

Accessing short-term memory with semantic and phonological information: A time-course analysis

BRIAN McELREE

University of California, Irvine, California

The time course for accessing short-term memory representations on the basis of item, phonological, and semantic information was measured with a speed-accuracy tradeoff (SAT) variant of a recognition probe task. Subjects studied 5-word lists and then judged whether a test word (1) was in the study list (item judgment); (2) rhymed with a word in the study list (rhyme judgment); or (3) was a synonym of a word in the study list (synonym judgment). Analysis of the SAT retrieval functions demonstrated that judgments based on phonological and semantic information were slower and less accurate than judgments based on composite item representations. The slower retrieval dynamics for rhyme and synonym judgments suggest that judgments based on component properties of the memory trace are mediated by either a generate or recall process distinct from the parallel or direct-access mechanism used in item judgments.

Recognition of events stored in long-term memory is typically viewed as exhibiting the property of *content addressability*, meaning that a recognition probe can be matched directly or in parallel to items in the memory store (Doshier & McElree, 1992). The type of retrieval mechanism that underlies the recognition of events assumed to be held in short-term memory (STM) has been more controversial, with proposals ranging from various direct-access or parallel mechanisms (e.g., McElree & Doshier, 1989; Ratcliff, 1978; Reed, 1976) to specialized serial scanning mechanisms (e.g., Sternberg, 1975). Recently, speed-accuracy analyses (outlined below) have suggested that STM item recognition is best viewed as a parallel or direct-access procedure, on a par with the type of mechanism used to recognize events stored in long-term memory. The current study extends this approach to cases where the STM representation must be accessed on the basis of component properties of the memory representation, namely phonological and semantic properties.

STM Item Recognition

The retrieval of item information from STM representations has been a long-standing issue in memory research since the appearance of Sternberg's (1966, 1969) classic reaction time (RT) studies of the recognition probe task. On the basis of the finding that mean RT was a linear function of the number of items studied, with approximately equal slopes for positive and negative responses, Sternberg (1966, 1969, 1975) argued that items were retrieved from STM by a serial exhaustive search or scan mechanism.

However, a number of alternative retrieval mechanisms have since been proposed to account for this pattern of mean RT, including serial self-terminating search mechanisms (e.g., Theios, 1973; Theios, Smith, Haviland, Traupman, & Moy, 1973), rate-varying parallel-access mechanisms (e.g., Murdock, 1971; Townsend & Ashby, 1983), a diffusion or random-walk mechanism (Ratcliff, 1978), and various direct-access mechanisms (e.g., Anderson, 1973; Baddeley & Ecob, 1973). Discriminating between alternative retrieval models has proved difficult in part because RT—in particular, mean RT—has been the primary object of interest. With an appropriate set of assumptions, radically different (e.g., serial and parallel) mechanisms have been shown to yield equivalent RT predictions (see Townsend & Ashby, 1983).

McElree and Doshier (1989) argued that a number of alternative retrieval mechanisms that lead to nearly equivalent RT predictions nevertheless give rise to divergent predictions concerning the full time course of retrieval. The cued-response SAT procedure is one means of measuring the full time course of retrieval. In this procedure, subjects are cued to respond (typically with a tone) at various points following the presentation of a recognition probe, which allows retrieval accuracy to be measured as a function of retrieval time. Full time-course retrieval functions generated by this procedure provide a measure of when accuracy departs from an initial chance level, the rate at which accuracy grows over retrieval time, and the ultimate or asymptotic level of performance. The asymptote is a measure of memory strength or the overall probability of recognition. The point at which accuracy departs from chance and the rate at which accuracy grows are joint measures of the speed or dynamics of retrieval. McElree and Doshier found that many extant retrieval models yield distinct predictions concerning the dynamics of retrieval as a function of (1) the size of the memory set and/or (2) the serial position or recency of the (positive) test probe. For example, Stern-

This research was supported by a University of California Faculty Research Grant (92/93-90-FRG). The author thanks B. A. Doshier, D. Hintzman, R. Proctor, and an anonymous reviewer for helpful comments and suggestions. Correspondence should be addressed to B. McElree, Experimental Psychology, 8th Floor, 6 Washington Pl., New York University, New York, NY 10003 (e-mail: bdm@xp.psych.nyu.edu).

berg's (1966, 1969, 1975) serial exhaustive scan model predicts that retrieval dynamics will slow (longer intercepts and/or slower growth rates) as the size of the memory set is increased, since the memory set must be exhaustively scanned. Serial self-terminating and rate-varying parallel models predict that retrieval dynamics will vary with the serial position or recency of the test probe, since recency is assumed to affect the order of the serial comparisons in the serial self-terminating model (Theios, 1973) or the rate of information accrual in rate-varying parallel models (Murdock, 1971; Townsend & Ashby, 1983).

McElree and Doshier (1989), however, found no evidence to support the claim that retrieval (speed) dynamics are affected by the size of the memory set. Rather, they found that the size of the memory set affected asymptotic accuracy only, with larger set sizes yielding lower overall asymptotic levels. The analysis of the retrieval functions for individual serial positions within lists of different sizes indicated that the impact of set size on asymptotic performance was primarily a consequence of the inclusion of less recent serial positions in the larger (averaged across serial position) set size functions. Asymptotic performance varied with the recency of the (positive) test probe, with less recent test probes resulting in lower asymptotic performance, consistent with forgetting with the passage of time and/or intervening study events. With the exception of the last item on the study list—a case of immediate repetition between study and test (see below)—the dynamics of retrieval for different serial positions within and across various set sizes were indistinguishable. These data are inconsistent with serial self-terminating models, in which recency determines the order of the serial comparisons (Theios, 1973), and with parallel retrieval models, in which the rate of information accrual depends on recency (Murdock, 1971; Townsend & Ashby, 1983). Uniform dynamics for different serial positions, as well as for different set sizes, are consistent with a direct-access mechanism, such as a simple strength accumulator (Reed, 1973) or a parallel retrieval mechanism with self-terminating decision rules, such as a simple random-walk or diffusion process (Ratcliff, 1978; McElree & Doshier, 1989). These retrieval mechanisms are in turn generally consistent with a large class of memory models that treat item recognition as an assessment of the global familiarity, strength, or goodness-of-match of the memory probe (Gillund & Shiffrin, 1984; Hintzman, 1984, 1988; Murdock, 1982, 1983).

Accessing Semantic and Phonological Information

The question addressed in the present experiment was whether component properties of the memory trace for a studied item can be similarly accessed with a direct or parallel retrieval mechanism. The mechanism by which phonological and semantic properties of a test probe provide access to STM representations was examined. Access to phonological properties of the memory trace was analyzed by examining judgments of whether a test probe rhymed with an element in the study list. Access to semantic properties of the memory trace was analyzed by examining judgments of whether a test probe was a synonym of an el-

ement in the study list. In each case, the SAT procedure was used to generate full time-course retrieval functions, and these functions were compared to the full time-course retrieval functions for standard item recognition.

Many models of memory, including models of short-term memory, view the memory trace as a bundle of features, a subset of which include semantic and phonological features (e.g., Bower, 1967; Drewnowski, 1980; Eich, 1982; Hintzman, 1988; Murdock, 1982, 1993; Nairne, 1988, 1990; Nosofsky, 1986; Underwood, 1969). Given such an assumption, one perspective might be to view episodic judgments based on phonological and semantic properties of the representation as differing from standard item recognition only in terms of the number of cues that are available to access the memory trace. When fewer cues are available, judgments based on component properties will yield a lower goodness of match between the test probe and the memory trace than judgments based on composite item information.

However, there are at least two reasons to question the assumption that the same retrieval mechanism is used for both types of judgments. One reason concerns the nature of the underlying memory representation. A composite representation, although consisting of a bundle of features, may not enable direct access to its component elements or features. Although no explicit proposal of this sort has been made for cases of component features of an item, Murdock (1993) has argued that in TODAM, a distributed memory model, composite representations formed by chunking or grouping items during study result in "opaque containers" (Johnson, 1972, p. 125) in which the component parts are not directly accessible to the retrieval processes that provide access to the composite representation. (See also Doshier & Rosedale, 1989, for evidence from a priming paradigm that highlights differences between composite and component representations.) Additional retrieval operations may be needed to effectively unpack the composite representation into its component parts.

