Journal of Experimental Psychology: General
1993, Vol. 122, No. 3, 291–315

0096-3445/93/$3.00

Serial Retrieval Processes in the Recovery of Order Information

Brian McElree and Barbara Anne Dosher

The retrieval of temporal order and item information from short-term memory (STM) are examined with the cued-response speed–accuracy trade-off (CR-SAT) procedure and a complementary reaction time (RT) task. The retrieval of order information was examined with a two-alternative forced-choice (2AFC), relative judgment of recency (JOR) task. Analyses of the pattern of mean RT, RT accuracy, and the overall shape of the RT distribution for correct JORs suggest that order information is retrieved by a serial retrieval mechanism. Analyses of SAT retrieval functions confirm that order information is retrieved by a recency-based, serial retrieval process. These results contrast with previous SAT analyses of STM item recognition (B. McElree & B. A. Dosher, 1989), which indicate that item information is retrieved by a parallel or direct-access mechanism. The dissociation between item and order information retrieval was further documented in a 2AFC item recognition SAT study.

Human memory records not only individual events but also relationships between events, including temporal, spatial, logical, and causal relationships. An adequate model of human memory must specify how both individual events and relationships between events are stored and subsequently retrieved. This article examines how temporal order information is retrieved from short-term memory (STM). Our primary focus is on the nature of the retrieval process, specifically, whether temporal order information is recovered by a serial or parallel retrieval mechanism.

We report evidence that indicates that order information is retrieved from STM with a slow serial process. Our prior work (Dosher & McElree, 1992; McElree & Dosher, 1989) used a cued-response speed–accuracy trade-off (CR-SAT) analysis to demonstrate fast parallel processes in the retrieval of item information from STM. Here, we used CR-SAT and collateral reaction time (RT) tasks to document strongly serial processing for the recovery of order information from STM under conditions comparable with those for item recognition. Furthermore, we show that the seriality of order (i.e., before and after) recovery is not simply a consequence of the need to consider more than one item. The contrast in retrieval for item and order information documented here has important implications for general memory models that attempt to account for both types of information (e.g., Lewandowsky & Murdock, 1989). The application of CR-SAT procedures to the issue of serial and parallel retrieval processes illustrates how SAT procedures in general provide a strong empirical basis for discriminating between serial and parallel processing architectures.

In the first part of this article we (a) review the evidence for parallel retrieval of item recognition from STM, (b) describe the major paradigm for testing retrieval of order information and review the pattern of findings in RT measures, and (c) provide details of the CR-SAT methods and the predictions of serial and parallel processing models for the resulting data.

Serial and Parallel Processes in Item Recognition

The departure point for studies of serial and parallel retrieval processes in STM retrieval is Sternberg's (1966, 1969, 1975) classic RT studies of short-term item recognition. Sternberg argued that items were retrieved from STM by a serial exhaustive scan. He based his arguments on the finding that correct RT was a linear function of the number of items studied, with approximately equal slopes for positive and negative responses.

Unfortunately, mean RT data alone cannot in general discriminate between serial and parallel processing mechanisms. Often, with an appropriate set of assumptions, serial and parallel processing organizations yield equivalent RT predictions (e.g., Townsend & Ashby, 1983). One response to this ambiguity has been to consider properties of RT data other than the mean, including, for example, RT variance (Shiffrin & Schneider, 1977; Townsend & Ashby, 1983) and the shape (Hockley, 1984; Hockley & Corballis, 1982; Ratcliff, 1978; Ratcliff & Murdock, 1976) or higher order moments of the RT distribution (Sternberg, 1973).

A second method for more strongly contrasting serial and parallel mechanisms is to examine the full time course of processing with speed–accuracy trade-off procedures, in particular with the CR-SAT paradigm (Corbett, 1977; Corbett & Wickelgren, 1978; Dosher, 1976, 1979, 1981, 1982; McElree

---

Brian McElree and Barbara Anne Dosher (now at Department of Cognitive Sciences, University of California, Irvine), Department of Psychology, Columbia University.

Experiment 2 was supported by BNS89–19498 (co AFOSR89-NL295). Experiments 1a and 1b were based on Brian McElree's doctoral dissertation, submitted to Columbia University (1990). Partial reports of this work were presented at the 1990 Annual Interdisciplinary Conference, Jackson Hole, Wyoming, and at the 1991 meeting of the Psychonomic Society, San Francisco, California.

Correspondence concerning this article should be addressed to Brian McElree, Department of Cognitive Sciences, Social Science Tower, University of California, Irvine, California 92717. Electronic mail may be sent to bdm@aris.ss.uci.edu.

& Dosher, 1989; Ratcliff, 1978; Reed, 1973, 1976; Wickelgren, 1977; Wickelgren & Corbett, 1977; Wickelgren, Corbett, & Dosher, 1980). Retrieval time is controlled by requiring subjects to respond immediately following a response cue presented at various times during retrieval. Accuracy can thereby be measured across the full time course of retrieval, from an initial chance level to an asymptotic level. The SAT asymptote is a measure of memory strength. Retrieval speed is measured jointly by the rate at which the function grows to asymptote and when it departs from chance. We refer to this measure of retrieval speed as the *dynamics of retrieval*. Serial and parallel models make different predictions concerning the dynamics of retrieval.

McElree and Dosher (1989) used CR-SAT to explicitly test a variety of serial and parallel retrieval models for a Sternberg item recognition task, including serial exhaustive (Sternberg, 1975; Treisman & Doctor, 1987) and serial self-terminating (Theios, 1973) scan models, parallel scan models (Murdock, 1971), and the diffusion or continuous random walk model (Ratcliff, 1978). The dynamics of retrieval did not vary with either set size (a list of three to six words) or serial position within the lists, clearly indicating that item retrieval is not serial but is a parallel (e.g., Ratcliff, 1978) or direct-access (e.g., Reed, 1976) process. Set size is predicted to control dynamics in serial exhaustive models; study position is predicted to control dynamics in serial self-terminating and some rate-varying (capacity-limited) parallel self-terminating models. Serial models are also inconsistent with observed RT distributions. Serial processing mechanisms predict that increasing the number of processes carried on in series should increase not just the longest RTs, but should increase the minimum, or leading edge RTs, as well. Hockley (1984; Hockley & Corballis, 1982) found minimal impact of list length on the leading edge; only the longest trials were affected by list length.

Relative Judgments of Recency (JORs)

Many events, particularly salient events, may be directly associated with specific times or dates or may be associated to other time-tagged events by way of a causal relation (Bower, 1972; Estes, 1985; Linton, 1975; Wickelgren, 1972). Discriminating between the recency of two such events requires little beyond a direct retrieval of time concepts associated with each event. Of interest here are cases in which simple inferential processes or an associated date are not available; Bower (1972) cites examples such as determining the order of the "arrival of pupils before a class, of concert-goers at a symphony hall, or in-bound aircraft at Kennedy airport" (p. 102). We analyze the recovery of order information by examining the speed and accuracy of recency judgments in the absence of inferential links or direct time coding.

Judgments of recency have been examined in two paradigms: absolute and relative. In an absolute JOR task, subjects judge the absolute position of elements within a study list (e.g., Hinrichs, 1970; Lockhard, 1969). Of primary interest is the relationship of judged to absolute recency. In a relative JOR task, subjects perform two-alternative forced-choice (2AFC) recency discriminations, choosing which of two items occurred more recently on a list. The measures of primary interest in the relative JOR task are the speed and accuracy of the order discrimination.

Untimed Accuracy

Early studies of relative JOR examined untimed accuracy as a means to determine the form or type of memory representation underlying JOR. One approach exemplified by Yntema and Trask (1963) assumed that the memory trace explicitly coded order information in the form of a temporal tag. Other approaches asserted that temporal order information was based on other properties of the memory representation, such as trace strength (Hinrichs, 1970; Morton, 1968; Peterson, 1967), trace fragility (Wickelgren, 1972, 1974), or attribute counts (Bower, 1972; Flexser & Bower, 1974). The relevant property is assumed to decline monotonically with the time since study, the number of items intervening between study and test, or both. In a relative JOR paradigm, subjects are assumed to follow a decision rule that chooses as the most recent the item with the largest strength, fragility, or attribute value.

RT

The strength, fragility, attribute, and time tag theories differ representationally but share common processing assumptions: Information associated with each probe item is retrieved, the relevant recency metrics are compared, and an appropriate response is executed. However, an examination of RTs calls in to question this type of comparison model.

Muter (1979, 1980) and Hacker (1980) independently examined RT and associated accuracy for relative JOR as a function of the study position of each of the items in the test probe. Mean correct RT was found to be inversely related to the study position of the more recent or later probe in the pair, increasing as the later probe was drawn from more remote positions, but was unaffected by the study position of the less recent or earlier probe. Accuracy dramatically decreased when the later probe was drawn from earlier study positions. Holding the study position of the later probe constant, accuracy slightly increased for more remote study positions of the earlier probe, that is, as the separation in recency between the two probe items increased. Hockley (1984) has also reported the same pattern of results. If a response was executed on the basis of a comparison of the respective recency metrics of the two probe items, then RT should vary with the difference in recency between the two probe items. To the contrary, the data indicate that processing of the probes is self-terminating on finding the first match to an item in the memory set.

Hacker's (1980) Model

Hacker (1980) proposed a retrieval model to account for the pattern of RT and accuracy in which comparison of the test probes with elements in the memory set is accomplished by a backward or recency-based, serial search or scan. Sternberg (1969) proposed a related, although forward rather than backward, serial scan model for the recovery of order information under somewhat different experimental circumstances (see McElree & Dosher, 1991, for a discussion of Sternberg's model and data). The model assumes a memory representation in which items are ordered by their position in the study list. At test, the probability that any particular item is still available in memory is represented in the model by an availability parameter ($0 \leq a_i \leq 1$). (Less recent items are typically less available.) Test probes are compared with elements in memory in a serial fashion, starting with the most recent and moving backward through the memory representation. The scan is self-terminating in that the first test probe that matches an item in the memory representation is chosen as the more recent. If the later probe is unavailable (with probability $pr = 1 - a_i$), the earlier probe is incorrectly chosen as the more recent. If both probes are unavailable, the subject is assumed to be guessing. The model estimates mean RT with three processing parameters: (a) a base time b reflecting average time to encode probes and execute a response, (b) a search time s that reflects the average time to compare test probes with an individual item in the memory representation, and (c) a guessing time g that estimates the average time to guess when both probes are unavailable. Equations for estimating probability correct, mean correct RT, and mean incorrect RT for JOR pairwise comparisons are detailed in the Appendix.

Hacker (1980) found that this relatively simple scanning model produced very good quantitative fits of the percentage correct, mean correct RT, and mean incorrect RT data. Moreover, although the serial scanning model was not fit to RT distributions directly, the shapes of the correct RT distributions reported by Muter (1980), Hacker (1980), and Hockley (1984) were generally compatible with a serial model in that the leading (fast) edge of the RT distributions is affected by the recency of the later test item (see the RT Distribution section of Experiment 1a for a detailed treatment).

Remaining Ambiguities in Interpretation

Unfortunately, as acknowledged by Hacker (1980) and Muter (1980), the mean and distributional data pattern for relative JOR judgments, although consistent with a serial scan model, do not necessarily rule out a parallel model. The mean RT and error data may be consistent with a model in which matches to items in memory are processed in parallel, but the time to complete each comparison (to a criterion level) varies with recency. Purely parallel models may be consistent with the RT distributions as well. In at least one model (Ratcliff, 1978), parallel retrieval processes that differ only in the degree of match (i.e., strength) can produce shifts in the leading edge of RT distributions.[1] To use RT means and distributions to argue definitively for either serial or parallel models requires precise quantitative development of both models.

We pursue another alternative here, which is to examine the full time course of retrieval by using the CR-SAT procedure. Full time course data can give converging information that may allow us to rule out serial or parallel models (cf. McElree & Dosher, 1989). The distinction focuses on the availability of partial information about order when subjects make a judgment after being interrupted early in processing. The serial backward scan model predicts that tests including a less recent later item should have longer initial periods during the scanning process when information is at chance.

SAT Methodology

The CR-SAT method interrupts the retrieval process with a cue to respond and measures accuracy at various points across the full time course of retrieval. Full retrieval functions generated from this method typically show a period of chance performance, followed by a period of rapid increases in accuracy, and finally an asymptotic accuracy level as retrieval time is further increased. Generally, three parameters suffice to describe these functions: (a) an asymptotic accuracy parameter reflecting overall memory limitations, (b) an intercept or minimal processing time, and (c) 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, & Kounois, 1988).

The CR-SAT method yields a strong empirical contrast of parallel and serial processing models because of its ability to separately assess the impact of experimental factors on dynamics and asymptotic performance. Differing numbers of serial processes will alter the dynamics of information accrual, independent of covarying or independent effects on asymptotic performance. The duration of the serial processing component controls the distribution of finishing times in cases in which the underlying processing is discrete or quantal, and most, if not all, serial processing models are cast in discrete terms (see Meyer et al., 1988). Because SAT dynamics reflect the distribution of finishing times for discrete processes, different finishing time distributions will result in SAT functions that rise to asymptote disproportionately or at different rates. The CR-SAT methodology has been profitably used to critically examine serial processing claims in a number of domains, including short-term item recognition (McElree & Dosher, 1989; Reed, 1976), perceptual matching (Ratcliff, 1981), retrieval interference (Dosher, 1981), long-term sentence recognition (Dosher, 1982), and semantic-verification tasks (Corbett & Wickelgren, 1978).

Differences in SAT dynamics are a necessary prediction of serial processing models (see Dosher, 1981, 1982, 1984; McElree & Dosher, 1989). Parallel processing models are intrinsically more flexible. Whereas parallel processing models are consistent with common dynamics (i.e., rise strictly proportional to asymptote), dynamics differences can often result from inherent differences in the rate of continuous information accrual (Dosher, 1984; McElree & Dosher,

---

[1] More detailed analyses of the relative magnitudes of shifts in the leading edge and tail of the RT distributions might rule out specific parallel or serial models.

