompatible with Assumption (a), and recent negative effects are apparently incompatible with Assumption (b). A number of possible changes might allow the model to accommodate our data. One possibility is to eliminate a set size effect by decoupling the memory set from the implicit search set by assuming that the search set is very large, large enough to include the recent negatives. Another is to weaken the impact of a memory-set determined search set by introducing large interitem differences in both match and mismatch processes. In this case, negatives intrude themselves into the search set by direct access matching.

One can accommodate equivalent dynamics for different serial positions within a set size by simply allowing separate mean resonance values for each position. However, slower dynamics for larger memory sets result from early increased false alarms because of more processes in negatives. Because set size does not appear to affect dynamics, and because recent negatives differ from distant negatives, it is tempting to assume that the effective search set is large and includes many comparisons for items not in the memory set. Ratcliff (1978) presented evidence that the search set may be considered arbitrarily large without affecting results very much, so long as the additional comparisons are so low in resonance that they always terminate in a mismatch very quickly in relation to the few likely contenders. Small residual set-size-based differences in asymptotic accuracy levels would be accounted for by differential study patterns. In this scenario, the search set is eliminated as a definition of to-be-remembered items. Items are basically evaluated on a strength dimension. As such, it is not clear that this scenario is preferable to a simpler single-comparison strength evaluation.

An alternative is to retain the notion of an implicit search set equal to the memory set but to assume that performance in all cases is dominated by some small number of most relevant item representations. Comparisons for serial positions leading to lower resonance values for positive trials might also lead to higher resonance values on negative trials. This outcome is counter to Ratcliff's (1978) assumptions, but it might be reasonable if the memory representations that best code a particular positive are also least confusable with other items. In this scheme, the memory set does not strongly affect dynamics because a few comparisons dominate performance, regardless of the size of the search set for both positive and negative probes. Recent negative performance would have to reflect direct-access contact of the probe with its own fairly recent memory representation. Similar contact for distant negatives would reflect less strength.

Either version of the diffusion model outlined earlier requires post hoc modifications to account for the retrieval benefit for the final list member and for the correction process for recent negatives. Both modifications represent weak or nonexistent set size mechanisms, which instantiate the observed dynamic equivalence of retrieval from memory lists of different lengths.

Other Models

We propose variants of the diffusion model to account for asymptotic and dynamic properties of retrieval from short-term memory lists. However, more generic direct-access models combined with a model of asymptotic strength differences may be equally plausible. Monsell (1978) argued against direct-access models, largely on the basis of interactions of set size and recent negatives. He explicitly manipulated recency of distractors across different list lengths by a cued forgetting paradigm. In that paradigm, a large list was presented on each trial, in which some list items appeared before a forget cue and some after. Items after the forget cue constituted the memory list, and items before it formed a pool for recent negatives. On some trials, unpredictably, the first list segment was tested, and so subjects could not ignore that early list portion until the forget cue actually occurred. Items were labeled according to their position from the end of the memory list (i.e., −1, −2, . . .). Monsell found that recent negatives in the same list positions (i.e., −8) yielded different RTs and error rates, depending on the length of the memory list, with only small recency effects. He argued that direct-access strength models could not simultaneously allow nearly overlapping serial position functions for positives and list-length effects on recent negatives (his Proposition IIb). His formulation assumed a monotonically decreasing transfer function between the difference of a sampled strength from the criterion and reaction time. He reasoned that set-size-dependent criterion shifts would account for set size effects on negatives but would introduce differences between positives. However, our SAT data show changes in strength for the same recency positions in different list lengths. The list-length, serial position factors are outlined as follows.