Another reason why different retrieval processes may be used for component recognition might lie in the type of evidence required to reliably determine whether a component property occurred in the study event. Even in the event that component properties provide parallel or direct access to STM traces, the degree of match between a component property of the test probe and an item's memory representation may not be sufficient to reliably discriminate targets from lures. Whether a component property provides enough evidence to reliably discriminate targets from lures clearly depends on the distinctiveness of the property in relation to potential (nonlist) lures. (In this respect, rhyme judgments may have a notable disadvantage, since phonological properties form a small set of features commonly shared among list members and lures.) Component properties that do not provide enough evidence for recognition may induce an alternative retrieval process.

What alternative retrieval processes could underlie component judgments? One effective strategy might be to use the component properties of the test probe to explicitly generate a candidate rhyme or synonym or a set of candidate

rhymes or synonyms (Becker, 1979), and then directly query the memory representation with the generated item(s). Here the assumption is that word recognition operations applied to the test probe will make available its relevant component (phonological and semantic) properties. The generation process uses these properties to retrieve candidate items from the mental lexicon. Once retrieved, the candidate items can be compared to the episodic list representation, via a parallel or direct-access operation, in order to effect an old–new discrimination.¹

A closely related strategy would be to use the relevant information contained in the test probe as a cue for an explicit recall operation. As above, the relevant component properties of the test probe can be extracted by word recognition operations. In this case, however, the extracted phonological or semantic information may be used as a cue for an episodic recall operation. Indeed, there is some precedent in the memory literature for assuming that a recall operation may underlie what is nominally a recognition task. Gronlund and Ratcliff (1989; see also Gronlund, Edwards, & Ohrt, 1993) and Clark (1992) have argued that associative recognition is mediated by a recall operation performed on the components of the paired associates. Hintzman and Curran (1994) have argued that a recall operation may be used in difficult recognition discriminations (e.g., when the plural of a study word is used as a lure). While the generate and recall strategies differ in a number of aspects, they do not yield any distinctive predictions in the current work and so are referred to here as a generate/recall strategy (see Discussion section).

As an alternative to the generate/recall strategy, recognition based on component properties may induce a serial sequence of operations, tantamount to a serial scanning process, in which a single composite representation is retrieved from memory, unpacked, and then compared to the test probe on the basis of the relevant component property. McElree and Doshier (1993) documented a strategy of this sort for the retrieval of order information in a relative judgment of recency (JOR) task. In this task, subjects were presented two test items from a sequentially presented study list and required to judge which item had occurred more recently. A number of models of order information posit that relative judgments of recency are mediated by a comparison of the relative strength (or number of attributes, or fragility) of the memory trace for each test item (see McElree & Doshier). Using the SAT procedure, along with collateral RT (mean, distributional, and accuracy) data, McElree and Doshier argued that recency judgments are not primarily mediated by an assessment of memory strength but rather by a relatively slow serial self-terminating retrieval process in which the memory set is scanned in the order of recency, and a response is executed as soon as one of the test items matches an element in the memory set (Hacker, 1980; Muter, 1979). The critical data that supported this recency-based, serial self-terminating mechanism were the differences in SAT dynamics for probes of varying serial positions. The intercept, or earliest point where the SAT function departed from chance, increased as the most recent item in the test probe was drawn from

more remote serial positions. The differences in SAT intercept were large, varying by as much as 500 msec for test probes drawn from the most recent position as compared to test probes from the beginning of the memory set. Direct-access and parallel retrieval mechanisms are inconsistent with the large recency-based differences in SAT intercept.

Retrieving order information is not, of course, directly comparable to retrieving component properties of the memory trace, since the relevant information to be retrieved is relational in the former case and decompositional in the latter case. Nevertheless, the McElree and Doshier (1993) study highlights the fact that even though a direct-access mechanism might provide information relevant to the judgment, the demands of some tasks appear to be more effectively met by a slower, serial retrieval process. Indeed, McElree and Doshier used strength estimates derived from a two-alternative forced-choice item recognition study to predict the upper limit on performance in a JOR task for the case that a difference in item strength was used as a metric for assessing the recency of occurrence. This simulation demonstrated that strength-based estimates of recency were available sooner than serially derived estimates of recency, since the former were based on a direct-access mechanism, but that strength-based performance vastly underestimated the observed performance derived from the slower serial comparison process. Additionally, the McElree and Doshier study serves to illustrate the signature pattern in SAT of a serial process: significant differences in the SAT intercepts for conditions differing in the number of implicit (serial) comparison operations.

Contrasting the Retrieval of Item, Phonological, and Semantic Information

Shulman (1970) examined mean RT and accuracy for judgments of whether a test probe was a homonym (rhyme), synonym, or identical (item) to an element in a 10-word list. With the exception of a small primacy effect, mean correct RT and error rates decreased with recency for each judgment. Overall, synonym judgments were slower than rhyme judgments (1,103 msec versus 846 msec, averaged across serial positions), which in turn were slower than item judgments (683 msec). The error rates were indistinguishable for rhyme and item judgments, but both judgments were more accurate than synonym judgments.

The Shulman (1970) study highlights the importance of serial position or recency in each type of judgment, as well as serving to document the large temporal differences among the judgments. The fact that recency had a similar effect on each judgment lends some support to the claim that each type of information is retrieved with the same sort of mechanism. The McElree and Doshier (1989) SAT data, as well as related SAT data (see, e.g., Wickelgren, Corbett, & Doshier, 1980, and Ratcliff, 1978), suggest that item recognition is mediated by a direct-access or parallel process and that recency effects in mean RT and accuracy are a function of underlying differences in the strength of the memory representations. (See McElree & Doshier, 1989, for an explicit strength-based forgetting model of recency

effects that jointly accounts for SAT and RT item recognition data.) The RT data for synonym and rhyme judgments reported by Shulman (1970) may also be consistent with a direct-access or parallel retrieval process: Longer mean RT may simply reflect differences in the accuracy of each type of judgment rather than differences in the manner in which the respective types of information are accessed. However, it is also possible that the longer RTs for component judgments are indicative of a different sort of retrieval operation, like the serial strategy outlined above and documented by McElree and Doshier (1993) for JOR or the more specialized generate/recall strategy. Without examining full time-course data, we have no principled means of discriminating between these alternatives.²

Serial position is the appropriate variable to test for various serial self-terminating mechanisms, as well as rate-varying parallel mechanisms, which may be consistent with the serial position profiles documented by Shulman (1970). (The size of the memory set is the variable that is critical for testing the applicability of a serial exhaustive scan mechanism. However, serial exhaustive scan mechanisms give no principled account of serial position effects; see McElree & Doshier, 1989, and Ratcliff, 1978. Consequently, this variable was not examined in the present experiment.) Serial self-terminating models account for recency effects by positing that the memory set is scanned in the order of recency (Theios, 1973). Rate-varying parallel models account for recency effects by assuming that more recent memory traces are retrieved more quickly (e.g., Murdock, 1971; Townsend & Ashby, 1983). Both models predict that SAT functions for more recent serial positions will be associated with shorter SAT intercepts and/or faster SAT rates of rise to asymptote.

The direct-access or parallel retrieval model suggested by McElree and Doshier (1989) contrasts with both serial self-terminating and rate-varying parallel models in predicting uniform SAT dynamics (intercept and rate) for different serial positions. Uniform dynamics for different serial positions are also predicted by the more specialized strategy in which candidate rhymes or synonyms are either generated or recalled. However, this model predicts that rhyme and synonym judgments will yield slower overall dynamics (SAT rate or intercept) than item judgments, since either an additional generation process precedes memory retrieval or a slower recall process is used.