1989) or from the type of decision rules governing information combination (Ratcliff, 1978). Discriminating between serial and parallel models in the presence of dynamics differences is therefore a more subtle process of precisely fitting predicted to observed retrieval functions. We approach this problem for JOR by deriving specific predictions of the Hacker model for SAT data.

Note that RT data alone are not sufficient to estimate dynamics because RT may represent an unknown mixture of dynamic and asymptotic effects. Empirically, RT has often been observed to covary with asymptotic SAT differences, even in the absence of differences in SAT dynamics (Dosher, 1982, 1984; McElree & Dosher, 1989; Ratcliff, 1978; Wickelgren, 1977). In other cases, RT covaries with SAT dynamics in the absence of asymptotic differences (Dosher, 1981; Dosher & Rosedale, 1989). Because asymptotic and dynamic effects have distinct theoretical interpretations, models constructed solely in response to RT data may substantially misrepresent the underlying processes. The approach of this article is to relate mean RT, RT distributions, and accuracy in an RT paradigm to SAT data on the time course of order judgments.

SAT Predictions From Hacker's Model

In this section, we present SAT functions predicted by the Hacker backward serial self-terminating model of relative JOR. Figure 1 illustrates the strong effect of recency in SAT. The derivations of the functions are outlined in the Appendix. In this figure, finishing time for each comparison process in the scan was exponentially distributed with a rate (S) corresponding to the mean search time for each comparison (s) in the RT model. Hence, the finishing time distribution for the entire scan is gamma distributed with an order equal to the number of serial comparison processes performed before a match is found (see Equation A4). The expected number of serial comparison processes for any test pair depends on the number of items remaining in the memory representation at test. The item availability parameters (a_i) provided the basis on which to weight gamma distributions to compute predicted accuracy at various retrieval times.

The SAT functions are derived for pairwise combinations of the last four study positions (i.e., [34, 35, 36, 45, 46, 56]) of six-item lists by using model parameters reported by Hacker (1980; see Figure 1 caption). Figure 1 presents predicted SAT functions in units of d'. The predicted SAT asymptotic accuracies show a pattern equivalent to accuracy in the RT studies of Hacker. The asymptotes of the SAT curves are strongly affected by the recency of the later probe and, to a much smaller degree, by the recency of the earlier probe. Lower asymptotic accuracy reflects errors produced when the later (more recent) probe is absent from the memory representation. The earlier probe affects asymptotes only when the later probe is unavailable; hence, errors with later probes from Position 6 are virtually nonexistent.

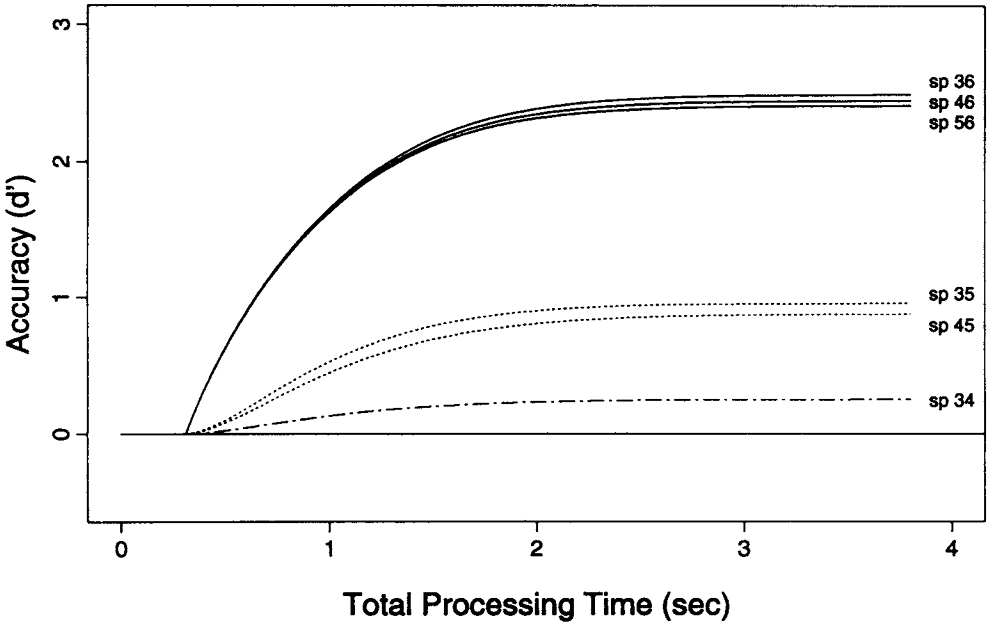

Figure 1. Predicted speed–accuracy trade-off functions for two-alternative forced-choice judgments of recency from Hacker's (1980) serial self-terminating scan model. (The curves show predicted d' accuracy as a function of processing time for the six pairwise comparisons of the last four serial positions [sp] in a list of six items. The curves were derived by using the method specified in the Appendix with the parameters reported in Hacker [1980, p. 664]. Parameter values: a_6 = .946, a_5 = .662, a_4 = .577, a_3 = .474, search time per item, s = 209 ms, base encoding time, b = 209 ms.)

Most critically, the study position of the later probe determines the dynamics of the SAT function by means of the serial scan mechanism. There are two parts to the predicted dynamics differences illustrated in Figure 1. The first is an apparent increase in the intercept as the later probe is drawn from less recent study positions. The second is a concomitant slowing of the rate of rise to asymptote. Although differential rate of rise may be consistent with either serial or rate-varying parallel processes, significant delays in the initial availability of information will be taken as strong support for a serial backward scan.

Experiments 1a and 1b: JOR

The RT results reported by Muter (1979, 1980), Hacker (1980), and Hockley (1984) are robust across different rates of presentation (2–16.7 items/s), different types of study materials (consonants and words), and different list lengths (4–13 items). The experiments reported here use set sizes of 6 consonants presented at modest rates of presentation (2–3.3 items/s). Test probes are constructed from pairwise combinations of all study positions, yielding 15 distinct types.

We were primarily interested in measuring properties of the full time course SAT data to see if they conform to the predictions of the serial, self-terminating model. Experiment 1b reports the SAT results. Experiment 1a reports an RT experiment in which sufficient data were collected to determine the pattern of mean RT, RT accuracy, and the concomitant shape of the RT distributions. The RT results replicate those reported by Muter (1979, 1980), Hacker (1980), and Hockley (1984) and allow explicit comparison with full SAT retrieval functions on an individual subject basis. RT data is subject to speed–accuracy trade-offs and does not provide a means of independently measuring the speed and ultimate accuracy of retrieval. As such, it cannot be used in isolation to evaluate unambiguously the applicability of serial versus parallel retrieval mechanisms. Nevertheless, principled relationships should exist between the RT and SAT data.

Method

Subjects

Four subjects participated in a total of 20 approximately 75-min experimental sessions. Subjects EC and LB were paid for their services. Subjects BM and GR were affiliated with the laboratory and volunteered their services. All subjects had normal or corrected vision.

Design and Stimuli

Each study list consisted of six consonants randomly selected (with no replacement). The test probe consisted of two consonants drawn from the study list. Across a session, each pairwise combination of study positions from the list was tested equally in a random

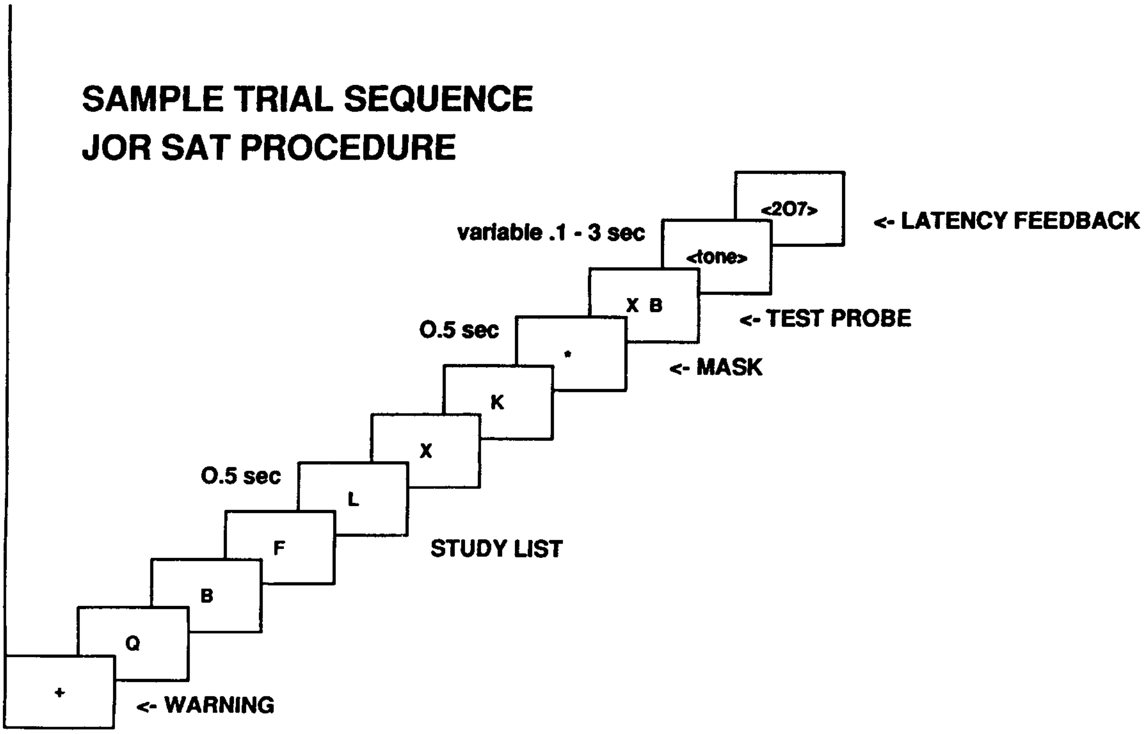

Figure 2. A sample trial sequence illustrating the speed–accuracy trade-off (SAT) variant (see Experiment 1b) of the relative judgment of recency (JOR) task. (In the reaction time variant of the task [Experiment 1a], the identical procedure was used up to the presentation of the test probe, at which point the tone was eliminated, and the test probe remained on the screen until the subject responded.)

order. The left–right order of presentation was completely counterbalanced within a session.

Each session consisted of two blocks of 210 trials. Across the 14 SAT sessions, this yielded for each subject a total of 56 trials per lag for each of the 15 pairwise contrasts. Across the six RT sessions, this yielded a total of 168 trials per contrast. One RT session was performed following every 3rd SAT session.

Procedure

Stimulus presentation, response collection, and feedback were controlled by an IBM–AT compatible computer. The consonants in a study list were sequentially presented in lower-case characters ($\approx 10 \times 6$ mm) in the center of the screen. The test probe consisted of two lower-case characters presented simultaneously in a side-by-side arrangement, separated by a 20-mm space.

The sequence of events that constituted a trial in each of the two tasks is schematically presented in Figure 2 and was as follows: (a) A centered, square fixation point was presented for 500 ms. (b) Each study consonant was sequentially presented for 300–500 ms, followed by a 50-ms blank screen. (c) An asterisk was presented for 500 ms to mask the final study item, alert subjects to a pending test, and help maintain fixation on the center of the screen. (d) The two items of the test probe were displayed in a left–right arrangement. (e) In the RT sessions, the probe items remained on the screen until a subject pressed one of two keys indicating which of the left–right test items was judged more recent. In the SAT sessions, the probe remained on the screen for either 0.15, 0.35, 0.55, 0.75, 1.0, 1.5, or 3 s at which point the screen cleared and a brief (10-ms) tone sounded to cue subjects to respond. Subjects responded by pressing one of two keys indicating which item was judged more recent. (f) Following a response, latency feedback was given. In the RT task, it consisted of the RT to the probe. In the SAT task, feedback consisted of the latency to respond to the interruption cue. Subjects were instructed to respond to the cue within 270 ms. They were informed that responses longer than this were too long and that responses under 130 ms were anticipations. A key press initiated the next trial.

All subjects were given at least 1-hr practice with each type of session. Initially the rate of presentation for items in the study list was set at 500 ms. For subjects BM and GR, this rate was maintained throughout all sessions. The rate was reduced to 400 ms for subject LB and to 300 ms for subject EC in an attempt to reduce near-perfect performance in some conditions.

Data Analysis

In a 2AFC task, such as the relative JOR task, accuracy can be transformed into a *d'* measure by either assuming symmetry or asymmetry of the response alternatives and an underlying signal distribution (Green & Swets, 1966). An asymmetric *d'* scaling accommodates bias to one of the response alternatives: Assuming equal variance Gaussian distributions, $d' = [z(1 \mid 1) - z(1 \mid 2)] \div 2^{1/2}$, where z is the standard normal deviate of the probability of responding that the most recent item was the first alternative, given that the most recent item was either the first $(1 \mid 1)$ or the second

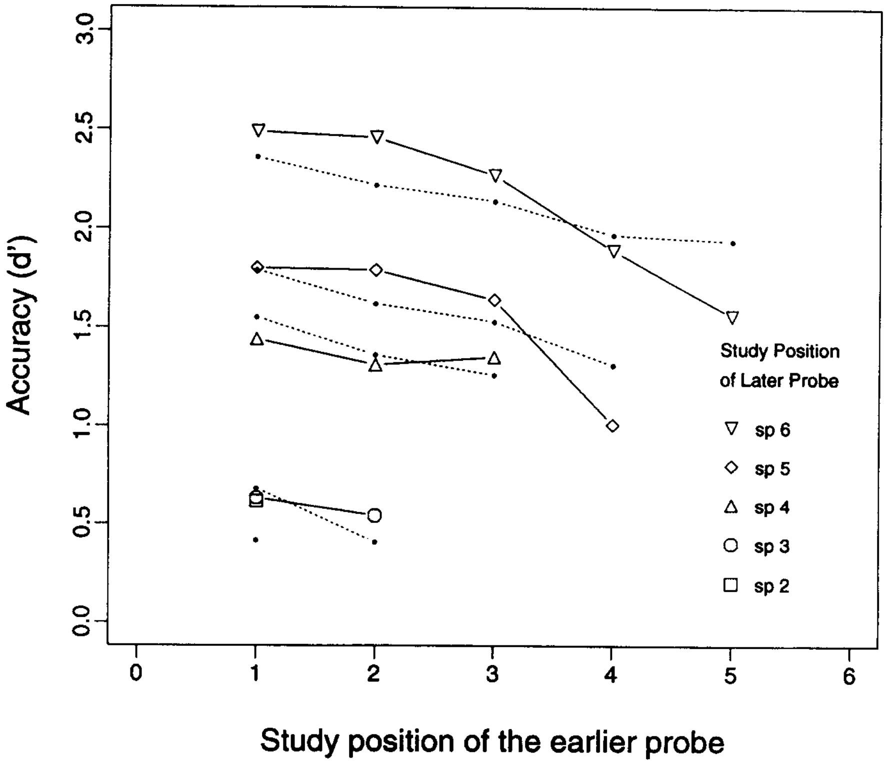

Figure 3. Average (over subjects) *d'* accuracy in the reaction time variant of the judgment of recency task (see Experiment 1a). (The study position [sp] of the earlier [less recent] test item is plotted on the abscissa. The curve parameter is the sp of the later [more recent] test item. Open symbols connected by solid lines show observed data. Dashed lines show fits of Equation A1 from Hacker's [1980] serial self-terminating scan model [see Discussion section and Appendix]).