SAT asymptotic accuracy for the various set sizes and serial positions bears striking similarity to the untimed accuracy data of Wickelgren and Norman (1966). They varied the list length and examined serial positions separately for short-term memory for three-digit numbers. A short-term memory acquisition-primacy model accounted well for their data. This model assumed a constant, item-dependent forgetting rate for items in all serial positions, but it assumed serial-position-dependent initial acquisition parameters. In other words, $d(k,L) = \alpha(k)\phi^{L-k}$, where L is the list length; k is the serial position counting from the beginning of the list; $d(k,L)$ is the untimed recognition accuracy for position k of list length L; $\alpha(k)$ is a serial-position-dependent initial acquisition value; ϕ is a constant per-item decay parameter; and $L - k$ represents the number of list items following the specified item. This model predicts linear forgetting functions on semilog axes, with equal slopes but possibly unequal initial acquisition intercepts for all serial positions. An analysis of our asymptotic accuracy data does not yield a strong test of linearity on semilog axes (each serial position occurred only twice in each experiment). However, the slopes were remarkably consistent, and the pattern of initial acquisition parameters and decay estimate was remarkably similar to that of Wickelgren and Norman (1966). The average asymptote data for the two experiments were fit together with this model, without regard to the differences in subjects, font, level of practice, and so on. The model fit accounted for .946 of the variance in asymptotic levels. Like Wickelgren and Norman, we found that the first element on a list had substantially higher acquisition strength; the strength of each successive position was slightly lower, $\alpha(1) = 6.3$, $\alpha(2) = 4.9$, $\alpha(3) = 4.7$, $\alpha(4) = 4.4$, $\alpha(5) = 4.3$, and $\alpha(6) = 4.1$ in d' units. Our fitted decay rate was .82, whereas theirs ranged from .60 to .80. (Wickelgren, 1970, later showed that the decay rate depended not just on subsequent items but on presentation rate. Also, it is possible to accommodate variance differences in the strength distribution. However, the approximation just given is adequate for our purposes.)

Predicted asymptotic strengths for various list lengths are shown in the four panels of Figure 15. These values are based on the estimated αs and ϕ listed earlier. Also shown are arbitrarily selected criterion values that nearly equate the distance of each comparable serial position from its respective criterion. The absolute distances of each of the distribution means from the relevant criterion are shown in Figure 16. Last, in Figure 17, the mean RT from Experiment 2 (a different set of subjects and practice levels) is plotted as a function of the distance from the criterion of each condition of Figure 16. The relation between the hypothetical controlling values and Experiment 2 RT is strong. As did Murdock (1985), we assume a monotonic and negatively accelerated "transfer function" between this distance and mean RT, shown as the smooth line in Figure 17. The data from serial position −1 (immediate repetition) fell below the transfer function, but this is consistent with the dynamic benefit that applies to immediate repetition conditions. In other words, an immediate repetition resulted in a substantial speeding in retrieval over and above the effects of asymptotic strength; hence this pure strength analysis does not incorporate the expected dynamic speeding for −1 serial positions. The transfer function accounts for about 89% of the variance for all conditions, excluding serial position −1. (A linear formulation including serial position −1 systematically underfits data in