Aside from examining the general impact of serial position on SAT retrieval functions for component and composite judgments, the retrieval functions for the last item on the list form a special focus of the present experiment. A number of SAT studies have documented one exception to the finding of uniform dynamics for different serial positions. This exception concerns the last serial position on the list, specifically, the most recently studied item. McElree and Doshier (1989) found that the retrieval dynamics for cases of immediate repetition in item recognition were 40%–50% faster than for study positions in which at least 1 item intervened between study and test. Wickelgren et al. (1980) measured retrieval functions for item recognition in super-span lists of 18 items and similarly reported

faster dynamics for the last item on the list and slower, uniform dynamics for every other serial position. Doshier (1982) also reported speeded dynamics for the last pair of words in a list of paired associates presented in an associative recognition task. In all these studies, the last item on a study list was a case in which there was an immediate repetition of an item from study to test. As originally suggested by Wickelgren et al. (1980), the fast dynamics for cases of immediate repetition appear to indicate that the test probe can be directly matched to the item(s) active in awareness, effectively circumventing the retrieval processes that are needed to restore an item in a passive state to active processing. Cowan (1993) has recently argued for a view of short-term memory that in general accords with these data, suggesting that short-term memory consists of recently activated long-term memory representation, a subset of which is the current focus of attention. The exact nature of the active representation that mediates this fast matching process remains an open question. McElree and Doshier (1989) noted, however, that the fast matching process must be mediated by a level of representation more abstract than a simple physical match (see Posner, Boies, Eichelman, & Taylor, 1969), since the matching advantage was not attenuated by (upper/lower) case variations between study and test. The question addressed here is whether a similar pattern will be observed for rhyme and synonym judgments. Such a finding would reinforce the notion that this dynamics advantage is a direct consequence of what is currently being processed or active in awareness, as opposed to a relatively low-level physical match.

METHOD

Subjects

Four undergraduates and 1 graduate student from the University of California, Irvine, served as subjects. All subjects were native speakers of English and had normal or corrected-to-normal vision. Each subject participated in 10 one-hour sessions, plus an additional 1-hour practice session that served as training for the SAT procedure. Three of the subjects were paid for their participation, and 2 subjects (B.C. and S.H., who were affiliated with the laboratory) volunteered their time.

Materials

One hundred word triplets of 1–4 syllables were used as stimuli. The triplets were composed of 1 study item (e.g., *tale*) that also served as a positive probe for item judgments, a rhyme of the study item (e.g., *bail*) that served as a positive probe for rhyme judgments, and a synonym of the study item (e.g., *story*) that served as a positive probe for the synonym judgments. The full set of 100 triplets is listed in Appendix A.

Design and Procedure

Stimulus presentation, response collection, and feedback were controlled by an IBM-compatible computer. Each trial consisted of the presentation of a 5-word study list, randomly drawn from the set of 100 words. Following the presentation of the study set, a test probe was presented and the subject was required to make an item, rhyme, or synonym judgment. To ensure that the encoding of the study list was the same across the three judgments, a high, medium, or low tone was presented after the study list to indicate to subjects that a rhyme, synonym, or item judgment, respectively, was required.

The sequence and timing of events that constituted a trial are schematically presented in Figure 1 and were as follows: (1) A centered, square fixation point was presented for 500 msec. (2) Each study word was sequentially presented for 400 msec, followed by 14 msec of blank screen. (3) The final item in the study list was masked by a sequence of nonletter symbols (e.g., #%@\$&@#) presented for 500 msec. Coincident with the onset of the mask, a high (3100-Hz), medium (1500-Hz), or low (100-Hz) tone sounded for 50 msec to indicate a rhyme, synonym, or item judgment, respectively. (4) The test probe was presented in the center of the screen. (5) The test probe remained on the screen for .128, .357, .500, .786, 1.2, or 3 sec, at which point the screen cleared and a second 50-msec tone (same frequency as the first tone) sounded to cue subjects to respond. Subjects responded by pressing one of two keys on a numeric keypad to denote either an old (“3” key) or new (“1” key) judgment. (6) Following a response, visual feedback on the latency to respond to the interruption cue was presented. Subjects were instructed to respond to the cue within 270 msec. They were informed that responses longer than this were too long and that responses under 130 msec were anticipations. The subject initiated the next trial with a key press.

Each of the 10 sessions consisted of 600 trials, divided into 4 blocks of 150 trials. Subjects were free to rest for a few minutes between blocks. Within each block, there was an equal number of positive and negative item, rhyme, and synonym judgment trials. Posi-

tive probes were drawn from each of the five serial positions equally often. The type of judgment, list status of the probe, serial position of the (positive) probe, and the interruption lag were randomized within a session.

Data Analysis

For each subject, the proportion correct for each condition at each response lag was used to derive a standard *d'* measure (equal variance-Gaussian model). Perfect performance in any condition was adjusted by a minimum-error correction that replaced a zero error rate with an error rate of 0.5%. This correction ensures that, given the sample size, the resulting *d'* values are in fact measurable (Macmillan & Creelman, 1991). (The potential impact of the correction process on model fits is discussed in the Results section.) Additionally, trials on which the latency to respond to the interruption cue was greater than 500 msec were excluded from the analysis. These trials constituted less than 2% of the total number of trials.

Standard SAT functions can be quantitatively summarized by the exponential approach to a limit equation (see Doshier, 1976, 1979; Reed, 1973, 1976; Wickelgren, 1977):

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

In Equation 1, the parameter λ serves to describe the asymptotic level of performance. The δ and β parameters jointly estimate the dynamics of the function. The former describes the intercept or the

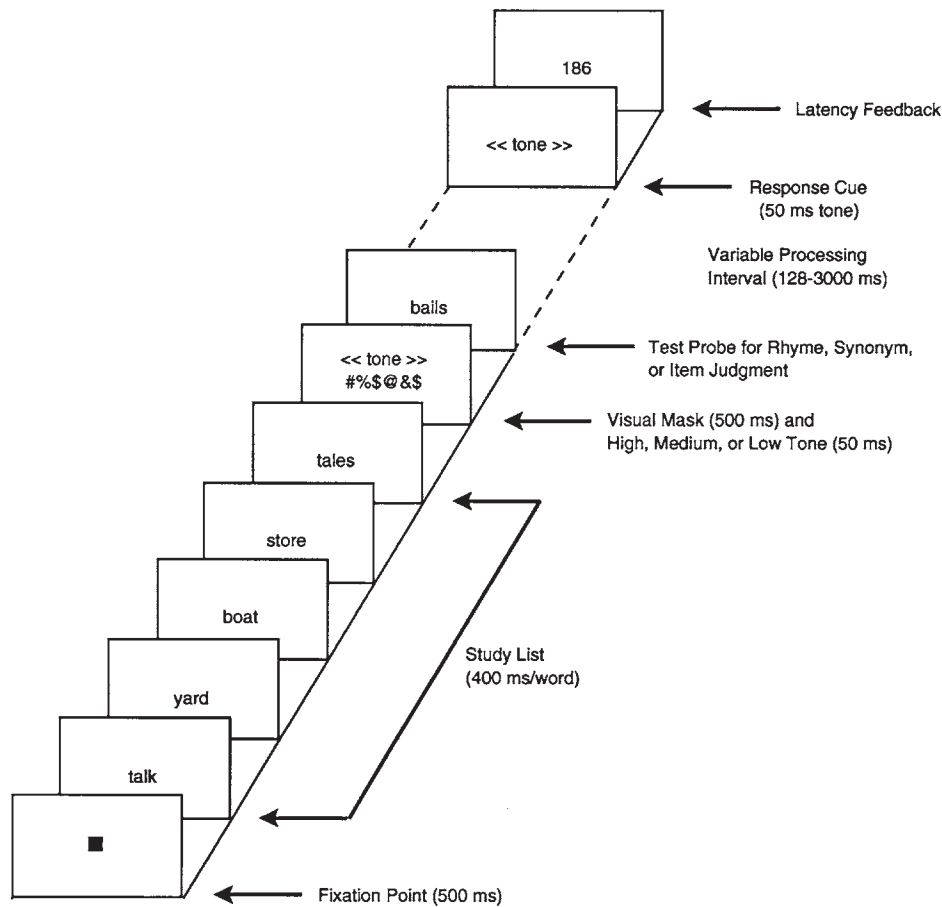


Figure 1. A sample trial sequence illustrating the cued-response speed-accuracy tradeoff variant of the item, rhyme, and synonym judgment tasks.

discrete point in time when performance begins to rise from chance. The β parameter indexes the speed or rate at which performance rises from chance to asymptote. Analysis of the SAT function in terms of Equation 1 allows an independent assessment of memory availability (λ) and the dynamics of memory retrieval (β and δ).³

An iterative hill-climbing algorithm (Reed, 1976) similar to STEPIT (Chandler, 1969) was used to fit Equation 1 to the empirical SAT functions. This algorithm minimized the squared deviations of predicted from observed data. Sets of competitive fits, varying the three parameters of Equation 1, were used to isolate differences between various SAT functions. The quality of the respective fits was assessed by three criteria, as follows: (1) the value of an R^2 statistic,

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

where d_i represents the observed data values, \hat{d}_i the predicted values, \bar{d} the mean, n the number of data points, and k the number of free parameters (Reed, 1973). This R^2 statistic is the proportion of variance accounted for by the fit, adjusted for the number of free parameters (k). (The same equation for *adjusted* r^2 is often used in multiple linear regression; see, e.g., Judd & McClelland, 1989.) (2) Evaluation of the consistency of parameter estimates across the 6 subjects. (3) Most importantly, examination of whether a fit yielded systematic (residual) deviations that could be accounted for with more parameters.