(1 | 2) alternative. (Empirical estimates of d' are bounded by a correction that substitutes a .001 error rate for perfect performance.) Analyses with symmetric d' scaling, asymmetric d' scaling, and proportion correct all yielded identical patterns. We report results in the asymmetric d' format, except when presentation order is of explicit interest in which proportion correct is used.

Descriptive equations (see Equations 2 and 3) and model-based equations (see Appendix) were fit to the data with an iterative hill climbing algorithm (Reed, 1973), similar to Stepit, that minimized the squared deviations of predicted from observed data. When fits with differing numbers of (nested) parameters were compared, we evaluated the quality of the respective fits with three criteria: (a) The value of an R^2 statistic was computed.

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

where the d_i are the observed data 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 (Reed, 1976). This R^2 statistic is the proportion of variance accounted for by the fit, adjusted for the number of free parameters [k]. It is the same equation for adjusted r^2 often cited in multiple linear regression (e.g., Judd & McClelland, 1989). (b) We evaluated the consistency of parameter estimates across subjects. (c) Most important, we examined whether a fit yielded systematic (residual) deviations that could be accounted for with more parameters.

Experiment 1a Results: RT JOR Task

Accuracy

Figure 3 presents the average (over subjects) d' in the RT task as a function of the study position of the respective probes in a test pair. Table 1 presents the corresponding individual subjects' data. Although expressed in d' rather than percent correct, these data are replications of Hacker (1980) and Hockley (1984) with somewhat different timing and display characteristics; hence, we present an abbreviated statistical treatment. The data shown in Figure 3 illustrate the two major phenomena of prior JOR accuracy data: (a) Accuracy decreased dramatically as the later probe was drawn from less recent study positions, $F(4, 13) = 15.7, p < .005$, collapsing across study position of the earlier probe. (b) Holding the later probe constant, the accuracy is also weakly affected by the difference in recency between the two probes. This trend was significant for pairs involving List Position 6, $F(4, 12) = 3.62, p < .05$, and Position 5, $F(3, 9) = 13.5$, $p < .05$. There were small 2%–6% differences in the accuracy for the two presentation orders (more recent item on the right vs. more recent item on the left) that differed for different list positions of the most recent item. However, these appear to be speed–accuracy trade-offs as the RT differences are in the opposite direction (see below). Explicit fits of Hacker's model to the RT-accuracy data are deferred to the Discussion section and after the RT mean and distributional analysis and the SAT data are presented.

Table 1
Accuracy (d') in Reaction Time Experiment 1a

Study position of earlier probe	Study position of later probe				
	2	3	4	5	6
Subject BM					
1	0.417	0.159	1.896	2.355	3.193
2		0.192	1.905	2.355	3.555
3			2.016	1.997	3.553
4				0.870	2.966
5					2.660
Subject EC					
1	0.721	1.116	1.436	1.534	2.355
2		0.677	1.472	1.655	2.545
3			1.436	1.586	2.211
4				1.054	2.077
5					1.896
Subject GR					
1	1.175	1.294	1.547	2.355	3.193
2		1.116	1.333	2.187	2.966
3			1.669	1.796	2.545
4				1.620	2.798
5					1.749
Subject LB					
1	0.227	0.124	1.054	1.366	1.950
2		0.263	0.745	1.322	1.788
3			0.629	1.299	1.705
4				0.669	0.971
5					0.701

Mean RT

The number of errors in many paired comparisons fell well below a level that would ensure reliable RT functions for errors: Eight conditions had less than 100 errors pooled across all subjects. Consequently, analysis and discussion of RT is restricted to correct trials only.

Average correct RT as a function of study position is presented in Figure 4. Table 2 presents the individual subject data on which the average data is based. Figure 4 clearly illustrates that correct RT was dependent on the study position of the later probe, $F(4, 12) = 12.49, p < .05$, collapsing across study position of the earlier probe. There was little impact of the earlier probe. Subjects were faster on average at responding to trials in which the presentation order accorded with the study order (868 ms vs. 928 ms), $F(1, 3) = 26.89$, $p < .05$. As noted earlier, accuracy varied in the opposite direction, suggesting that this difference likely reflects a speed–accuracy trade-off.

The pattern of correct RT cleanly replicates the results reported in prior studies (Muter, 1979, 1980; Hacker, 1980; Hockley, 1984). The effect of the study position of the later probe on mean RT is decidedly nonlinear and is adequately fit ($R^2 = .965$) as a logarithmic function of recency: RT = 510 + 428 × log(sp), where sp denotes the number of items, including the test item, that intervene between study and test. Strict serial models often predict linear functions for experimental variables that determine the number of component serial processes (e.g., set size in Sternberg's [1966] serial

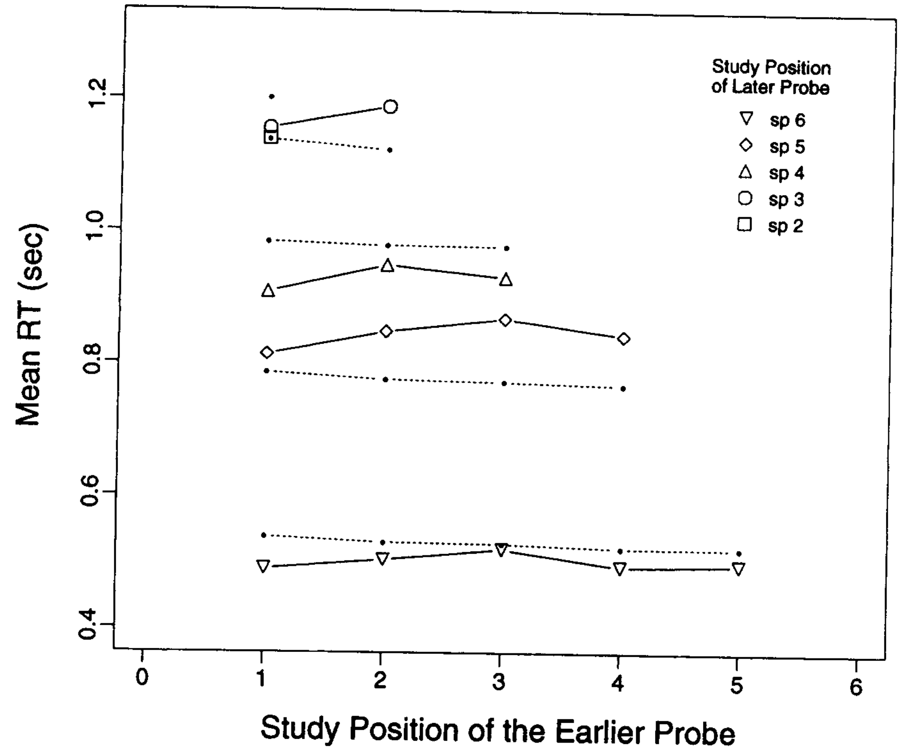

Figure 4. Average (over subjects) mean reaction time (RT) in the RT variant of the judgment of recency task (see Experiment 1a). (The study position [sp] of the earlier [less recent] test item is plotted on the abscissa. The curve parameter is the sp of the later [more recent] test item. Open symbols connected by solid lines show observed data. Dashed lines show fits of Equation A2 from Hacker's [1980] serial self-terminating scan model [see Discussion section and Appendix]).

exhaustive scan model). However, nonlinearity is an intrinsic property of serial mechanisms of the type proposed by Hacker (1980) in which the expected number of serial comparison processes is weighted by graded memory availability parameters. For this class of models, a nonlinear function is predicted in all cases in which availability in memory monotonically declines with study position.

RT Distributions

Increasing the number of serial matching or comparison processes should shift the entire distribution, including the leading edge, the mode, and the skewed right tail, toward longer times (Hockley, 1984; Hockley & Corballis, 1982; Ratcliff & Murdock, 1976; Sternberg, 1975). However, differences in mean RT often result from changes in the skewed right tail of the distribution only. Increases in the tail of the distribution are consistent with models in which only a proportion of trials is affected. Serial processing components are assumed to be common to all trials (Ratcliff & Murdock, 1976).

A number of methods have been proposed to fit and model RT distributions (see Luce, 1986b). RT distributions of recognition judgments appear to be extremely well fit by the convolution of a Gaussian and an exponential distribution (Ratcliff & Murdock, 1976, following Hohle, 1965):

$$f(t) = \frac{e^{-[(t-\mu)/\tau]+\sigma^2/2\tau^2}}{\tau(2\pi)^{1/2}} \int_{-\infty}^{[(t-\mu)/\sigma]-\sigma/\tau} e^{-y^2/2}\, dy. \qquad (2)$$

The exponential component, with parameter τ, captures the tail of the distribution: Large τs reflect long tails. The Gaussian, with its two parameters, namely the mean (μ) and variance (σ), roughly quantify the mode and leading edge of the RT distribution.

RT distributions were calculated for individual subjects and for group data for each of the 15 test probe conditions. Figure 5 shows group RT distributions for all positions of the later (more recent) probe, collapsing over the position of the earlier probe. Rather than defining equal-interval bins and calculating response frequency, a more stable method was used that defines 15 equal-probability bins and adjusts the width of time bins accordingly (Ratcliff, 1979). Group distributions were based on vincentized averaging (Ratcliff, 1979).

The study position of the later probe produced large and stable effects on all three aspects of the RT distributions, as summarized by the three parameters of the ex-Gaussian equation. Figure 5 shows that as the probe is drawn from more remote study positions, the leading edge and mode of

Table 2
Correct Mean Reaction Times (RTs; in Milliseconds) and Standard Deviations in Experiment 1a

	Study position of later probe									
	2		3		4		5		6	
Study position of earlier probe	RT	*SD*	RT	*SD*	RT	*SD*	RT	*SD*	RT	*SD*
					Subject BM					
1	1,459	50	1,408	44	928	37	854	26	501	19
2			1,389	49	914	34	829	28	482	19
3					930	34	870	29	521	22
4							992	42	467	14
5									552	31
					Subject EC					
1	995	51	1,061	59	930	49	831	49	383	11
2			1,200	47	951	41	898	40	460	22
3					980	45	887	43	457	22
4							870	34	443	19
5									423	15
					Subject GR					
1	1,320	51	1,370	52	1,254	55	1,069	52	635	35
2			1,448	53	1,384	61	1,136	57	638	32
3					1,216	55	1,177	54	657	32
4							953	44	623	32
5									577	25
					Subject LB					
1	655	45	694	38	475	19	459	23	428	13
2			620	41	481	23	492	19	424	12
3					497	26	511	40	433	17
4							514	25	418	11
5									407	15

the distributions (μ and σ) are shifted toward longer times. The right tails of the distributions (τ) are stable beyond Study Position 6, but there is a sharp discontinuity between the most recent position and all others.[2] Following the pattern of mean RT, the position of the earlier probe had little impact on the shape of the RT distribution (for a full analysis of all conditions for individual subject data, see McElree & Dosher, 1991). Muter (1980), Hacker (1980), and Hockley (1984) report similar shifts in the leading edge and mode with the later probe.

Differences in the leading edge and mode of the distributions, in the context of the observed pattern of mean RT and accuracy, are consistent with a serial self-terminating retrieval mechanism (Hacker, 1980; Hockley, 1984; Ratcliff & Murdock, 1976; Sternberg, 1973). Unfortunately, they are also consistent with certain classes of parallel self-terminating mechanisms.

The shape of the RT distribution also may be contingent on speed–accuracy criteria (Luce, 1986a; see also, Ratcliff, 1978, on criterion shifts and the predicted shape of the RT distribution for the random walk model). For example, whereas 3 of the 4 subjects' data were completely consistent with the average pattern in Figure 5, subject LB's data showed little change in the leading edge and the mode of the distributions but large changes in the right tail. LB's responses were much faster, but far less accurate, than the other 3 subjects. LB's empirical SAT retrieval functions are not different from other subjects'. So we conclude that her RT data reflect a shift in criterion rather than a different retrieval process. We suggest that this examination of individual subject's data provides a clear demonstration of the need to interpret RT distributional patterns only in the light of collateral measures of controlled retrieval speed.

Experiment 1b Results: SAT Retrieval Functions

The serial and rate-varying parallel accounts of JOR are both compatible with even the detailed analysis of RT data. The full time-course SAT data will discriminate between these differing accounts of JOR. This section presents a descriptive summary analysis of the results from the SAT task.

SAT Retrieval Functions

Asymptotic accuracy. Figure 6 presents d' values averaged across subjects and presentation orders at the longest interruption point (3 s) as an empirical measure of SAT asymptotic accuracy. Presentation order did not systematically affect asymptotic accuracy, $F(1, 3) < 1$, $p > .05$. As

---

[2] Position 6 is accorded special treatment in subsequent model fits.