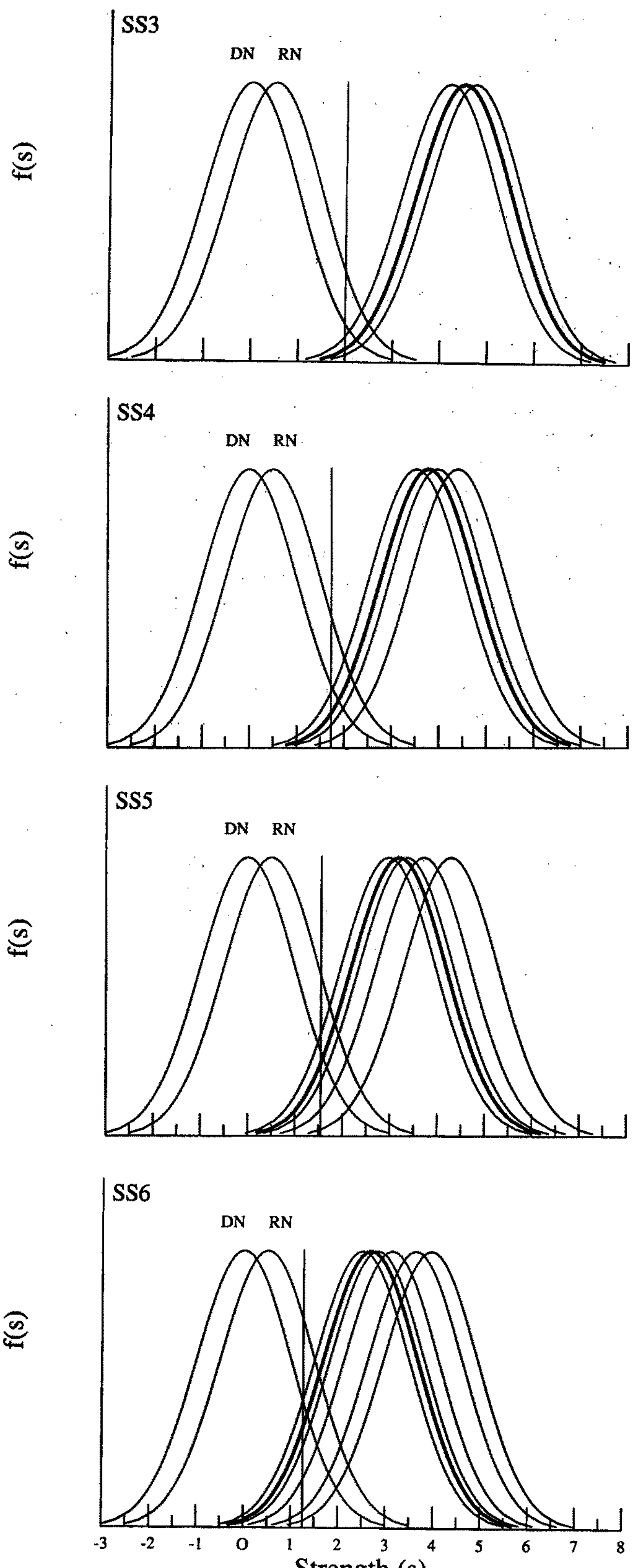

Figure 15. Predicted strength distributions for positive serial positions, recent negatives (RN), and distant negatives (DN) derived from empirical estimates of the acquisition (α) and decay (ϕ) parameters in the "acquisition-primacy" model of Wickelgren and Norman (1966). (The four panels show set sizes [SS] three to six, respectively. Items in the primacy position are denoted by the heavy lines; the remaining serial positions, from the most recent to the most remote, yield a consistent ordering from right to left on the strength axis.)

the midrange, but it accounts for a comparable amount of the variance.)

Figure 17 slightly oversimplifies the relation between observed RT, the transfer function, and distance from the criterion. In fact, the transfer function would need to be convolved with the item distributions in the four panels of Figure 15, and error trials would have to be removed in order to do a careful fit of RT data. Also, the criterion values and transfer function were not systematically optimized. Nonetheless, the ability of the strength model for asymptotic accuracy to account for RT data is impressive. Qualitatively, this account predicts approximately superimposed positive serial position functions in RT. (The shape of the serial position functions is slightly different from that shown in Figure 16, because different serial positions fell on different portions of the curved transfer function. Also, serial position −1 received an added dynamic benefit.) Small set size effects on SAT accuracy are built into the d' strength values. Criteria are lower for longer set sizes, hence predicting the effect of set size on negative RT. Some interaction of recent and distant negatives with set size (Monsell, 1978) is possible, depending on the shape of the transfer function. Recent negatives are shifted overall toward smaller average distances to the criterion and hence may occupy a higher slope region of the transfer function.

One possible mechanistic account of the retrieval process that would be qualitatively consistent with the RT transfer function of Figure 17 is the diffusion model discussed earlier. This is a continuous accrual mechanism in which the drift rate is determined by an item's resonance or match value (strength). Another possible mechanism is Hockley and Murdock's (1987) noise-sampling model. In this model, a single memory strength sample is obscured by independent added noise samples in a series of sampling epochs. A decision is made whenever the sample plus noise exceeds a high criterion for "yes" responses or falls below a low criterion for "no" responses. Unlike the diffusion model, this model does not cumulate information over retrieval time; instead, the criteria are moved closer together over time. This model is generally consistent with the RT transfer function of Figure 17; however, the model requires post hoc modifications in order to account for the full trade-off from chance to asymptotic accuracy of SAT curves. A number of other mechanistic accounts might also be compatible with the RT transfer function.

In summary, Monsell (1978) apparently assumed that list length would not affect the strength estimates for items with a given recency, and hence he excluded a strength model. However, as shown here, fairly simple strength models of retrieval accommodate both the RT pattern and the SAT data. Furthermore, the strength model is the same as that shown previously (Wickelgren & Norman, 1966) to apply to untimed accuracy in short-term lists.