RESULTS

Appendix B lists the means and standard deviations across subjects of the lags plus latencies, hit rates, and false alarm rates for each judgment. Individual subject data are available from the author.

Asymptotic Accuracy

The longest interruption time (3 sec) gives an empirical measure of asymptotic recognition performance. Figure 2

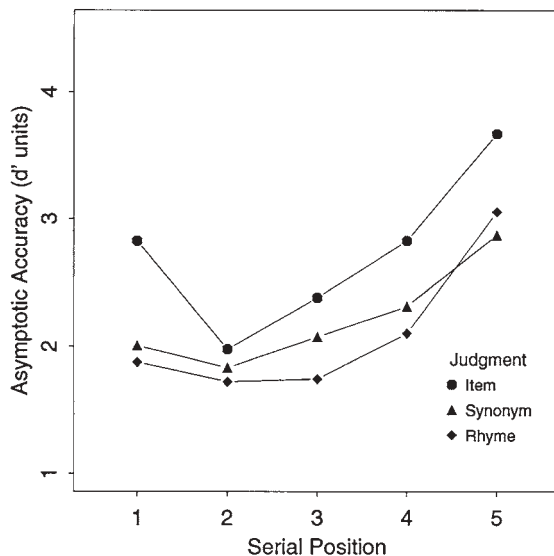


Figure 2. Observed average asymptotic d' values as a function of serial position for item, rhyme, and synonym judgments.

shows the average (over subjects) asymptotic d' for each judgment as a function of serial position. An analysis of variance (ANOVA) on the subjects' d' scores for the longest interruption time, with judgment (item, rhyme, and synonym) and serial position (1–5) as within-subject factors demonstrated that the type of judgment significantly affected asymptotic accuracy [$F(2,8) = 6.68$, $MS_e = 0.4295$, $p < .05$]. Item judgments were more accurate than either synonym or rhyme judgments, averaging (across serial position) 2.7 d' units as compared to 2.2 d' units for synonym judgments and 2.1 d' units for rhyme judgments. The small advantage for synonym judgments over rhyme judgments was nonsignificant in an ANOVA that directly contrasted the two judgments [$F(1,4) \approx 1$].

Serial position also had a strong and significant effect on asymptotic accuracy [$F(4,16) = 43.1$, $MS_e = 0.0936$, $p < .05$]. As inspection of Figure 2 shows, the impact of serial position was non-linear, producing a standard bowed function with a strong recency and somewhat smaller primacy component. The nonlinearity was evidenced by regression analyses in which a simple linear component produced (standard) r^2 values of .404 for the item judgments, .609 for rhyme judgments, and .752 for synonym judgments. Adding a quadratic component significantly increased the r^2 values to .932 for item judgments, .986 for rhyme judgments, and .987 for synonym judgments.

Overall, the pattern of SAT asymptotic accuracy accords well with the mean RT and RT-accuracy data reported by Shulman (1970). In both cases, serial position had a pronounced effect on all three judgments. Shulman (1970) reported that item judgments were faster and/or more accurate than either rhyme or synonym judgments, and this finding is paralleled here in the higher asymptotic levels for item judgments. Shulman (1970) also reported that rhyme judgments were faster and more accurate than synonym judgments. There was no evidence for a comparable difference here. This discrepancy may be in part because Shulman used homonyms as stimuli for the rhyme judgments, whereas simple (nonhomonym) rhymes were used here. The phonological structure of the former completely overlaps with the relevant study item, whereas nonhomonym rhymes differ by at least 1 syllable. This difference may have resulted in relatively lower performance for rhyme in the present experiment. It should also be noted that, in both studies, there are varying amounts of orthographic overlap between particular rhyme probes and the relevant study item. Thus it is possible that orthographic similarity partially contributes to rhyme judgments.

Full Retrieval Functions

Figure 3 presents the average (over subjects) SAT retrieval functions (d' as a function of processing time) for the item (panel A), rhyme (panel B), and synonym (panel C) judgments. There are two issues of primary concern. The first is the impact of serial position on the retrieval functions within each of the three types of judgments. The second is differences among the three judgments. The first issue was addressed by systematically varying the three parameters of Equation 1 to find the best fit of the 5 serial

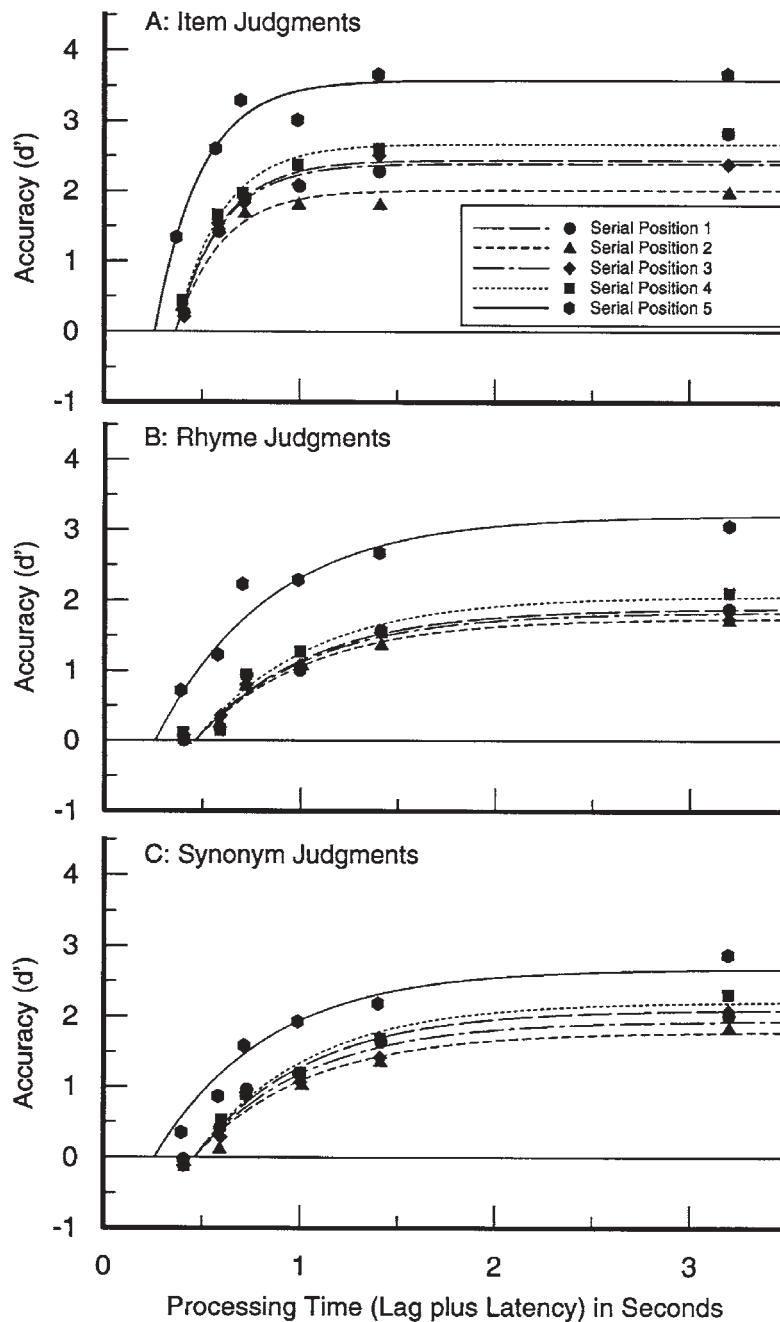


Figure 3. Average d' accuracy (symbols) as a function of processing time for item (A), rhyme (B), and synonym (C) judgments. Smooth curves in each panel show the best fits of Equation 1 with (the average) parameters listed in Table 1.

position functions within each judgment. After identifying the best fit for each judgment, all 15 conditions (3 judgments \times 5 serial positions) were fit simultaneously in order to isolate differences among the three judgments. In all cases, the fits were performed on individual subject data. Fits of the average (over subjects) data were used to summarize consistent patterns across subjects.