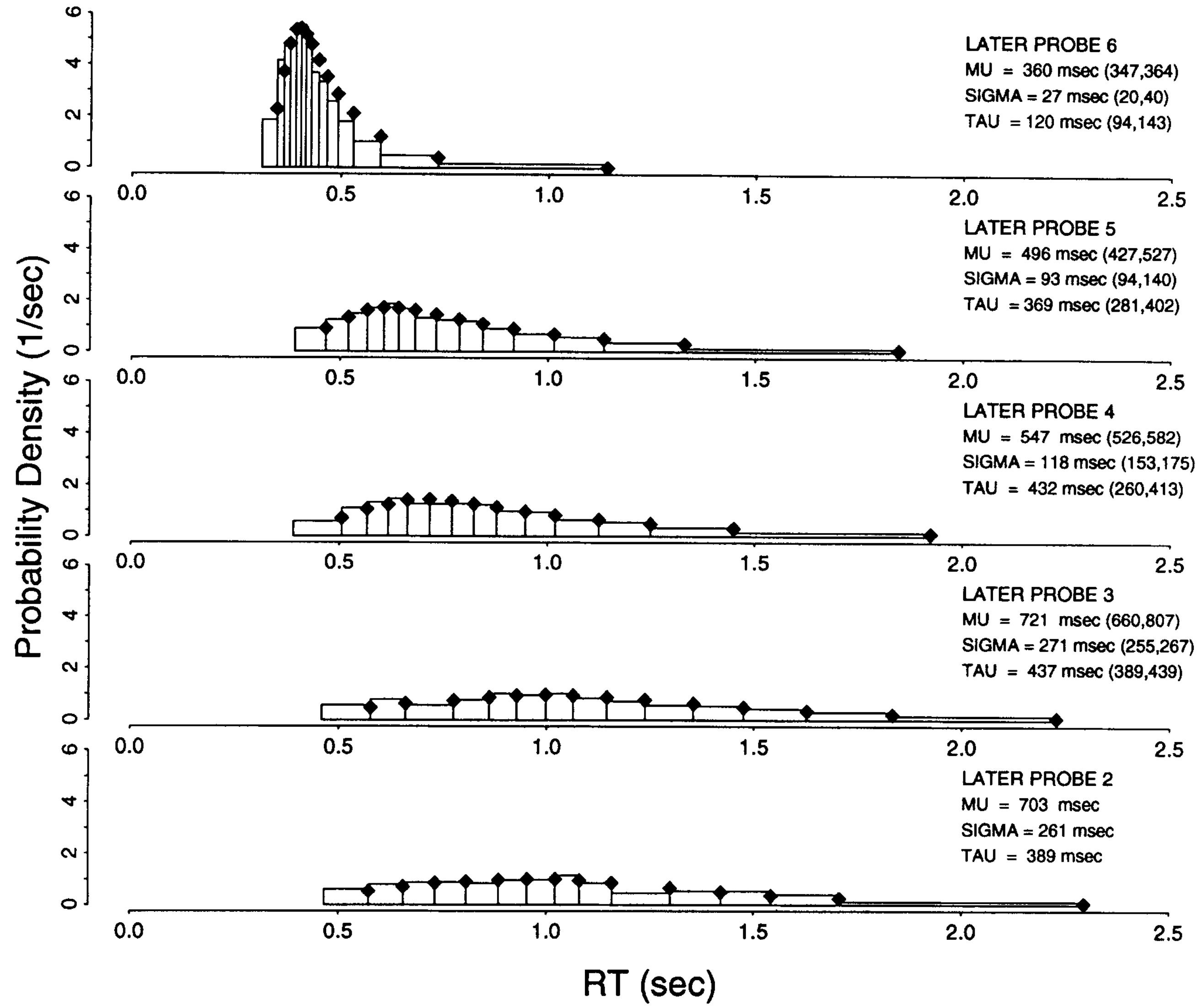

Figure 5. The average (over subjects) reaction time (RT) distributions as a function of the study position of the later (more recent) test probe, collapsing over the study position of the earlier probe. (The bars show the estimated probability density at each of 15 time quantiles spanning the range of RTs. Solid diamonds show fits of the ex-Gaussian model [see Equation 2]. The estimated values of μ, σ, and τ [and range over subjects] are listed in each panel).

with the RT task, proportion correct significantly varied with the study position of the later probe, $F(4, 12) = 9.8$, $p < .05$, collapsing over the position of the earlier probe. Unlike the RT task, however, accuracy did not significantly vary with the earlier probe when the study position of the later probe was held constant; all Fs < 1. Accuracies observed in the RT task were substantially lower than those at SAT asymptote, indicating that subjects in the RT task trade speed for accuracy.

Retrieval dynamics. Figure 7 presents the SAT functions averaged over subjects for each pairwise combination of the study positions. Panel A shows SAT functions for tests in which Item 6 is most recent, Panel B for tests in which Item 5 is most recent, and so forth. Estimated d' is graphed as a function of total processing time (i.e., the average time from test onset to response). Latency to the interruption tone depended on the recency of the later probe (see McElree & Dosher, 1991, for individual subject's data and statistics). The latencies averaged over lag and comparisons were 181 ms, 202 ms, 208 ms, 222 ms, and 222 ms for tests involving Position 6, 5, 4, 3, and 2, respectively. Condition differences of this magnitude are not typical for SAT data. However, excluding latency in the SAT analysis would systematically underestimate—though far from eliminate—dynamics differences between conditions, which are an order of magnitude larger than the latency differences.

Empirical SAT retrieval functions of the form in Figure 7 can be closely approximated by an exponential approach to a limit (Dosher, 1976, 1981, 1982, 1984; McElree & Dosher, 1989; Reed, 1976; Wickelgren, 1977;

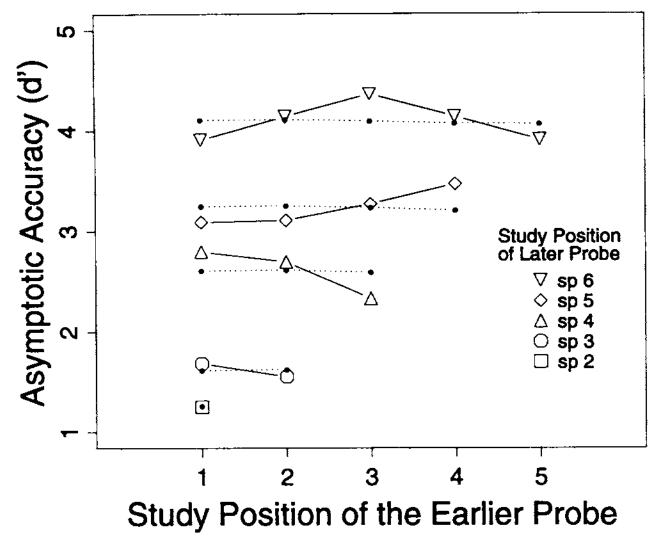

Figure 6. Average (over subjects) asymptotic d' accuracy in the speed–accuracy trade-off (SAT) variant of the judgment of recency task (see Experiment 1b). (The study position [sp] of the earlier [less recent] test item is plotted on the abscissa. The curve parameter is the sp of the later [more recent] test item. Open symbols connected by solid lines show observed data. The longest interruption point [3 s] was used as an empirical estimate of SAT asymptotic accuracy. Dashed lines show fits of Equation A1 from Hacker's [1980] serial self-terminating scan model [see Discussion section and Appendix]).

Wickelgren et al., 1980):

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}), \qquad t > \delta \quad \text{else} \quad 0, \qquad (3)$$

where λ represents the asymptotic accuracy level, δ the intercept or time before which accuracy is at chance level, and β represents the exponential rate parameter, indexing the speed with which accuracy rises from chance to asymptotic level. The parameters β and δ jointly serve to describe the dynamics of retrieval. Equation 3 quantitatively summarizes the impact of study position on the dynamics and asymptote of the SAT function.[3]

Both a rate-varying parallel comparison model and a backward serial self-terminating scan model predict differences in SAT rate of rise to asymptote (β). If the substantial differences in mean RT for JOR reflect a serial model, then we must observe large shifts in intercept (δ). If rate-varying parallel processing accounts for the differences in mean RT, then we should observe no shifts in intercept but rather substantial shifts in rate.

Examination of the data in Figure 7 reveals large differences in SAT intercept when the functions depart from chance performance. Considering the first interruption point (0.15 s), accuracy for contrasts involving later probes from Position 6 (Panel A) is well above 1.0 d' units, whereas contrasts involving Positions 2 and 3 (Panel D) are at chance and remain so until the third interruption point (0.55 s). Contrasts involving Positions 5 (Panel B) and 4 (Panel C) yield intermediate results. These shifts in intercept, quantified in the fits of the exponential, rule out parallel models, including those with variation in processing rate.

To fit all the data in Figure 7 with the exponential, it was necessary for the asymptotes, rates, and intercepts to vary with condition. Fits that ignored differences in any of the three parameters produced systematic misfits and relatively low R^2 values. Through extensive comparisons of competitive fits, a 5 asymptote, 15 rate, and 3 intercept (5λ-15β-3δ) model emerged as the best description of the data.

Estimated asymptotes differed only with the position of the later probe as shown in Figure 6. The position of the earlier probe had no measurable effect.

Estimated intercepts varied over an exceptionally wide range, differing by as much as ½ s (186–683 ms) over test probes. Although we can not rule out a fully graded pattern of five intercepts—one for each position of the later probe—only a subset of these differed enough to be significant. The best competitive fit yielded three clusters of intercepts: (a) a low intercept for contrasts with a later probe from the most recent study position (Position 6) estimated at 186 ms in the average data (ranging from 151 to 280 ms across subjects); (b) a middle intercept estimated at 283 ms (ranging from 233 to 500 ms across subjects) for contrasts with the later probe from Position 5 or 4; and (c) a long intercept estimated at 683 ms (ranging from 577 to 1,021 ms across subjects) for later probes from Positions 2 and 3.

In addition to the large intercept differences, the speed of rise to asymptote (β) also depends on the position of the later probe. Separate examination of each panel in Figure 7 reveals that accuracy is better at earlier interruption points when there is more separation in recency between the test probes. This secondary effect was reflected in the rate estimates.

The 5λ-15β-3δ fit accommodates these three major properties of the data. The parameter estimates and R^2 statistics are shown in Table 3 for the average data and for individual subjects. The lines shown in Figure 7 correspond to this fit. All subjects show essentially the same patterns of data.

The largest estimated dynamics (i.e., rate and intercept) difference is between probes involving the most recent position (Position 6) and all other positions. Test probes involving an item from Position 6 are cases of an immediate repetition of a studied item and an element in the test pair. Wickelgren et al. (1980) and McElree and Dosher (1989) observed a similar rate advantage for immediate repetitions in an item recognition task, as did Dosher (1981) for a paired associate recognition task. Wickelgren et al. suggested that the most recently studied item remains in an active or primed state if no interfering mental activity intervenes between study and test. Recognition of an active item reflects a matching process in which normal retrieval processes may be circumvented. The analysis of the RT distributions also supports this view. The τ parameter of the ex-Gaussian model, reflecting the positive skew of the distribution, was markedly smaller for cases of immediate

---

[3] An alternative and similar equation results from the time-bounded diffusion process (Ratcliff, 1978; see also Dosher, McElree, Hood, & Rosedale, 1989, and McElree & Dosher, 1989).

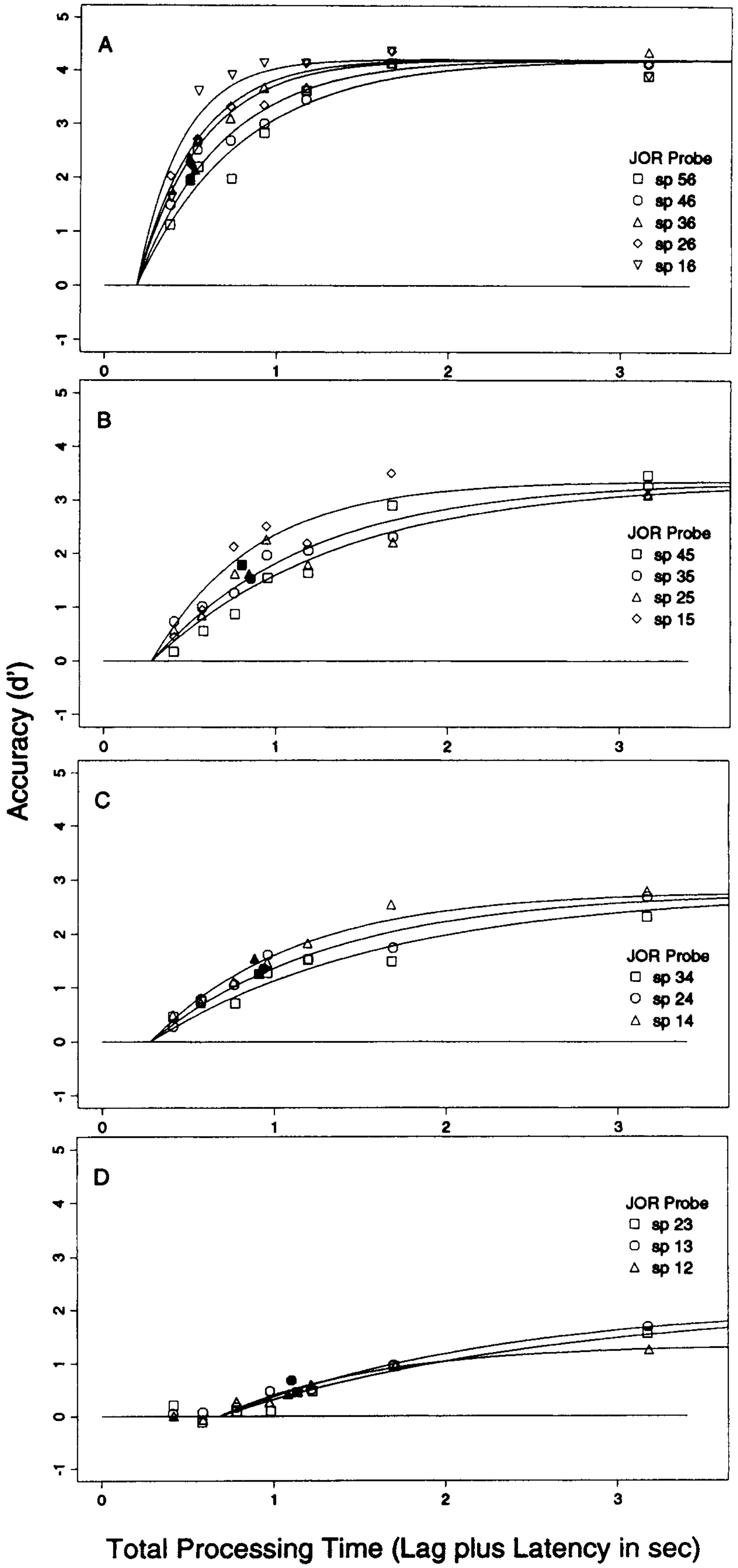
A
JOR Probe
□ sp 56
○ sp 46
△ sp 36
◇ sp 26
▽ sp 16
B
JOR Probe
□ sp 45
○ sp 35
△ sp 25
◇ sp 15
C
JOR Probe
□ sp 34
○ sp 24
△ sp 14
D
JOR Probe
□ sp 23
○ sp 13
△ sp 12
Accuracy (d')
-1
0
1
2
3
4
5
0
1
2
3
Total Processing Time (Lag plus Latency in sec)