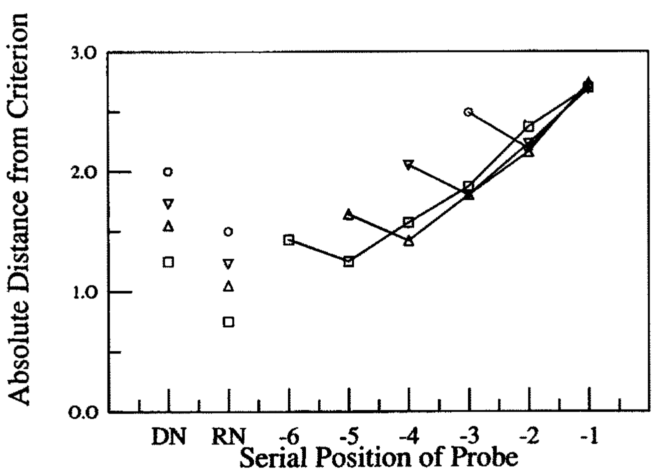

Figure 16. Absolute distance of each strength distribution mean from its set-size-contingent criterion.

RT and SAT Paradigms

As described in the Introduction, one possible concern is that the SAT paradigm alters the retrieval strategies of the subject, resulting in differences in processing between SAT and RT paradigms. For example, subjects might be induced to use a terminating variant of a scan process in order to prepare for interruption in SAT. To the contrary, we believe that SAT performance and RT performance in our short-term memory retrieval experiments reflect the same set of basic memory processes. Consider the position that subjects use (as we found) a direct-access retrieval mechanism—coupled with classic forgetting mechanisms—in the SAT paradigm, but that they use an exhaustive scan mechanism in the RT paradigm. One must then separately account for the strong relation between the SAT results and the RT results illustrated in Figure 17. We argue that the systematic variance

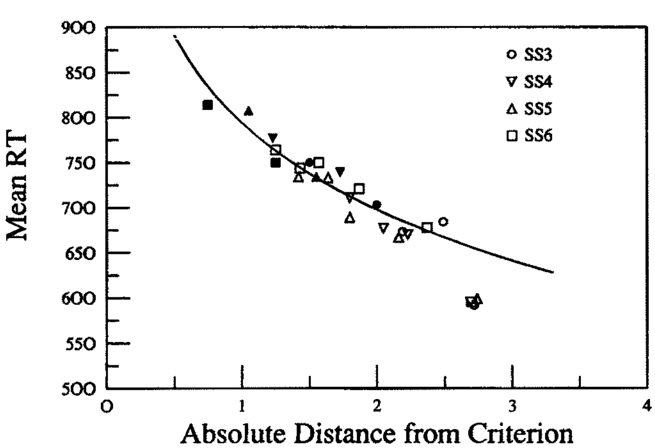

Figure 17. A monotonic, negatively accelerated transfer function mapping distance from the criterion into mean reaction time (RT). (The points that lie below the function are the respective positive probes from the most recent serial position and may reflect speeded dynamics.)

in the RT results follows directly from some aspect of SAT retrieval parameters: either from properties of strength distributions (asymptotic levels) or, in the case of immediate repetition, from the additional sharp increase in rate of retrieval. Thus we propose a single mechanism that can generally account for the SAT performance, for the relation of RT to that performance, and for the main properties of the RT results themselves. Not only is there no evidence that compels us to propose a different mechanism (either exhaustive or self-terminating scan) to account for RT data, but such a proposal would also require an explanation of why SAT parameters and RT data correlate so highly and, more generally, why subjects would elect to use a mechanism in which intermediate information is not available in RT, when a direct-access mechanism is available for use in SAT.

This strong relation between parameters of the SAT functions and RT results is not unusual (i.e., Dosher, 1982, 1984a; Dosher & Rosedale, 1989), although there may be some domains in which strong paradigm-specific strategic differences would occur. Furthermore, this discussion should not be taken to imply that RT data can be predicted without additional estimated parameters from the SAT data. At a minimum, some temporal offsets that estimate the time devoted to detection and processing of the interruption cue are required. (For a related discussion of partial information dynamics in decomposition methods, see Ratcliff, 1988.) Nonetheless, we assert that our model conclusions likely generalize to other paradigms such as RT.