To quantify the impact of serial position on the retrieval functions for each judgment, sets of competitive fits were

performed that systematically varied the three parameters of Equation 1. These fits ranged from a null model, in which the 5 serial position functions were fit with a single λ (asymptote), β (rate), and δ (intercept) parameter, to a fully saturated 15-parameter model, in which each serial position was allotted a unique λ , β , and δ parameter. This analysis yielded a clear and consistent pattern across the three judgments. In each case, the best fit to the data was a $5\lambda-1\beta-2\delta$ model. With one exception (noted below), models

with fewer parameters than the $5\lambda-1\beta-2\delta$ model all yielded lower R^2 statistics and produced systematic (residual) misfits of the functions. More embellished models also produced lower R^2 values and did not yield any consistent parametric differences across subjects. The $5\lambda-1\beta-2\delta$ model allocated a separate λ parameter for each serial position (reflecting the strong serial position effects in Figure 2), a single β parameter for all serial positions, and one δ parameter for the last serial position (Position 5) and another δ parameter for every other serial position. Two δ parameters were needed to capture the exceptionally fast dynamics for the last item on the list—a case of immediate repetition between study and test. The intercept for judgments concerning the last item on the list was substantially shorter than for every other serial position, with an average advantage of 112 msec for item judgments, 238 msec for rhyme judgments, and 132 msec for synonym judgments.

The $5\lambda-1\beta-2\delta$ model produced the highest adjusted R^2 value (Equation 2) in fits of the average data for item (.959), rhyme (.964), and synonym (.967) judgments. More importantly, this model produced the most precise fits of the individual subjects' data. In the case of item judgments, the $5\lambda-1\beta-2\delta$ model yielded the highest R^2 values for 4 of the 5 subjects, with values ranging from .738 to .944. Subject B.G. had a marginally higher R^2 value for a simpler $5\lambda-1\beta-1\delta$ model (R^2 value of .853 vs. .849), but nevertheless showed a 40-msec advantage for the last serial position when fit with the $5\lambda-1\beta-2\delta$ model. In the case of rhyme judgments, the $5\lambda-1\beta-2\delta$ model also yielded the highest R^2 values for 4 of the 5 subjects, with values ranging from .896 to .943. Subject B.G. was again better fit by an alternative model, but in this case it was a $5\lambda-2\beta-1\delta$ model that produced an R^2 value of .864 as compared with .820 for the $5\lambda-1\beta-2\delta$ model. Like the other subjects, B.G. showed a dynamics advantage for the last item on the list, but it was better captured by the rate (β) than the intercept (δ) parameter (see below). A $5\lambda-1\beta-2\delta$ model was also found to give the best fit for 4 of the 5 subjects' synonym judgments, yielding R^2 values ranging from .782 to .927. In this case, subject S.G. showed a dynamics advantage for the last item on the list that was better captured in rate ($5\lambda-2\beta-1\delta$) than intercept (R^2 value of .937 vs. .924).

The $5\lambda-1\beta-2\delta$ model provides the best fit of the serial position functions for each type of judgment. In order to directly determine whether there were differences in time course between the three judgments, the data from all three judgments were simultaneously fit with Equation 1. For this analysis, each serial position within a judgment was allotted a separate λ parameter, and the two dynamics (β and δ) parameters were systematically allotted to different judgments and/or serial positions. The best fit of the average and individual subject's data was a $15\lambda-2\beta-3\delta$ model. This model allocated one β (rate) parameter to item judgments and another β parameter to both rhyme and synonym judgments. The differences in dynamics between item judgments and rhyme and synonym judgments also necessitated allotting one δ (intercept) parameter to item judgments and another δ parameter to both rhyme and synonym judgments. Additionally, consistent with the analy-

sis presented above, a separate δ parameter was needed for the fast judgments involving the last serial position (Position 5), but only one parameter was needed for all three judgments. The $15\lambda-2\beta-3\delta$ model produced an R^2 value of .967 when fit to the average data, with R^2 values ranging from .843 to .938 across the 5 subjects. No model with fewer parameters produced as high an R^2 value for the average or any individual subject's data. One more embellished model, a $15\lambda-3\beta-4\delta$ model (viz., one β and δ parameter for each judgment, plus an additional δ parameter for all judgments of Serial Position 5), produced a marginally higher R^2 value (.970) for the average data and Subject B.C. (.916 vs. .907). However, this model was rejected as a viable description of the data since it occasioned clear tradeoffs between the rate (β) and intercept (δ) parameters.

For the most recent item on the list, Serial Position 5, the $15\lambda-2\beta-3\delta$ model assigned a single intercept for all three types of judgments. Yet, for all other serial positions, this model allotted a separate intercept for item judgments and another for both rhyme and synonym judgments. The conclusion that all three judgments of the most recent item may be associated with a common intercept has important theoretical implications (see Discussion section). The validity of the $15\lambda-2\beta-3\delta$ model is reinforced by the fact that no consistent differences in intercept emerged when the intercepts were allowed to vary for different judgments of the last item on the list, as in either a $15\lambda-2\beta-4\delta$ model (δ_1 = item judgments of Serial Position 5, δ_2 = item judgments of Serial Positions 1–4, δ_3 = rhyme/synonym judgments of Serial Position 5, and δ_4 = rhyme/synonym judgments of Serial Positions 1–4) or in more embellished models such as a $15\lambda-2\beta-6\delta$ model, where the intercepts for rhyme and synonym judgments are segmented further. In the case of the $15\lambda-2\beta-4\delta$ model, for example, the intercept was estimated to be earlier for item as compared to rhyme/synonym judgments in 2 subjects (8 and 66 msec). However, this pattern reversed in the other 3 subjects. For these subjects, the intercept for rhyme/synonym judgments was estimated to be earlier than item judgments (7, 8, and 15 msec). The differences in intercept for item versus rhyme/synonym judgments across subjects can be used to construct a confidence interval around this model's intercept estimates. The 95% confidence interval for the mean difference in intercepts for item as compared to rhyme/synonym judgments is approximately -32 to 50 msec. Since zero is included in the interval, there are no statistical grounds on which to reject the hypothesis that the intercepts for judgments of Serial Position 5 are equal across the three types of judgments. Similar inconsistent and nonsignificant parameter estimates are also observed in more embellished models such as the $15\lambda-2\beta-6\delta$ or other models that allot more β parameters. Consequently, there is no evidence in these data to support the notion that the intercepts vary for different judgments of the most recent item on the list.