Table 3
Exponential Fits of Speed–Accuracy Trade-Off Judgment of Recovery Data

	Subject														
	AV			BM			EC			GR			LB		
Study position	λ	β	δ	λ	β	δ	λ	β	δ	λ	β	δ	λ	β	δ
5–6	4.21	1.64	.186	4.21	1.57	.151	4.11	2.98	.215	4.12	2.68	.257	4.45	0.83	.217
4–6	4.21	2.02	.186	4.21	1.55	.151	4.11	5.79	.215	4.12	2.42	.257	4.45	1.79	.217
3–6	4.21	2.66	.186	4.21	3.03	.151	4.11	3.07	.215	4.12	3.67	.257	4.45	1.99	.217
2–6	4.21	2.93	.186	4.21	5.12	.151	4.11	3.39	.215	4.12	2.96	.257	4.45	1.79	.217
1–6	4.21	3.87	.186	4.21	3.54	.151	4.11	4.69	.215	4.12	4.29	.257	4.45	4.74	.217
4–5	3.36	0.89	.276	2.78	0.35	.285	3.39	1.54	.493	4.19	1.03	.394	3.28	1.37	.172
3–5	3.36	1.07	.276	2.78	2.21	.285	3.39	0.78	.493	4.19	0.87	.394	3.28	1.49	.172
2–5	3.36	1.08	.276	2.78	1.46	.285	3.39	0.92	.493	4.19	0.73	.394	3.28	2.30	.172
1–5	3.36	1.67	.276	2.78	2.04	.285	3.39	1.10	.493	4.19	1.82	.394	3.28	2.89	.172
3–4	2.79	0.72	.276	2.52	0.82	.285	3.82	0.42	.493	3.26	0.61	.394	2.12	1.32	.172
2–4	2.79	0.94	.276	2.52	1.38	.285	3.82	0.87	.493	3.26	0.67	.394	2.12	1.52	.172
1–4	2.79	1.19	.276	2.52	2.33	.285	3.82	0.66	.493	3.26	1.23	.394	2.12	1.68	.172
2–3	2.13	0.51	.685	1.23	0.47	.967	2.18	0.61	.577	3.77	0.29	.639	1.81	1.18	.971
1–3	2.13	0.62	.685	1.23	0.45	.967	2.18	0.61	.577	3.77	0.62	.639	1.81	0.56	.971
1–2	1.34	1.13	.685	0.50	1.29	.967	2.07	0.64	.577	1.78	1.54	.639	1.18	0.25	.971
R^2		.964			.900			.923			.853			.871	

repetition (Position 6) and did not systematically vary with study positions beyond the most recent. A matching process presumably would be both faster and less variable than a standard retrieval operation yielding a lower overall positive skew.

Even if conditions involving an immediate matching process (Position 6) are excluded from consideration, the intercept shifts among the remaining conditions are quite large (400 ms). Subsequent sections provide detailed discussions of the implications of the intercept and rate shifts for parallel models and evaluate the serial model of Hacker (1980) as an account of these time-course data.

Discussion

Empirical Summary

The RT task replicated all of the principal results reported in the previous studies (Muter, 1979, 1980; Hacker, 1980; Hockley, 1984). Accuracy increased dramatically as the later probe was drawn from more recent study positions and to a much lesser extent as the separation in recency between the later and earlier probe was increased. Mean RT was affected by the study position of the later probe, decreasing logarithmically with recency, yet was uninfluenced by the earlier probe. Fits of the descriptive ex-Gaussian model demonstrated that differences in mean RT primarily reflected differences in the mode and leading edge of the RT distribution. The positive skew of the RT distribution remained fairly stable beyond comparisons involving items from Study Position 6 (a case of immediate repetition between study and test).

The study position of the later probe had a large impact on both the asymptotic and dynamic components of the SAT retrieval functions. Asymptotic accuracy increased dramatically as the later probe was drawn from more recent study positions. The dynamics of retrieval also varied directly with the recency of the later probe, reflected in substantially shorter intercepts and faster rates of approach to asymptote for more recent (later) probes. The study position of the earlier probe did not affect asymptotic levels but did introduce small but reliable differences early in retrieval. Accuracy rose to asymptote at a faster rate as the earlier probe was drawn from more remote positions where the separation in recency between probes was largest.

Cross-Tasks Comparisons

There is, in general, a strong relationship between data of the RT and SAT tasks in the current experiments. This is consistent with prior studies that have run comparable experiments using both tasks (Dosher, 1982; Dosher et al., 1989; McElree & Dosher, 1989; Reed, 1976). We begin with an examination of mean RT and accuracy, then further discuss aspects of the RT distributions.

Figure 7. Average (over subjects) *d'* accuracy in the judgment of recency (JOR) task (see Experiment 1b) as a function of processing time (i.e., lag of the interruption cue plus latency of response to the cue). (Open symbols show observed *d'* for earlier probes paired with later probes from Study Positions [sps] 6 [Panel A], 5 [Panel B], 4 [Panel C], 3, and 2 [Panel D]. Solid lines show the best fitting exponential retrieval functions [see Equation 3] by using the parameter values for the average data listed in Table 3. The corresponding filled symbols show reaction time [RT] data [observed *d'* at observed RT] from Experiment 1b).

To the degree that both RT and SAT tasks reflect common encoding and retrieval processes, RT points should lie near the corresponding SAT curves (Dosher, 1982; Reed, 1976; Wickelgren, 1977). RT points (d' at observed RT) are shown as solid symbols on the plots of the full SAT retrieval functions in Figure 7. (The RT points are for mean correct RT; points including error RTs are similarly located, perhaps showing slightly more spread over conditions.) Inspection of Figure 7 shows that the RT points indeed lie quite close to observed SAT functions. There are a number of reasons why we may not necessarily observe perfect alignment of RT points with the SAT retrieval functions (see Dosher, 1982). In this case, however, the alignment is remarkably close.

Figure 7 illustrates that SAT asymptotes are higher than the corresponding RT accuracies across all conditions. This is again a standard finding in direct comparisons between the two methods (Dosher, 1982; Reed, 1976). In an RT task, subjects typically select points on the speed–accuracy operating characteristic that trade relatively large gains in speed for modest decrements in accuracy. Here, possibly because of the very lengthy processing for the most difficult conditions, the RT accuracies show substantial decrements ranging from about half of SAT asymptote for tests involving the more recent probes to about one quarter of SAT asymptote for tests of the first few list items. Under continuous retrieval models, trading speed for accuracy in an RT task reflects the setting of a criterion or criteria for information necessary to complete a comparison and elicit a response. Under a discrete retrieval model, the trade-off results either from fast guessing or from approximate time deadlines, which serve to replace longer responses with simple or partially informed guessing.

One superficial difference between SAT and RT accuracy concerns the effect of the study position of the earlier probe: RT accuracy significantly improved as the separation in recency between the two probes increased, whereas the SAT functions show no differences at asymptote as a function of the earlier probe. Differences in RT accuracy for earlier probes appear to reflect underlying dynamics differences rather than differences in availability. Extrapolating from the SAT functions, we predict that if subjects in an RT task operated close to asymptote, then these differences in accuracy would be eliminated or greatly reduced. In a weakly related observation, Muter (1980) reported no effects of the earlier probe for high-confidence responses.

Considering the RT-task data in more detail, we found that shifts in the leading edge and mode of the RT distributions parallel large dynamic differences in the SAT data (measured as intercept δ and rate β in Equation 2) as a function of the later probe. Previous cases of cross-subject and experiment comparisons (Hockley, 1984; Hockley & Corballis, 1982; McElree & Dosher, 1989) of item recognition show that large SAT asymptotic differences in the presence of minimal dynamics differences were coupled with RT-distribution shifts in positive skew or tail and minimal shifts in leading edge or mode. It is tempting, then, to associate shifts in leading edge and mode with large SAT dynamics effects. Unfortunately, empirically the relationship between characteristics of RT distributions and particular properties of the full time-course measures may be significantly more complex. For example, Dosher (1984, and unpublished RT-distribution data) found shifts in leading edge and mode of RT distributions for visual embedding (a form of masking), which corresponded to shifts in rate, but not intercept, of comparable SAT functions. In contrast, subject LB in the current experiments illustrates a case in which both intercept and rate of the SAT data show substantial shifts as a function of position of the probes, but there are minimal shifts in the leading edge and mode of the RT distributions. (The leading edge was not a function of fast guessing for LB; accuracy for RTs in the fastest quarter of the trials was almost as high as for slower trials.) So shifts in leading edge of the RT distribution cannot be directly mapped into shifts in SAT intercept on an empirical basis.

From a theoretical perspective, under discrete retrieval models (including Hacker's serial scan model), RT distributions and SAT dynamics should both reflect the same underlying distribution of finishing time distributions; early portions of the RT distributions and intercept and rate of the SAT are both associated with the initial portions of the finishing time distribution. However, Ratcliff's (1978) model predicts that shifts in leading edge and mode may occur as a function of asymptotic differences alone. Hence, model-based conclusions are also premature. Additional case studies that compare RT distributions and SAT dynamics, as well as explicit modeling of trade-offs in speed and accuracy in RT paradigms, are necessary before we can propose guidelines for the interpretation of RT distributions in the absence of converging information about SAT functions.

Parallel Models

Both Muter (1979) and Hacker (1980) acknowledge that the pattern of mean RT may be consistent with a parallel self-terminating mechanism. Comparison processes could act in parallel across the memory set, racing to reach a criterion. Subjects select as the most recent the item whose strength or attribute count first exceeds a criterion or threshold. If the rate of information accumulation is faster for more recently studied items, mean RT will be largely controlled by the study position of the later probe. As it stands, this account has not been developed sufficiently to determine whether such a model is capable of accounting for not only observed mean RT but also RT distributional patterns (see McElree & Dosher, 1991, for a discussion of a diffusion model account). However, one very salient feature of the SAT data is incompatible with parallel models—specifically, the clear and quite profound differences in SAT intercept that arise with large differences in the study position of the later probe. Although this result does not rule out contemporaneous processing models in which some processes are delayed in onset (cf. Shaw & Shaw, 1977), the intercept shifts do rule out all strictly parallel models. All strictly parallel models are unable to account for substantial differences in intercept in

Table 4
Estimated Availability (a_i) Parameters From the Speed–Accuracy Trade-Off (SAT) and Reaction Time (RT) Tasks

	Study position						
Subject	1	2	3	4	5	6	R^2
				SAT task			
BM	0.524	0.503	0.609	0.898	0.924	0.986	0.973
EC	0.899	0.885	0.875	0.961	0.964	0.996	0.806
GR	0.050	0.764	0.877	0.943	0.982	0.998	0.827
LB	0.980	0.778	0.824	0.914	0.976	0.996	0.805
Average	0.758	0.765	0.813	0.933	0.964	0.994	0.986
				RT task			
BM	0.100	0.241	0.228	0.866	0.816	0.979	0.911
EC	0.144	0.460	0.583	0.788	0.792	0.921	0.894
GR	0.368	0.702	0.741	0.824	0.897	0.950	0.773
LB	0.010	0.065	0.138	0.482	0.643	0.713	0.817
Average	0.091	0.356	0.444	0.749	0.796	0.899	0.891

other than an ad hoc manner.[4] By definition, in strictly parallel-processing models all comparisons are initiated at the same time and hence are associated with a common intercept. This class of models can be clearly rejected by the SAT data.

Serial Models

SAT intercept differences are the critical feature of the data that motivate a serial model. In this section, we present fits of the Hacker serial self-terminating model to mean RT, RT accuracy, and the SAT data to demonstrate that a variant of this particular serial model is capable of accounting for many, though not all, facets of the data.

Accuracy data. The serial self-terminating search model proposed by Hacker (1980) attributes differences in accuracy to the availability of items in the search or memory set. The availability parameters for each study position in the list were estimated from the accuracy data by using Equation A1 in the Appendix.

Table 4 presents availability parameters estimated from SAT asymptotic performance and RT accuracy. Predicted d' accuracy is shown in dashed lines in Figure 3 for RT and in Figure 6 for SAT. Consistent with standard assumptions concerning item decay, displacement, or both, the parameters monotonically and dramatically decline with more remote study positions.[5] The model clearly captures the large and reliable differences attributable to the later probe. The study position of the earlier probe has no effect on the SAT asymptotic performance in Figure 6 and has small effects on RT accuracy in Figure 3. In the model, the recency of the earlier probe affects performance only when the later probe is unavailable. Because all items appear in the memory representation with very high probability, there is little room for an impact of the study position of the earlier probe on SAT asymptotes. Absolute levels of the estimated availability parameters are lower, and the rate of decline is much faster for the RT task (see Table 4). Subjects in the RT task operate on a point on the SAT curve that is substantially less than the maximal asymptotic level of performance. The RT accuracies are lower than SAT asymptotes by a larger proportion for less recent items (cf. Figure 7). The estimates of availability from RT data are therefore contaminated by the speed of access.