Conclusions

There is no evidence in our data for an impact of set size on the time course of recognition from short-term memory. Furthermore, asymptotic accuracy differences between set size can, in fact, be well accounted for by a simple acquisition-primacy model in which apparent set size effects are merely a by-product. The primacy-acquisition model, coupled with some simple direct-access retrieval model and a match benefit for immediate repetition, accommodates both SAT and comparable RT data quite well. Nonetheless, it might be possible to argue that the retrieval process in circumstances like ours, with short retention intervals, emphasizes serial position differences that might be less apparent under typical "scan" circumstances with rehearsal and a longer retention interval (Sternberg, 1975).

References

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review, 80,* 417–438.

Atkinson, R. C., Herrmann, D. J., & Westcourt, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 101–146). Hillsdale, NJ: Erlbaum.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1, pp. 243–293). San Francisco: Freeman.

Aube, M., & Murdock, B. B., Jr. (1974). Sensory stores and high speed scanning. *Memory & Cognition, 2,* 27–33.

Baddeley, A. D., & Ecob, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for the high speed exhaustive scan hypothesis. *Quarterly Journal of Experimental Psychology, 25*, 229–240.

Briggs, G. E. (1974). On the predictor variable for choice reaction time. *Memory & Cognition, 2*, 575–580.

Burrows, D., & Okada, R. (1971). Serial position effects in high-speed memory search. *Perception and Psychophysics, 10*, 305–308.

Burrows, D., & Okada, R. (1975). Memory retrieval from long and short lists. *Science, 188*, 1031–1033.

Clifton, C., & Birenbaum, S. (1970). Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology, 86*, 69–76.

Corballis, M. C. (1967). Serial order in recognition and recall. *Journal of Experimental Psychology, 74*, 99–105.

Corballis, M. C. (1975). Access to memory: An analysis of recognition times. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press.

Corballis, M. C., Kirby, J., & Miller, A. (1972). Access to elements of a memorized list. *Journal of Experimental Psychology, 94*, 185–190.

Corbett, A. (1977). Retrieval dynamics for rote and visual image mnemonics. *Journal of Verbal Learning and Verbal Behavior, 16*, 233–246.

Corbett, A., & Wickelgren, W. A. (1978). Semantic memory retrieval: Analysis by speed–accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology, 30*, 1–15.

Dosher, B. A. (1976). The retrieval of sentences from memory: A speed–accuracy study. *Cognitive Psychology, 8*, 291–310.

Dosher, B. A. (1979). Empirical approaches to information processing: Speed–accuracy tradeoff or reaction time. *Acta Psychologica, 43*, 347–359.

Dosher, B. A. (1981). The effect of delay and interference: A speed–accuracy study. *Cognitive Psychology, 13*, 551–582.

Dosher, B. A. (1982). Sentence size, network distance, and sentence retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 173–207.

Dosher, B. A. (1984a). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 541–574.

Dosher, B. A. (1984b). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed–accuracy study. *Cognitive Psychology, 16*, 519–555.

Dosher, B. A., & Rosedale, G. R. (1989). Combined retrieval cues as a mechanism of priming. *Journal of Experimental Psychology: General, 118*, 191–211.

Forrin, B., & Cunningham, K. (1973). Recognition time and serial position of probe item in short-term memory. *Journal of Experimental Psychology, 99*, 272–279.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.

Grossberg, S., & Stone, G. (1986). Neural dynamics of attention switching and temporal-order information in short-term memory. *Memory & Cognition, 14*, 451–468.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96–101.

Hockley, W. E. (1984). Analysis of response time distribution in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 598–615.

Hockley, W. E., & Corballis, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology, 36*(2), 189–212.

Hockley, W. E., & Murdock, B. B., Jr. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94*, 341–358.

Juola, J. F., & Atkinson, R. C. (1971). Memory-scanning for words versus categories. *Journal of Verbal Learning and Verbal Behavior, 10*, 522–527.

Kristofferson, M. W. (1972). When item recognition and visual search functions are similar. *Perception and Psychophysics, 12*, 379–384.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychological Review, 95*, 183–237.

McNicol, D., & Stewart, G. W. (1980). Reaction time and the study of memory. In A. T. Welford (Ed.), *Reaction times* (pp. 253–307). London: Academic Press.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10*, 465–501.

Morin, R. E., De Rosa, D. V., & Shultz, V. (1967). Recognition memory and reaction time. *Acta Psychologica, 27*, 298–305.

Murdock, B. B., Jr. (1971). A parallel-processing model for scanning. *Perception and Psychophysics, 10*, 289–291.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.