Table 1 presents the parameter estimates from the $15\lambda-2\beta-3\delta$ model for the average and individual subjects' data. The smooth functions in Figure 3 show the fit of the ex-

Table 1
Exponential Parameter Estimates

| Parameter | Serial Position | Subject | | | | | Composite (Average) Data |
|-------------------|-----------------|---------|-------|-------|-------|-------|--------------------------|
| | | B.C. | B.G. | M.K. | S.G. | S.H. | |
| Item Judgments | | | | | | | |
| λ | 1 | 2.02 | 2.68 | 2.83 | 2.36 | 2.43 | 2.44 |
| λ | 2 | 1.91 | 2.16 | 1.83 | 2.16 | 1.97 | 2.01 |
| λ | 3 | 2.18 | 2.67 | 2.28 | 2.65 | 2.24 | 2.39 |
| λ | 4 | 2.53 | 2.79 | 2.38 | 2.97 | 2.71 | 2.67 |
| λ | 5 | 3.71 | 3.75 | 3.14 | 4.07 | 3.43 | 3.58 |
| β | 1-5 | 5.47 | 2.19 | 3.97 | 6.62 | 4.91 | 4.29 |
| δ_1 | 5 | .217 | .234 | .254 | .278 | .333 | .258 |
| δ | 1-4 | .386 | .254 | .331 | .402 | .418 | .366 |
| Rhyme Judgments | | | | | | | |
| λ | 1 | 1.49 | 1.89 | 1.71 | 2.16 | 2.21 | 1.89 |
| λ | 2 | 1.22 | 1.86 | 1.67 | 2.17 | 1.83 | 1.75 |
| λ | 3 | 1.29 | 1.82 | 1.85 | 2.37 | 1.82 | 1.84 |
| λ | 4 | 1.80 | 1.56 | 2.14 | 2.74 | 2.09 | 2.06 |
| λ | 5 | 2.76 | 3.07 | 3.35 | 3.57 | 3.34 | 3.22 |
| β | 1-5 | 1.85 | 1.89 | 1.33 | 1.83 | 2.04 | 1.73 |
| δ_1 | 5 | .217 | .234 | .254 | .278 | .333 | .258 |
| δ_2 | 1-4 | .521 | .508 | .348 | .450 | .533 | .461 |
| Synonym Judgments | | | | | | | |
| λ | 1 | 1.20 | 2.51 | 1.51 | 2.68 | 2.34 | 2.05 |
| λ | 2 | 1.08 | 1.91 | 1.34 | 2.45 | 2.14 | 1.79 |
| λ | 3 | 1.43 | 2.00 | 1.44 | 2.82 | 1.98 | 1.94 |
| λ | 4 | 1.39 | 2.24 | 1.85 | 3.06 | 2.49 | 2.21 |
| λ | 5 | 2.08 | 2.37 | 2.22 | 3.69 | 2.96 | 2.68 |
| β | 1-5 | 1.85 | 1.89 | 1.33 | 1.83 | 2.04 | 1.73 |
| δ_1 | 5 | .217 | .234 | .254 | .278 | .333 | .258 |
| δ_2 | 1-4 | .521 | .508 | .348 | .450 | .533 | .461 |
| R^2 | | 0.907 | 0.879 | 0.843 | 0.938 | 0.936 | 0.967 |

Note—Parameters with the same subscript number were jointly estimated; see text for details. The column labeled “Composite (Average) Data” represents fits of the average (over subjects) data.

ponential Equation 1 to the average data using the parameter estimates in Table 1. The $15\lambda-2\beta-3\delta$ model served to quantify the large differences in retrieval dynamics for item versus rhyme and synonym judgments. Excluding the last serial position, item judgments were found to have a shorter intercept and faster rate of rise than either rhyme or synonym judgments. The intercept for item judgments was estimated to be on average 95 msec shorter than for either rhyme or synonym judgments, with differences ranging from 17 to 254 msec across subjects. The mean rate ($1/\beta$) for item judgments was 344 msec faster than for either rhyme or synonym judgments, with differences ranging from 72 to 499 msec across the 5 subjects. There was no evidence to suggest that the dynamics for rhyme or synonym judgments differed.

Figure 3 shows that judgments for test probes from the last serial position (Position 5) are well above chance at the earliest interruption point, when tests of other serial positions are at or near chance. As noted, this difference in dynamics was best captured by allotting a separate intercept parameter for judgments concerning Serial Position 5. The average difference in intercept (δ) for the last versus less recent serial positions was 108 msec in item judgments (ranging from 20 to 169 msec across subjects) and

203 msec in rhyme and synonym judgments (ranging from 94 to 304 msec across subjects). This retrieval advantage for the last item studied replicates similar findings reported by McElree and Doshier (1989, 1993) and Wickelgren et al. (1980). McElree and Doshier (1989, 1993) found that varying the rate (β) parameter gave a (marginally) better fit of the differences in dynamics than did varying the intercept parameter. As in the present analysis, Wickelgren et al. (1980) found that the retrieval advantage for the last studied item was slightly better captured by the intercept than the rate parameter. In some circumstances, a difference in intercept versus rate can have important theoretical significance. (For example, McElree & Doshier, 1993, argued that differences in intercept but not in rate can motivate a serial processing model.) Here, however, there is no reason to assume that anything hinges critically on whether the advantage for the last item ultimately lies in intercept or rate. Moreover, it is important to note that whether a dynamics difference is expressed in intercept or rate depends on the size and variability of the differences, as well as constraints internal to the data sets. A dynamics difference with large variability (across trials) may be better expressed in rate rather than in intercept. As the size of the dynamics difference increases relative to the vari-

ance, the same underlying effect may be better expressed in intercept than in rate. Additionally, a few SAT points at or close to the chance level can more strongly force an intercept rather than a rate difference. Either of these factors may have contributed to the slight discrepancies between the current study and the McElree and Doshier (1989) study.

Issues Concerning d' Correction

As noted, perfect performance in any condition was adjusted by a minimum-error correction process that replaces a zero error rate with an error value of 0.5%. This correction process has the desirable property of ensuring measurable d' values (see Macmillan & Creelman, 1991). However, it is important to note that such a correction process can introduce systematic artifacts into model fits of the full time-course data. If the true d' value were higher than the maximally allowed value, the correction process would have imposed an artificial ceiling on the asymptotic d' values. This in turn can serve to inflate rate estimates (see McElree & Doshier, 1989; McElree, 1993). Fortunately, there was only one condition in this experiment that was subject to a correction process, namely the late interruption points for item judgments of Serial Position 5. (All subjects had at least one of the late interruption lags for Serial Position 5 that was subject to the correction.) Consequently, it is possible that the parameter values for item recognition judgments of Serial Position 5 underestimate the true asymptotic values and overestimate the true rate parameter. It seems unlikely, however, that the correction process dramatically affected the rate estimates found here. This would imply that the rate is slower for item recognition judgments of Serial Position 5 than for rhyme and synonym judgments of the same serial position. Since every other serial position for rhyme and synonym judgments was associated with slower rates (and longer intercepts) than item judgments, it seems unlikely that the pattern would have reversed in the case of Serial Position 5.

DISCUSSION

The item recognition task replicates previous SAT studies of the time course of item retrieval from STM (McElree & Doshier, 1989, 1993; see also Wickelgren et al., 1980). Serial position had a strong effect on asymptotic accuracy: Recognition accuracy depends on the recency of study, with less recent items yielding lower recognition accuracy, coupled with a small primacy advantage for the first item on the list. There was no evidence to suggest that serial position directly affected the dynamics of memory retrieval. Uniform retrieval dynamics, specifically, equal SAT intercepts (δ s) and rates (β s), are most consistent with the notion that different list positions can be accessed directly or in parallel, despite differences in the underlying strength of the various STM representations. The only exception to this finding concerns the relatively fast dynamics for the last item studied. When no item intervenes between study and test, the test probe may be directly matched to the item that is currently active in awareness (Wickelgren et al., 1980), effectively circumventing the need for a retrieval

operation and thereby resulting in an exceptionally fast judgment.

The purpose of the present experiment was to compare memory retrieval based on item information with retrieval based on component phonological (rhyme) and semantic (synonym) properties of the memory trace. One hypothesis considered was that component properties, like composite item information, can be accessed directly or in parallel. Alternatively, it was suggested that the composite representation may not allow direct access to its component parts, or that component properties may not effectively serve to discriminate targets from lures. In either case, a different retrieval process may be needed. Two alternative retrieval mechanisms were suggested. Following McElree and Doshier (1993), component judgments may induce a serial comparison process in which items in the memory representation are serially retrieved and compared to the test probe. Alternatively, judgments could be based on the product of a generate/recall operation performed on the relevant features of the test probe.