Retrieval dynamics. In this section, we model the time-course and RT data with an extension of the Hacker serial search model. In fitting the SAT functions, we used the estimates of the six availability (a_i) parameters (see Table 4, top), together with several processing time parameters. Details of how the SAT functions were derived from the Hacker model are presented in the Appendix. Briefly, we assumed a base time for encoding and response. The search time for each comparison was modeled as an exponential distribution of comparison times, with mean corresponding to Hacker's search

---

[4] It might be assumed that recency serves to speed the encoding of the test probe (cf. Sternberg, 1975) and thereby affects the intercepts of the comparison processes. However, such an assumption is implausible because recency does not appear to affect encoding in other tasks such as item recognition (see Experiment 2; McElree & Dosher, 1989; Reed, 1976; Wickelgren et al., 1980) in which the same principle should hold.

[5] Although the parameters in Table 4 show a typical pattern of declining with more remote study positions, there is no evidence of a primacy effect for the first item in the list as is evident in the observed data. This may be due in part to parameter trade-offs in the fitting of Equation A1. Note that the primary list position enters into Equation A1 only as a component (a_j) in the estimated contribution of correct guessing for cases in which the other later item (a_i) in the test probe is not available. The availability estimate for the primary position is therefore derived only in relation to the availability estimates of the other list positions and as such may be subject to rather extreme trade-offs.

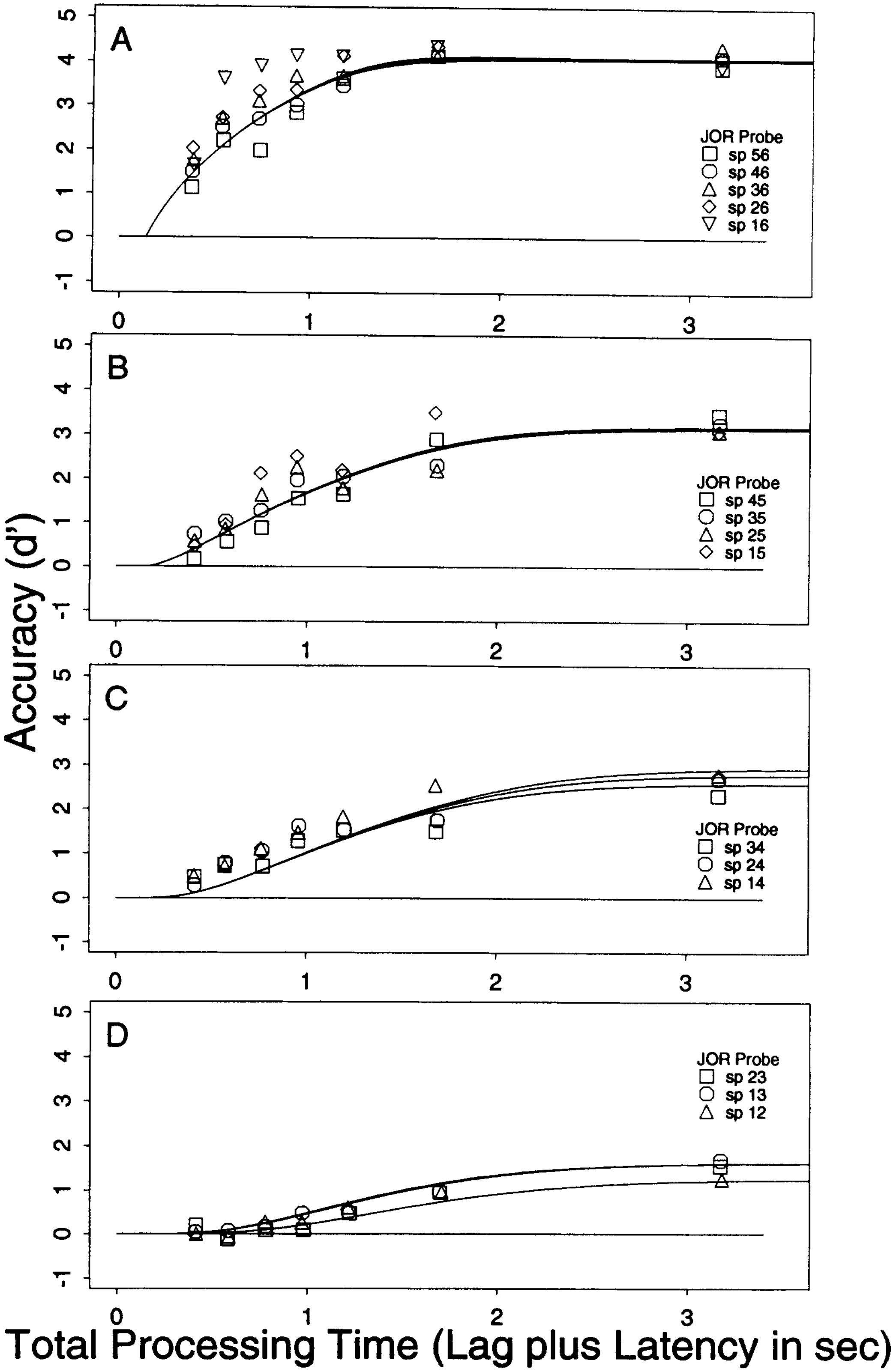

Figure 8. Predicted speed–accuracy trade-off functions from the embellished serial self-terminating scan model. (Open symbols represent average observed *d′* accuracy as a function of total processing time [i.e., interruption plus latency] for earlier probes paired with later probes from Study Position [sp] 6 [Panel A], 5 [Panel B], 4 [Panel C], 3 and 2 [Panel D]. Solid lines show predicted values based on the average parameter estimates in Table 4 [top panel] and Table 5 [top panel]. JOR = judgment of recency.)

time s.[6] We altered Hacker's model in two ways: first, with the addition of an immediate match process and second, in the treatment of guessing time.

Previously, it was argued that cases of immediate repetition may be plausibly viewed as a matching rather than a search or scan process, which is consistent with other SAT studies of short-term memory (Dosher, 1981; McElree & Dosher, 1989; Wickelgren et al., 1980). If a matching strategy was operative for cases of immediate repetition, then the retrieval dynamics for contrasts involving later probes from Study Position 6 should derive from a different underlying finishing time distribution. Consonant with this notion, we allowed a separate match process (assumed also to be exponentially distributed) for cases when the most recently studied item was presented in the test probe and available (a_6) in memory. Models that excluded the immediate match process resulted in systematic deviations and lowered R^2. Guessing was assumed to occur whenever information about both probes was unavailable either due to insufficient processing time or loss from the memory representation. Because of technical complications in joint estimation of comparison and guessing times for full distributions, we assumed that additional guessing time is subsumed in the latency to the interruption cue.

The resulting nine-parameter model for SAT yielded a good overall fit for all subjects, with an average R^2 of .930, ranging from .764 to .911 across subjects. Matching times were (with one exception) estimated to be substantially faster than the corresponding search time.[7] This model's predicted retrieval functions for the average data are shown as the smooth curves in Figure 8. The estimated parameters from this model are presented in Table 5.

Mean correct RT from the RT task was fit with a comparable version of the Hacker model. In fitting RT, we included a guessing time parameter as specified in the original Hacker formulation (see Appendix), along with a base time parameter, b, a search time parameter, s, and a matching time parameter for tests involving Position 6.[8]

The dashed lines in Figure 4 show the predicted correct RT for the average data. The model clearly captures the invariance of RT with earlier probe positions, yielding near-perfectly flat functions. Increments in mean RT with the recency of the later probe are also adequately captured, although there is some slight tendency to overestimate the differences for middling study positions and for the comparisons involving the primacy position. The model yielded R^2 values of 0.951 for the average data, ranging from 0.813 to 0.949 across subjects. The parameter values for the average and individual subjects are listed in Table 5.

Table 5
Estimated Base, Search, and Match Parameters (in Milliseconds) From the Speed–Accuracy Trade-Off (SAT) and Reaction Time (RT) Tasks

Subject	Base	Search	Match	Guess	R^2
		SAT parameters			
BM	100	322	190	—	0.801
EC	367	336	157	—	0.911
GR	399	274	201	—	0.833
LB	100	243	260	—	0.764
Average	136	299	202	—	0.930
		RT parameters			
BM	300	245	199	500	0.949
EC	200	238	123	100	0.858
GR	500	251	121	100	0.853
LB	372	30	13	272	0.813
Average	300	251	118	100	0.951

The estimated search parameters and, to a lesser degree, the matching parameters are lower for the RT data as compared with the SAT data. This is consistent with the trading of speed for accuracy in the RT performance. The Hacker model as currently formulated lacks a mechanism for the trading of speed for accuracy in the RT task and for predicting RT distributions. Nevertheless, the search times in both tasks are reasonably comparable with the 209 ms reported by Hacker (1980). In fits of the average and 3 of the 4 subjects' data, the estimated base encoding time, b, is substantially lower in the SAT task than in the RT task. The encoding time parameter in the RT model estimates the mean time to encode the target item and execute a response. In SAT, the comparable parameter estimates the minimum time needed to produce above-chance performance. Consequently, the SAT parameter is expected to be lower.

Inadequacies of the serial self-terminating model. The model listed in Table 5 and shown as smooth curves in Figure 8 gives an excellent account of the large shifts in SAT intercepts and moderate shifts in SAT rate resulting from the study position of the later, or most recent, test probe. However, there are systematic deviations between the model and the data that involve the impact early in retrieval of the study position of the earlier probe. The impact of the earlier probe, as the functions approach asymptote, are either eliminated (Positions 6 and 5) or greatly reduced (Positions 4, 3, and 2). The model does predict small dynamics differences in the

---

[6] A single search parameter, s, assumes that all comparison processes are independently and identically distributed. This assumption of course can be modified in a number of ways. For example, comparisons within a scan that result in a match might be distributed differently from those that result in a nonmatch. Alternatively, comparison rates may vary for different study positions, perhaps varying directly with availability (cf. Murdock, 1971; Muter, 1979). Neither of these more embellished models, however, substantially improved the quality of fit and, in fact, resulted in lower R^2 in the average and across individual subjects.

[7] The reversal in subject LB's data are due to the relatively slow rising function for the comparison involving items from Position 6 and 5. When this contrast is excluded from the fit, mean matching time is estimated at 235 ms, whereas mean comparison time is estimated at 270 ms.

[8] RT measures do not provide a direct assessment of retrieval speed and consequently do not provide direct support for the fast matching process evidenced in SAT. Nevertheless, identical study conditions were used in the RT and SAT experiments, so a similar matching strategy for cases of immediate repetition was almost surely operative in the RT task. Some independent support for a matching strategy in the RT task derives from the contracted right tail of the RT distribution for Position 6 (see Figure 5).

comparison of probes from Positions 1, 2, and 3. These small differences reflect guessing in the context of low overall availability parameters. However, the smooth functions in Figure 8 clearly illustrate that, for all other positions, the model predicts near-identical retrieval dynamics, regardless of the position of the earlier probe.

These differences in dynamics as a function of the earlier probe are small in comparison with those observed for the later probe. We believe they reflect the limited use of item information when scan-derived order information is unavailable. Instead of randomly guessing whenever the interruption cue occurs before a particular scan is complete or when the scan terminates without finding a match, subjects may respond on the basis of relative differences in available item strength (Hinrichs, 1970; Morton, 1968; Peterson, 1967) or attribute counts (Bower, 1972; Flexser & Bower, 1974). Because strength or the number of attributes covaries with recency, a greater proportion of correct responses will occur as the distance or separation in recency between the two probe items is increased and this effect will be greater for more recent later probes. Scan-related information becomes available later in retrieval, and so the contribution of item information should diminish near asymptote.

A similar reliance on item information may have occurred in the RT task, although in response to a different experimental demand. When RT data are situated in speed–accuracy coordinates (see Figure 7), it appears that a time sensitivity, or weak time deadline, affects the choice of when to respond. Item information may control the response on the proportion of trials that are terminated before adequate order information is retrieved. In the RT task, unlike the SAT task, we do not have any direct evidence for the use of item information. However, given the convergence between the two tasks—conjoint RT-accuracy points temporally align approximately with the corresponding SAT dynamics (see Figure 7)—it is likely that a similar strategy was operative in the RT task. In closing, we note that similar intrusions of item information have been documented in other related contexts. Ratcliff and McKoon (1989) reported a reliance on item information early in the time course of recognizing relational information from memorized sentences. Their data suggest that responses early in retrieval stem from item information but later in retrieval shift to relational information as it becomes available.

Summary

Although the pattern of mean RT and shapes of RT distributions are suggestive of a serial retrieval mechanism for the recovery of order, it is the large shifts in SAT intercept that rule out parallel mechanisms and provide clear evidence for the serial retrieval of order information. A modified form of Hacker's serial self-terminating model provides a good account of mean RT and RT-accuracy data. When tested against SAT data, it adequately predicts asymptotic performance and the large differences in dynamics (i.e., intercept and rate of rise) controlled by the recency of the later probe. Its primary failings are in modeling the smaller differences in rate of rise that are attributable to the recency of the earlier probe. However, these differences plausibly reflect the limited use of item information when order information is not available.

In the next section, we situate the findings concerning the recovery of order information in a larger memory context. In particular, we contrast the recovery of order information with the recovery of item information.

Experiment 2: 2AFC Item Recognition

SAT examinations of the retrieval of item information from STM (McElree & Dosher, 1989; Wickelgren et al., 1980) have consistently found that the recency of the test item primarily affects asymptotic memory strength. These studies found identical dynamics (intercept δ and rate β parameters in Equation 3) for all test items—regardless of item recency or list length—except for the most recently studied item. Tests involving this item were cases in which no other item intervened between study and test, allowing a direct, fast match of the test probe to the item active in awareness. The lack of an effect of either serial position or list length on retrieval dynamics is incompatible with the class of serial (exhaustive and self-terminating) retrieval mechanisms. Rather, item recognition is mediated by a parallel or direct-access retrieval mechanism. This is generally in agreement with a number of recent memory models, such as MINERVA2 (Hintzman, 1984, 1988), SAM (Gillund & Shiffrin, 1984), and TODAM (Murdock, 1982, 1983), which posit that recognition judgments are made by computing a global familiarity or goodness-of-match value.