Murdock, B. B., Jr. (1985). An analysis of the strength–latency relationship. *Memory & Cognition, 13*, 511–521.

Murdock, B. B., Jr., & Anderson, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 145–194). Hillsdale, NJ: Erlbaum.

Murdock, B. B., Jr., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology, 94*, 284–290.

Murdock, B. B., Jr., & Franklin, P. E. (1984). Associative and serial-order information: Different modes of operation? *Memory & Cognition, 12*(3), 243–249.

Nickerson, R. S. (1972). Binary-classification reaction time: A review of some studies of human information processing capabilities. *Psychonomic Monograph Supplements, 4*(17, Whole No. 65).

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in short-term memory. *Journal of Mathematical Psychology, 6*, 192–208.

Okada, R., & Burrows, D. (1978). The effects of subsidiary tasks on memory retrieval from long and short lists. *Quarterly Journal of Experimental Psychology, 30*, 221–233.

Pachella, R. G. (1974). The interpretation of reaction time information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halstead Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs, 76*(1, Pt. 2).

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs, 79*(1, Pt. 2).

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology, 21*, 178–184.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review, 95*, 238–255.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.

Reed, A. V. (1973). Speed–accuracy trade-off in recognition memory. *Science, 181*, 574–576.
Reed, A. V. (1976). The time course of recognition in human memory. *Memory & Cognition, 4*, 16–30.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review, 84*, 1–66.
Seamon, J. G., & Wright, C. E. (1976). Generative processes in character classification: Evidence for a probe encoding set. *Memory & Cognition, 4*, 96–102.
Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652–654.
Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*, 421–457.
Sternberg, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1–32.
Theois, J. (1973). Reaction time measurement in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 44–85). New York: Academic Press.
Theois, J., Smith, P. G., Haviland, S. E., Traupman, J., & Moy, M. C. (1973). Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology, 97*, 323–336.
Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
Treisman, M., & Doctor, E. (1987). Memory scanning: A comparison of the dynamic stack and exhaustive serial scan models, with an extension of the latter. *Acta Psychologica, 64*, 39–92.
Wickelgren, W. (1970). Time, interference, and rate of presentation in short-term recognition memory for items. *Journal of Mathematical Psychology, 7*, 219–235.
Wickelgren, W. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67–85.
Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes–no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 189–202.
Wickelgren, W. A., Corbett, A. T., & Dosher, B. A. (1980). Priming and retrieval from short-term memory: A speed–accuracy tradeoff analysis. *Journal of Verbal Learning and Verbal Behavior, 19*, 387–404.
Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3*, 316–347.

Received August 10, 1988
Revision received February 28, 1989
Accepted February 28, 1989 ■

Beutler, Levin, Tesser, and Miller Appointed New Editors, 1991–1996

The Publications and Communications Board of the American Psychological Association announces the appointments of Larry E. Beutler, University of Arizona; Joel R. Levin, University of Wisconsin; Abraham Tesser, University of Georgia; and Norman Miller, University of Southern California, as editors of the *Journal of Consulting and Clinical Psychology*, the *Journal of Educational Psychology*, the Attitudes and Social Cognition section and the Interpersonal Relations and Group Processes section of the *Journal of Personality and Social Psychology*, respectively. As of January 1, 1990, manuscripts should be directed as follows:

- For *Consulting and Clinical* send manuscripts to Larry B. Beutler, *Journal of Consulting and Clinical Psychology*, Department of Psychiatry, University of Arizona, College of Medicine, Tucson, Arizona 85724.

- For *Educational* send manuscripts to Joel R. Levin, Department of Educational Psychology, University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706.

- For *JPSP: Attitudes* send manuscripts to Abraham Tesser, Institute for Behavioral Research, University of Georgia, 548 Boyd Graduate Studies, D. W. Brooks Drive, Athens, Georgia 30602.

- For *JPSP: Interpersonal* send manuscripts to Norman Miller, Department of Psychology, Seeley G. Mudd Building, University of Southern California, University Park, Los Angeles, California 90089.

Manuscript submission patterns make the precise date of completion of 1990 volumes uncertain. Current editors will receive and consider manuscripts until December 1989. Should any 1990 volume be completed before that date, manuscripts will be redirected to the newly appointed editor-elect for consideration in the 1991 volume.