Consider first the possibility of a serial mechanism. As noted, a serial exhaustive mechanism (e.g., Sternberg, 1975) is inconsistent with strong serial position effects on mean RT. The most viable serial mechanism is a serial self-terminating comparison process in which memory representations are scanned in an order determined by recency of study. As illustrated by McElree and Doshier (1989, 1993; see also Reed, 1976), a recency-based, serial self-terminating mechanism will lead to pronounced effects of recency on retrieval dynamics. The observed SAT function for rhyme and synonym judgments, like those for the item judgments, provide no evidence for a serial self-terminating mechanism. With the exception of a fast matching process for the last studied item, there were no observable differences in SAT dynamics (intercepts and rates) for different serial positions in both rhyme and synonym judgments. Additionally, the lack of an effect of recency on retrieval dynamics is inconsistent with parallel rate-varying models in which different comparison processes are associated with different rates of information accrual (e.g., Murdock, 1971; Townsend & Ashby, 1983).

The uniform dynamics for different serial positions suggest that rhyme and synonym judgments are mediated by a type of direct-access or parallel retrieval process. This may be the case either (1) because memory representations can be directly accessed on the basis of component properties or (2) because the test probe is used to generate/recall a candidate rhyme or synonym that may be subsequently matched against the memory representation. These theories differ in that (2) necessarily predicts that rhyme and synonym judgments will be associated with slower dynamics than item judgments, since an additional generation operation precedes the memory comparison process or simply because a slower recall process is used.

The SAT cross-judgment analysis demonstrated that rhyme and synonym judgments are indeed associated with slower dynamics than item judgments. Slower retrieval dynamics for component judgments were expressed as slower rates of rise to asymptote (β s) and longer intercepts

(δ s). Consider first the differences in SAT rates. While the difference in rate for component versus item judgments is consistent with the notion that a slower generate/recall process is used in the former judgments, it is also consistent with possible ways in which some direct-access or parallel models could be extended to deal with cases in which the memory representation is accessed on the basis of component properties. Consider, for example, Ratcliff's (1978) diffusion or (continuous) random-walk model, perhaps the most extensively tested model of memory retrieval. In this model, the rate of rise of an SAT function is determined by the ratio of the variance of the diffusion drift rate over the variance of the relatedness (degree of match) of the test probe and the memory representation (see note 3). SAT rate differences can arise if two conditions differ in either variance term. This model could account for a slower rate for component judgments by assuming that the variance is lower for component judgments than for composite item judgments, since the former are based on a subset of features only. For this or similar reasons, a difference in SAT rate alone does not provide unequivocal evidence against the potential extension of this model to judgments based on component properties.

However, a direct-access or parallel model cannot account straightforwardly for the differences in intercept observed in this study. The intercept for synonym and rhyme judgments for serial positions other than the most recent (Position 5) were on average 95 msec later than for item judgments. In a direct-access or parallel retrieval model, this delay in intercept could reflect the time to extract the relevant subset of (phonological and semantic) features in the test probe that are subsequently used to directly query the memory representation. However, no comparable intercept difference was found in Serial Position 5, where no study item intervened between study and test. This speaks against the notion that the delayed intercepts simply reflect the time to extract the relevant features of the test probe, since this type of operation should also be involved with judgments concerning the last item on the list.

The overall pattern of rates and intercepts is most consistent with a generate/recall strategy. Here, the slower dynamics for judgments based on component properties can be taken to be a direct consequence of the generation or recall operation. Perhaps the simplest version of this account is one in which the intercept differences directly reflect the time needed to generate/recall a candidate item on the basis of features extracted from the test probe. As stated previously, this account assumes that word recognition processes applied to the probe display make available the relevant component features of the test item. As these features become available, they serve as input to a generation or recall operation. No comparable intercept differences are observed in judgments of the most recent item (Serial Position 5—a case of immediate repetition between study and test). No generate/recall operation is needed since the relevant item to be generated/recalled is the item that is currently active in awareness. However, rhyme and synonym judgments in this condition are associated with a slower SAT rate than item judgments because

they are based on a subset of features and thus require the subject to extract the relevant features from the test probe before matching the selected features to the item active in awareness.

The account outlined above assumes that subjects generate or recall 1 item only in the course of making a judgment. However, if the first item generated or recalled does not adequately match an item in the memory representation, then additional items may need to be generated and/or recalled. Additional generation/recall operations may well affect the observed dynamics of an SAT function. In general, the intercept will be controlled by the minimum number of generation/recall operations needed across trials. Thus, additional operations will affect the intercept only if they are needed on nearly all trials. It is more likely that additional operations are required on a subset of trials only. In this case, the impact of additional operations will be seen on SAT rate only, causing a slower rate of rise than what would be observed if only one match were required (see McElree, 1993, for SAT simulation of a stochastic model with analogous properties).

Before accepting the interpretation of the SAT data outline above, there is one other alternative interpretation that needs to be considered.⁴ Suppose that, without foreknowledge of the required judgment, subjects prepare for item recognition, switching to synonym or rhyme judgments only when explicitly cued. Here, the longer intercept (and slower rate) for rhyme and synonym judgments simply reflect the time needed to switch judgments. However, there are a number of reasons why such an interpretation seems implausible in the current situation. First, subjects were given the cue for the type of judgment 500 msec before the test item appeared. The 500-msec foreperiod was chosen because pilot subjects indicated that with practice they were able to adequately prepare for the required response. Second, the rationality of such a default strategy is questionable. Subjects were fully informed that they would be required to make each type of judgment equally often. Why then would subjects opt for a strategy that would require switching two thirds of the time and would, by hypothesis, introduce substantial time delays? Subjects in this experiment were highly practiced, with each subject performing 1 hour of practice and 10 experimental hours. If the time-course differences observed here were purely a consequence of a strategic set, it seems reasonable that experienced subjects could have modified their set in a manner that would have eliminated these rather large time differences. Finally, and perhaps most critically, this notion is inconsistent with the lack of an intercept difference for judgments concerning the last item on the list. If the intercept difference reflects an item recognition set, we should have observed comparable differences in all serial positions.

It is suggested, therefore, that the data reported here are most consistent with the notion that component judgments are mediated by either a generate or a recall strategy. Clearly, these are potentially rather distinct types of operations. For example, the recall operation posited here uses component properties of the test item as a cue to directly query the episodic list representations in order to recover

a candidate rhyme or synonym from the list. The exact details of such an operation need to be specified within the context of an explicit memory model (e.g., MINERVA2, Hintzman, 1984, 1988; SAM, Gillund & Shiffrin, 1984; TODAM, Murdock, 1982, 1983) and are clearly beyond the scope of this article. The generation operation may differ from the recall operation in that component properties of the test item are used to recover candidate items from the subject's semantic or lexical representations rather than from episodic representations. It is only after generating candidate items that the subject is assumed to access episodic memory by matching the candidate item(s) against the specific list representations. Of course, this potential difference between the two operations is a meaningful one only to the degree that episodic representations are distinct from semantic (lexical) representations, and this has been a matter of some debate over the past few decades. Note also that current work on episodic priming in generation tasks like stem and word fragment completion suggests that prior exposure of a word may substantially increase the probability that this word will be produced in the generation task (see, e.g., Richardson-Klavehn & Bjork, 1988). Thus, even if episodic representations are independent of semantic representations under some meaningful construal, current evidence suggests that there may be sufficient cross-talk between systems to render the distinctions outlined above empirically vacuous.

Another potential difference between the two strategies is that the generate strategy requires that the candidate items be matched to the list representation in a manner analogous to standard item recognition. The recall strategy does not necessarily require an additional recognition operation following recovery of a list member, since, in principle, a subject could base an *old-new* judgment on the success/failure of the recall operation. However, although not logically necessary, a recall process may be followed by a recognition process in which the recovered item is used as a means of providing further evidence to support a *yes/no* decision. Again, although these are two theoretically distinct operations, there may be few empirical differences between a pure generation operation and an episodically based recall operation.

Finally, prior studies of recognition in the short-term domain have suggested that when there is no item intervening between study and test, a recognition probe can be directly matched against the item currently active in awareness. The present experiment further extends this finding to cases in which the relation between the study and tested item is not one of identity but rather only an abstract relation defined by the similarity in component (phonological or semantic) features. This finding reinforces the original suggestion of Wickelgren et al. (1980) that the retrieval advantage for the last study item is mediated by a match of the test probe to the item that is the current focus of active processing and, in particular, is not a consequence of a relatively low-level physical match (see Posner et al., 1969). This finding has direct implications for models of STM. Minimally, it suggests that an adequate model needs to draw an architectural distinction between passive and active

STM representations, with the latter defined as the focus of attention or active processing. Cowan (1993) has argued on independent grounds for just such a distinction.