Comparison of the item recognition and JOR tasks appears to indicate that item and order information are retrieved by different mechanisms, one direct or parallel and the other serial. A dissociation in the retrieval of item and order information places constraints on global models of memory and is of direct interest to models that have explicitly attempted to simultaneously account for the storage and retrieval of both types of information (e.g., TODAM, Lewandowsky & Murdock, 1989). However, the JOR task uses a 2AFC paradigm to assess the retrieval of order information, whereas prior SAT studies of item recognition have used a yes–no paradigm. To provide a minimal experimental contrast between the two types of judgments, we report an SAT study of item recognition using a 2AFC paradigm. Subjects study a list of sequentially presented consonants followed by a test probe consisting of two test items. Test probes consist of a new, nonlist item and an old item drawn from one of the six study positions in the list.

Method

Subjects

Four subjects participated in a total of eight 60-min experimental sessions. Two of the subjects, BM and LB, participated in Experiments 1a and 1b and in the SAT item recognition studies reported in McElree and Dosher (1989). All subjects except BM were paid for their services.

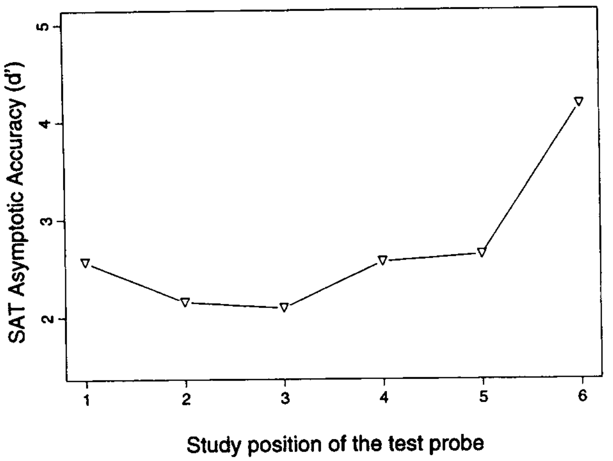

Figure 9. Average (across subjects) speed–accuracy trade-off (SAT) asymptotic accuracy in asymmetric *d*′ units for the two-alternative forced-choice item recognition task (see Experiment 2). (The longest interruption point [3 s] was used as an empirical estimate of asymptotic accuracy.)

Design, Stimuli, and Procedure

With a few exceptions noted below, the design, stimuli, and procedure were the same as Experiment 1b. The test probes consisted of two consonants, an old consonant from one of the study positions and a new consonant drawn at random from the 14 consonants not presented in the study set. Each session consisted of two blocks of 210 trials that sampled each study position equally often in each test (right–left) presentation order at each of the seven interruption points. Across the eight sessions, this yielded for each subject a total of 40 trials for each factorial combination of study position, test presentation order, and interruption lag.

The sequence of events in a trial are essentially the same as those in Experiments 1a and 1b. One exception concerns the interruption points where the test probe remained on the screen for either 0.1, 0.25, 0.4, 0.55, 0.9, 1.2 or 3 s. The rate of presentation for items in the study list varied from 0.2 to 0.3 s/consonant across subjects.

Results and Discussion

It is frequently suggested that 2AFC is a better or simpler method than a yes–no paradigm for assessing recognition memory, under the assumption that 2AFC eliminates issues of bias (see also Macmillan & Creelman, 1991, for a comparison of yes–no and 2AFC tasks). However, detailed analysis of the item recognition data—unlike the JOR data—indicate that not only is bias not eliminated, but the 2AFC presentation induces complex effects because of processing order of the two alternatives that would be avoided in yes–no assessment. A listing of an individual subject's data (latencies, proportion correct, and asymmetric *d*′) as well as a complete analysis of these presentation order effects are presented in McElree and Dosher (1991; see also Hockley, 1984, for the impact of a 2AFC task on RT distributions). Here, we focus directly on the issue of contrasting item versus order information.

SAT Retrieval Functions

Asymptotic accuracy. Performance at the longest interruption point (3 s) gives an empirical estimate of asymptotic performance. Figure 9 presents the average (over subjects) asymptotic performance by study position in asymmetric

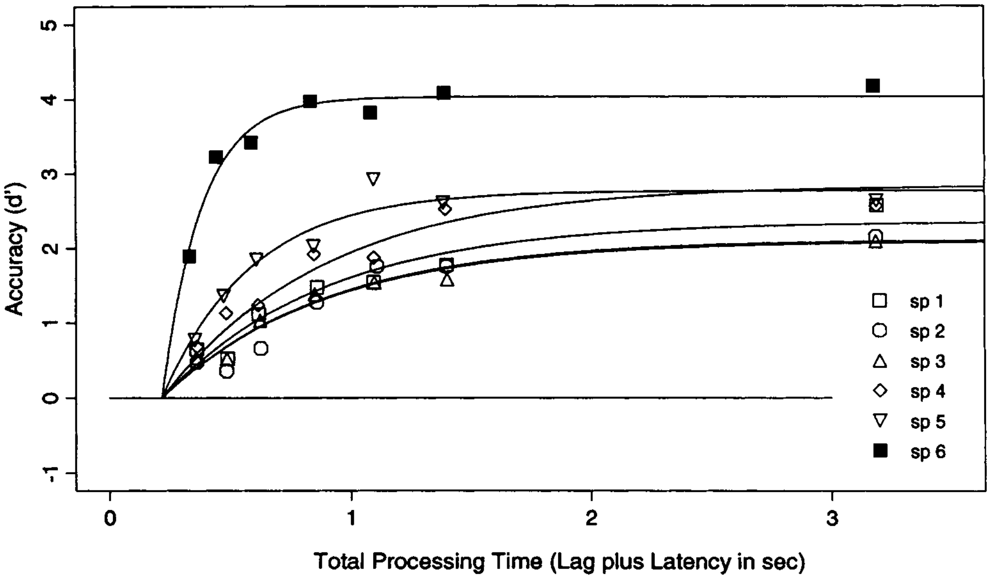

Figure 10. Average (over subjects) *d*′ accuracy in the two-alternative forced-choice item recognition task (see Experiment 2) as a function of processing time (i.e., lag of the interruption cue plus latency of response to the cue). (Symbols show observed [asymmetric] *d*′ accuracy for each study position [sp] of the old item in the test pair. Solid lines show the best fitting exponential retrieval functions [see Equation 3] by using the parameter values for the average data listed in Table 6.)

d' units. Accuracy significantly varied with study position, $F(5, 15) = 3.24$, $p < .05$, yielding a typical bowed serial position function. As with the results reported in McElree and Dosher (1989), asymptotic accuracy levels varied directly with recency coupled with a small primacy effect for Serial Position 1. McElree and Dosher (1989) found that asymptotic serial position functions across memory set sizes of three to six words were aptly described by a simple strength-forgetting model originally proposed by Wickelgren and Norman (1966) for untimed recognition data. The data in Figure 9 are consistent with earlier data for yes–no recognition. However, because these data are derived from only one set size, they provide few internal constraints to test specific forgetting models.

Retrieval dynamics. Figure 10 presents average (across subjects) full SAT functions for each study position.

Competitive fits systematically varied the three parameters of the exponential retrieval equation, Equation 3. The best fitting functions for the d' data, both in terms of the consistency of parameter estimates across subjects and high R^2 values (see Equation 1), allocated a separate asymptote (λ) for each serial position, a separate rate parameter (β) for Serial Position 6, Serial Position 5, and for the set of Positions 1–4, and one intercept (δ) (a 6λ-3β-1δ fit). Table 6 lists the parameter estimates and resulting R^2 for the average and individual subject's data. The smooth lines in Figure 10 present this best fitting exponential function for the average d' data.

Consistent with the observed data, the estimated asymptotes (λs) graded directly with recency of study, with a small primacy advantage for the first item on the list. We observe here a large rate (β) advantage for judgments involving the most recent item, Serial Position 6, estimated at 6.02 for the average data. The rate parameter for Serial Position 5 was estimated to be substantially slower than the most recent, 2.70 in the average data, but still consistently faster than the earlier serial positions, 1.46.

The observed rate advantage for Serial Position 6 replicates the fast-matching pattern seen in the yes–no item recognition studies of Wickelgren et al. (1980) and McElree and Dosher (1989) and a yes–no paired-associate recognition task of Dosher (1981). The data here depart from the pattern reported in other studies, in that there is an apparent rate advantage for the next-to-most recent position, Serial Position 5, when compared with other less recent positions. Its magnitude is much smaller than what is observed for Serial Position 6 but nevertheless appears consistent across subjects (see Table 6). We do not attribute this pattern to an extended recency advantage in dynamics for items from List Position 5 but rather to suppression of dynamics for Positions 1–4 because of presentation order effects that are documented in McElree and Dosher (1991).

Table 6
Exponential Fits of Two-Alternative Forced-Choice Item Recognition Data

	Subject				
Variable	AV	BM	EC	HS	LB
Asymptote (λ)					
Serial Position 1	2.35	2.08	3.42	2.12	1.98
Serial Position 2	2.12	1.58	3.41	1.78	1.91
Serial Position 3	2.09	1.45	3.14	2.19	1.80
Serial Position 4	2.83	1.17	4.50	2.61	3.15
Serial Position 5	2.77	1.33	3.98	3.64	2.81
Serial Position 6	4.03	4.50	4.50	4.50	4.40
SP 1–4 rate (β)	1.47	1.27	0.90	1.86	1.91
SP 5 rate (β)	2.70	3.25	0.94	3.36	4.00
SP 6 rate (β)	6.02	5.10	1.14	10.0	3.91
Intercept (δ)	.216	.146	.129	.251	.199
R^2	.967	.951	.873	.920	.894

Note. SP = serial position.

Discussion

Our purpose in examining SAT functions for 2AFC item recognition was to determine whether the forced-choice task yields a pattern of results different from what has been observed in prior yes–no item recognition tasks. Our primary concern was that the force-choice procedure itself may induce a serial mode of processing. However, the results of the experiment were generally consistent with prior yes–no item recognition studies. There is no suggestion from this study that the 2AFC procedure itself is responsible for the large dynamics (in particular, intercept) differences observed in the JOR task.

General Discussion

Retrieval of Item and Order Information

Summary of Results

In both the JOR and item recognition tasks, asymptotic performance reliably decreased as the relevant test item was drawn from less recently studied positions, except for a one-item primacy effect, which was somewhat more pronounced in the item recognition task. Recency is the primary determinant of the strength or availability of the STM representations in both JOR and item judgments.

Recency affects retrieval dynamics only when processing order information. In the JOR task, recency induced very large shifts in SAT intercept and rate. In contrast, in the item recognition tasks, as reported by McElree and Dosher (1989), Wickelgren et al. (1980), and in the current Experiment 2, recency has little effect on dynamics.

In the JOR task, the substantial dependence of SAT dynamics on recency reflects a serial retrieval mechanism. SAT intercepts for probes composed of less recent items were delayed as much as 500 ms relative to those composed of recent items. Delayed SAT intercepts are inconsistent with retrieval mechanisms in which probe items are compared in parallel with all elements in STM. The analysis of the RT distributions also supports—but by itself does not force—this conclusion. Models in which serial comparison processes are ordered by the recency of test probe(s) predict that recency should affect the leading edge of the RT distributions. This pattern was clearly demonstrated for order judgments in Experiment 1b, as it was in the prior analyses of

Muter (1979), Hacker (1980), and Hockley (1984).

In contrast to the JOR results, the observed dynamics for item recognition rule out serial retrieval mechanisms. Serial scanning models for item recognition, such as the serial self-terminating stack model of Theois (1973) and the conveyor-belt model of Murdock (1974), predict rather strong effects of study position on retrieval dynamics (McElree & Dosher, 1989). In prior yes–no tasks (McElree & Dosher, 1989; Wickelgren et al., 1980) and in the 2AFC task of Experiment 2, the dynamics for item recognition implicate a parallel or direct-access retrieval mechanism. Common dynamics (i.e., equal SAT intercept and rate) were observed for all study positions except for the most recent position, a case of immediate repetition, which exhibited very rapid dynamics. These SAT item recognition data are consistent with a simple parallel diffusion random walk model (Ratcliff, 1978) or a direct-access strength accumulator (Reed, 1976; see, McElree & Dosher, 1989).

Alternative Routes for Retrieving Order Information

Is serial processing the only means of retrieving order information? Most models of JOR prior to the work of Muter (1979, 1980) and Hacker (1980) assumed that order was assessed by a direct comparison of strength or feature counts of both test items. Muter (1979, 1980) and Hacker (1980) rejected these comparative processing accounts of relative JOR because RT did not vary with the difference in recency between the two test items. However, in Experiment 1b, evidence for comparative processing of the probes was found for early SAT interruption points less than 1 s. The d' early in retrieval was higher for pairs that differed most in study position. This evidence weakens the conclusion of Muter and Hacker: Under time pressure, a direct comparison of item information (e.g., strength, attributes, or trace fragility) may provide an alternate and relatively fast means of assessing recency. Because access to item information is direct or in parallel, the time course is sufficiently rapid that item information is almost always available prior to serial-derived order information.

If, in fact, a direct comparison of item information provides an alternative and potentially faster route to assessing order, one may question the usefulness of a slower serial process. Consider, however, that order does not always directly covary with strength or feature counts. A direct assessment of item information in many cases will lead to nonveridical performance and a serial retrieval process may be the only means of recovering correct order information. Even in cases in which strength does covary with recency, a direct comparison process may lead to lower accuracy than a slower serial process. Our argument is based on extrapolating performance from the results of the 2AFC item recognition task and then comparing these to the observed JOR performance of Experiment 1b.

JOR performance is far superior to what is predicted if order judgments were based on item strength exclusively. Figure 11 directly compares observed JOR performance and JOR performance predicted from strength comparisons alone. Observed performance is plotted in open symbols and the upper curves in each panel, whereas JOR performance predicted from item strength alone is shown in lower curves. The strength-based predictions were calculated in the following way. Evidence concerning the strength or feature counts of the test items is retrieved, and subjects choose as the most recent the item that has the largest strength value. In the 2AFC item recognition study of Experiment 2, old items (A, B, . . . , F) were paired at test with a new, nonlist item (X). The d' values in Figure 10 provide an estimate of the normalized distance of the old from the new distribution ($\mu_{(AX)}$, $\mu_{(BX)}$, . . . , $\mu_{(FX)}$). Assuming equal variance Gaussian distributions, we can estimate the difference in strength be-

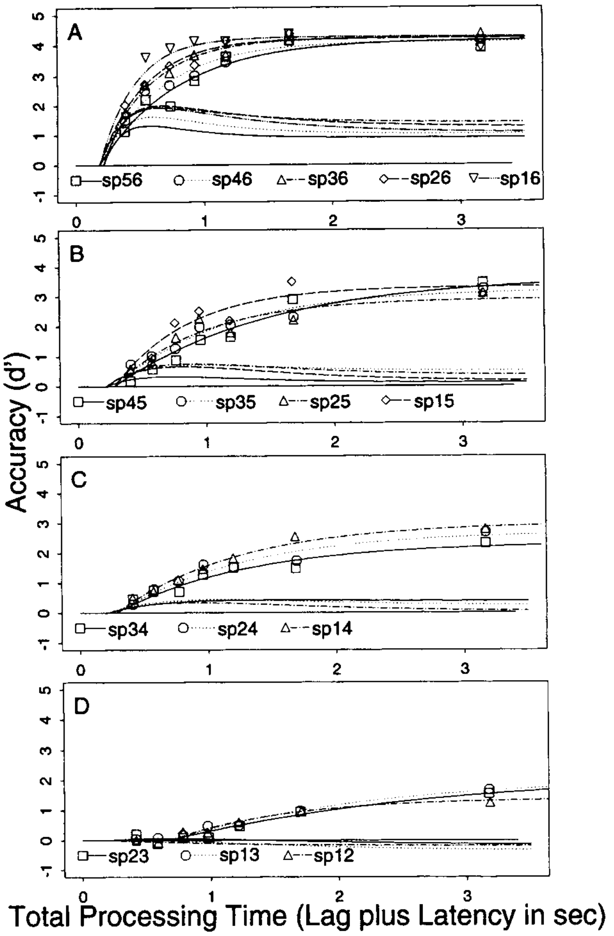

Figure 11. A comparison of observed judgment of recency (JOR) performance with predicted performance based on a direct-difference item-based strength metric (see text). (Open symbols show average [across subjects] observed d' accuracy as a function of processing time [i.e., lag plus latency] from the JOR task of Experiment 1b [replotted from Figure 7]. The monotonic functions fitted to the symbols show the best fitting exponential retrieval functions [see Equation 3] by using the parameter values for the average data listed in Table 3 [replotted from Figure 7]. The JOR functions predicted from a direct assessment of item strength are shown in the lower nonmonotonic curves in each panel. These functions are plotted in line types that match the corresponding fits of the observed data. See text for the method of computing predicted item-based performance. sp = study position.)

tween any pairing of old items with a subtractive method, that is, $d'_{AB} \approx [z(\mathrm{AX}) - z(BX)] \div 2^{1/2}$. This difference estimates the accuracy of performance if a response were based on a comparative assessment of item information alone. The full time course functions for the estimates of JOR performance based on item information were derived by using Equation 2 with the exponential parameters fitted to the item recognition data of Experiment 2 (see Table 6).

Two features of Figure 11 deserve attention. First, across all test pairs, estimated performance based on item information yields vastly poorer asymptotic performance than does the observed data. In fact, as a direct consequence of the primacy effect in item recognition, predicted d' is negative for test probes pairing the first item on the list with Study Positions 2 and 3. Second, item-based performance, although yielding lower asymptotic performance, shows a small advantage early in retrieval. This is a consequence of the faster retrieval dynamics for item as compared with order information. Of course, this cross-experiment comparison only approximates the possible contributions of an item-based process as compared with a serial retrieval process in a JOR task. The encoding and representation of items in the two situations may differ substantially. However, if subjects maximized the encoding and representation of item information in the item recognition task, then the derived functions in Figure 11 provide a reasonable estimate of item-based performance.[9] We conclude that a judgment based strictly on item information is a rather poor indicator of recency, even in a situation in which item strength for the most part directly covaries with recency.

The asymptotic performance in the JOR task reflects the retrieval of order or position information, not just item strength. However, the time course for the retrieval of this more specific information is slower than item-based judgments because it is based on a serial rather than parallel comparison process. Our SAT functions primarily reflect the slower, serial-derived order information, with some smaller contributions of strength-based assessments early in retrieval. In general, however, subjects might opt to forego a slow serial retrieval process in favor of a fast direct comparison process in circumstances in which speed is at a higher premium than accuracy.

Implications for General Memory Models

The general conclusion of this study is that two distinct retrieval mechanisms are used to retrieve item and order information. The strong effect of recency on retrieval dynamics in the JOR task—especially on SAT intercept—indicate that order information is retrieved by a serial mechanism. The absence of an effect of recency on retrieval dynamics in item recognition indicates that item information is retrieved by a parallel or direct-access mechanism. The latter is consistent with a number of general memory models that assume that item recognition is mediated by computing a global strength statistic (e.g., Gillund & Shiffrin, 1984; Hintzman, 1984, 1988; Murdock, 1982, 1983). In order retrieval, we have demonstrated that the large impact of recency on the SAT dynamics can be accounted for by a modified version of the serial self-terminating scan model of Hacker (1980). However, similar serial accounts may be given for these data in the context of more general, explicit memory models. In models such as TODAM (Lewandowsky & Murdock, 1989; Murdock, 1982, 1983) order information, unlike item information, is not directly retrievable from the memory trace. Rather, as developed by Lewandowsky and Murdock, order information is a derived property that is recovered at retrieval by a serial-chaining operation that capitalizes on pairwise associative information encoded at study.

Unfortunately, general memory models such as TODAM, along with other competing models such as SAM and MINERVA2, have been developed primarily on the basis of untimed memory performance. Quantitative application of these models to time-course data requires substantial development of the temporal properties of the assumed retrieval processes. Nevertheless, we note that in principle models such as TODAM appear to provide an organic or intrinsic explanation for the observed differences in retrieval dynamics for item and order information. That is, differences in retrieval stem from constraints on memory storage that can be motivated on independent grounds (Dosher & Rosedale, 1989; Lewandowsky & Murdock, 1989; Murdock, 1982, 1983). Further theoretical and empirical work is clearly needed to test the adequacies of these more general approaches.

---

[9] Rates of presentation were faster in the 2AFC item recognition task than in the JOR task. This may result in slight underestimates of item-based performance.

References

Bower, G. H. (1972). Stimulus sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* New York: V. H. Winston.

Corbett, A. (1977). Retrieval dynamics for rote and visual image mnemonics. *Journal of Verbal Learning and 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. (1984). 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., & McElree, B. (1992). Memory search: Retrieval processes in short-term and long-term recognition. In L. R. Squire (Ed.), *Encyclopedia of learning and memory.* New York: Macmillan.

Dosher, B. A., McElree, B., Hood, R. M., & Rosedale, G. R. (1989). Retrieval dynamics of priming in recognition memory: Bias and discrimination analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 868–886.

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

Estes, W. K. (1985). Memory for temporal information. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 149–168). Berlin: Springer-Verlag.

Flexser, A. J., & Bower, G. H. (1974). How frequency affects recency judgments: A model of recency discriminations. *Journal of Experimental Psychology, 103,* 706–716.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection and psychophysics.* New York: Wiley.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6,* 651–675.

Hinrichs, J. V. (1970). A two-process memory-strength theory for judgments of recency. *Psychological Review, 77,* 223–233.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

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.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69,* 382–386.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96,* 25–57.

Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376–404). San Francisco: Freeman.

Lockhard, R. S. (1969). Recency discriminations predicted from absolute lag judgments. *Perception & Psychophysics, 6,* 42–44.

Luce, R. D. (1986a). Response time distribution in memory search: A caution. In F. Klix & H. Hagendorf (Eds.), *Human memory & cognitive capabilities: Mechanisms & performances: Part A.* Amsterdam: North-Holland.

Luce, R. D. (1986b). *Response time: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General, 118,* 346–373.

McElree, B., & Dosher, B. A. (1991). *Serial comparison processes in judgments of order in STM.* (Tech. Rep. No. MBS 91-33). Irvine, CA: Irvine Research Unit in Mathematical Behavioral Sciences, University of California at Irvine.

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

Morton, J. (1968). Repeated items and decay in memory. *Psychonomic Science, 10,* 219–220.

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

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

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. (1983). A distributed memory model for serial order information. *Psychological Review, 90,* 316–338.

Muter, P. A. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 5,* 160–169.

Muter, P. A. (1980). *Response latencies in discriminations of recency.* Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

Peterson, L. R. (1967). Search and judgment in memory. In B. Kleinnmuntz (Ed.), *Concepts and the structure of memory.* New York: Wiley.

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

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88,* 212–225.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21,* 139–155.

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 and Cognition, 4,* 16–30.

Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 201–211.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

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. (1973, November). *Evidence against self-terminating memory search from properties of the RT distribution.* Paper presented at the meeting of the Psychonomic Society, St. Louis, MO.

Sternberg, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32.

Theios, 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. 43–85). San Diego, CA: Academic Press.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, England: 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. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9,* 418–455.

Wickelgren, W. (1974). Single-trace fragility theory of memory dynamics. *Memory and Cognition, 4,* 775–780.

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.

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior, 2,* 65–74.

Appendix

RT and SAT Predictions From Hacker's (1980) Serial Self-Terminating Model

RT Predictions

The Hacker model assumes that recency is assessed by a backward or recency-based, serial self-terminating scan. Errors are attributed to the loss of items in the scanned memory representation. For tests of items *i* and *j*, accuracy is modeled by Equation A1:

$$P_{ij} = a_i + .5(1 - a_i)(1 - a_j). \qquad \text{(A1)}$$

The probability of a correct JOR (P_{ij}) is determined by a memory availability parameter a_i (where i denotes the later probe in a test probe consisting of items i and j), incremented by a guessing factor for cases in which both items are unavailable. The P_{ij} can be converted to d' units by assuming equal-variance Gaussian distributions and the standard decision rule for 2AFC, $d' = 2^{½}[z(P_{ij})]$.

Correct RT is modeled by Equation A2:

$$RT_{ij} = b + \frac{1}{P_{ij}}\left\{a_i\left[\left(1 + \sum_{k=1}^{i-1} a_k\right)s\right] + .5(1 - a_i)(1 - a_j)\left[\left(\sum_{\substack{k \neq i,j \\ k=1}}^{n} a_k\right)s + g\right]\right\}. \qquad \text{(A2)}$$

The first parameter, *b,* estimates a base time for processes constant across all comparisons, such as encoding of the test probes, response execution, and so on. A second processing parameter, *s,* gives the average expected serial search time for each item in the scan. Search time is weighted by the availability parameter for the later probe, a_i, and the expected number of items that must be searched prior to reaching this item. The remaining component estimates the expected duration of cases in which the search fails to find a match yet ends in a correct guess. The parameter *g* estimates the duration of the guessing process. The matches and guesses are normalized by the overall probability of a correct response for the particular pairwise comparison *ij.*

Errors occur only when the the later item is unavailable to the scan operation. If the earlier item is also unavailable, the subject guesses randomly. If the earlier item is available, then error RT is controlled by its position in the scan. The formula for incorrect RT is given in Equation A3:

$$RT_{ij} = b + \frac{1}{(1 - P_{ij})}\left\{(1 - a_i)a_j\left[\left(1 + \sum_{\substack{k \neq i \\ k=1}}^{j-1} a_k\right)s\right] + .5(1 - a_i)(1 - a_j)\left[\left(\sum_{\substack{k \neq i,j \\ k=1}}^{n} a_k\right)s + g\right]\right\}. \qquad \text{(A3)}$$

In fitting the equations, accuracy and RT data are not jointly used to estimate availability parameters. Rather, the availability parameters are first estimated from the accuracy data. These parameters are then substituted into Equation A2 and the three processing parameters, *b, s, g,* are estimated.

SAT Predictions

SAT asymptotic accuracy is modeled by Equation A1, which is converted to d' units. The predicted growth of accuracy as a function of interruption time was computed with the assumption that the time to compare the two test items with an item in the memory set was exponentially distributed with rate *S,* corresponding to the *s* in the RT model. If all items in the study set were available in the memory set, the probability correct as a function of retrieval time, *t,* is simply gamma distributed, with an order, α, equal to the recency of the later probe in the test pair and offset by a base time, B:

$$P(T \leq t) = \frac{S^{\alpha}}{(\alpha - 1)!} \int_{0}^{t-B} e^{-St'} t'^{\alpha-1}\, dt', \qquad t > B \quad \text{else} \quad 0. \tag{A4}$$

Equation A4 is for a cumulative gamma distribution of order α. The gamma distribution represents the convolution of α independently and identically distributed exponential variables, each of which represents the completion time for a single comparison. The exponential (and hence the gamma) is not the only distribution for single stage completion times that might be assumed, but it is a frequently used distribution that has the advantage of tractability (e.g., see Townsend & Ashby, 1983).

Because the model assumes that each item has a specific probability of being available (a_i) in any particular scan, the overall expression that is needed to compute SAT functions will not reflect a pure gamma but rather a probabilistic mixture of gammas with different values of α (i.e., different numbers of exponential comparison processes). The gammas of particular orders, determined by the expected number of items in the search or memory set, must be weighted by the respective availability parameters and summed to compute the proportion correct as a function of retrieval time. Likewise, the proportion of incorrect responses is computed by weighting gammas of particular order for all possible states in which the earlier probe is available (with probability $p = a_j$) and the later probe is unavailable (with probability $p = 1 - a_i$). Finally, proportion correct must be incremented by guessing for cases in which (a) the interruption cue occurred before the scan was complete and (b) the scan was complete but both the later and earlier test items were unavailable. Guessing accuracy was 50%. The predicted proportion correct as a function of retrieval time was converted to d'.

Received December 17, 1991
Revision received December 9, 1992
Accepted December 14, 1992 ■