CONCLUSION

The data reported here suggest that recognition judgments based on phonological and semantic properties of the memory representation have a later retrieval intercept than do (composite) item recognition judgments. The delayed intercept is consistent with the notion that either a generate or a recall strategy is used for accessing STM representations from component properties. This result adds to a growing literature that documents boundary conditions on the parallel or direct-access retrieval mechanism that is thought to mediate simple item recognition (Clark, 1992; Gronlund et al., 1993; Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994; McElree & Doshier, 1993). Why recognition judgments based on component properties do not appear to be based on a simple parallel or direct-access mechanism remains an open question. Two potential reasons were suggested here, both of which have some precedent in the recognition literature.

One suggestion was that the memory trace for an item might be "opaque" in the sense of not being accessible on the basis of component features. If this notion is correct, it indicates that episodic representations have a property that is clearly not shared by semantic representations, since the ability to generate a rhyme or synonym is *prima facie* evidence that semantic representations are accessible on the basis of component features. This type of explanation attributes the use of an alternative retrieval operation to constraints intrinsic to the (episodic) memory representation. In this sense, it is roughly analogous to McElree and Doshier's (1993) claim that order information must be retrieved by a slow serial retrieval operation because the memory representation does not code order information in a manner that enables recovery with a parallel or direct-access mechanism. Gronlund et al. (1993) considered similar arguments concerning the recovery of position information.

Alternatively, it was suggested that STM representations might be accessible on the basis of component properties but that a generate/recall operation is used to provide more evidence to support a response. This explanation does not appeal to intrinsic constraints on the memory representation but rather only to the effective discriminability of component information. This type of explanation is similar to Hintzman and Curran's (1994) argument that a recall-like operation is needed for difficult recognition discrimination. Determining whether a generate/recall operation is a consequence of decision factors or intrinsic constraints on memory representations will require further empirical work as well as specific, model-based considerations.

REFERENCES

- ANDERSON, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, **80**, 417-438.
- BADDELEY, A. D., & ECOb, J. R. (1973). Reaction time and short-term memory: Implications of repetition effects for high speed exhaustive

- scan hypothesis. *Quarterly Journal of Experimental Psychology*, **25**, 229-240.
- BECKER, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 252-259.
- BOWER, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1, pp. 230-325). New York: Academic Press.
- CHANDLER, J. P. (1969). Subroutine STEPIT—Finds local minimum of a smooth function of several parameters. *Behavioral Science*, **14**, 81-82.
- CLARK, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, **20**, 231-243.
- COWAN, N. (1993). Activation, attention, and short-term memory. *Memory & Cognition*, **21**, 162-167.
- 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. (1982). Sentence size, network distance and sentence retrieval. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **8**, 173-207.
- DOSHER, B. A., & MCELREE, B. (1992). Memory search. In L. R. Squire (Ed.), *Encyclopedia of learning and memory* (pp. 398-406). New York: Macmillan.
- 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.
- DREWNOWSKI, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General*, **109**, 208-250.
- EICH, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, **89**, 627-661.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GRONLUND, S. D., EDWARDS, M. B., & OHRT, D. D. (1993). *Item information precedes position information in recognition*. Paper presented at the annual meeting of the American Psychological Society, Chicago.
- GRONLUND, S. D., & RATCLIFF, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 846-858.
- HACKER, M. J. (1980). Speed and accuracy of recency judgments for event in short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **6**, 651-675.
- 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.
- HINTZMAN, D. L., & CURRAN, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory & Language*, **33**, 1-18.
- JOHNSON, N. F. (1972). Organization and the concept of a memory code. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 125-159). Washington DC: V. H. Winston.
- JUDD, C. M., & MCCLELLAND, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego: Harcourt Brace Jovanovich.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MCELREE, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory & Language*, **32**, 536-571.
- 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. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, **122**, 291-315.
- MURDOCK, B. B., JR. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, **10**, 289-291.
- MURDOCK, B. B., JR. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- MURDOCK, B. B., JR. (1983). A distributed memory model for serial order information. *Psychological Review*, **90**, 316-338.
- MURDOCK, B. B., JR. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, **100**, 183-203.
- MUTER, P. A. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **5**, 160-169.
- NAIRNE, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, **16**, 343-352.
- NAIRNE, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, **18**, 251-269.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
- POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs*, **79** (1, Pt. 2).
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- REED, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, **181**, 574-576.
- REED, A. V. (1976). The time course of recognition in human memory. *Memory & Cognition*, **4**, 16-30.
- RICHARDSON-KLAVEHN, A., & BJORK, R. A. (1988). Measures of memory. *Annual Review of Psychology*, **39**, 475-543.
- SHULMAN, H. G. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **9**, 499-508.
- STERNBERG, S. (1966). High speed scanning in human memory. *Science*, **153**, 652-654.
- STERNBERG, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, **57**, 421-457.
- STERNBERG, S. (1975). Memory-scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **27**, 1-32.
- 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. 44-85). New York: Academic Press.
- THEIOS, J., SMITH, P. G., HAVILAND, S. E., TRaupMAN, J., & MOY, M. C. (1973). Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, **97**, 323-336.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- UNDERWOOD, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.
- WICKELGREN, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, **41**, 67-85.
- WICKELGREN, W. A., CORBETT, A. T., & DOSHER, B. A. (1980). Priming and retrieval from short-term memory: A speed-accuracy trade off analysis. *Journal of Verbal Learning & Verbal Behavior*, **19**, 387-404.

NOTES

1. There are a number of issues that a detailed model of this sort would need to address, most of which are beyond the purview of the present work. For example, what is the nature of the generation process? If more than one candidate item is generated, are the items compared to the episodic representation in a parallel or serial fashion? If the latter, what determines the order of the comparison?

2. If the recency effects in synonym and rhyme judgments are a function of a serial mechanism rather than of differences in memory strength, as in the case of item judgments, we might expect the impact of serial position to be more pronounced in the former cases (assuming that the serial comparison process is relatively slow). Shulman (1970) reported that the slope of the function that relates mean RT to recency was larger for synonym than for item judgments, but found no differences in slope for

rhyme and item judgments. However, given that rhyme and item judgments differ in mean RT but not error rates, the latter comparison is extremely suspect. Note also that slope differences are far from diagnostic, since such differences may reflect the differential impact of forgetting on the various types of information.

3. The empirical SAT functions can also be adequately fit with a descriptive equation that derives from a time-bounded diffusion process, as developed by Ratcliff (1978):

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

This equation, like the exponential Equation 1, has three parameters: λ represents the asymptotic accuracy level, δ the intercept, and v^2 is the combined random-walk variance term formed by the ratio of drift rate variance (S^2) over item relatedness variance (η^2). v^2 , like β in Equation 1, indexes the speed with which accuracy rises from chance to asymptote. Fits of diffusion equation showed the same pattern as fits of Equation 1, although Equation 1 gave marginally better fits, as is typical for STM recognition data (see McElree & Doshier, 1989). To conserve space, only the fits of the exponential are reported. The parameter estimates for the diffusion equation are available from the author.

4. This interpretation was suggested by a reviewer, R. Proctor.

APPENDIX A
Experimental Materials

| Item | Synonym | Rhyme | Item | Synonym | Rhyme |
|-----------|------------|------------|---------|------------|---------|
| ALONE | SINGLE | PHONE | KNOW | UNDERSTAND | BOW |
| ANGLER | FISHERMAN | WRANGLER | LANK | GAUNT | SANK |
| ANSWER | REPLY | CANCER | LAW | STATUE | RAW |
| APPEAL | PRAYER | SEAL | LIMB | BRANCH | BRIM |
| BECKON | SUMMON | RECKON | LINGER | DALLY | FINGER |
| BELIEF | CONVICTION | RELIEF | METAL | STEEL | PETAL |
| BIBLE | SCRIPTURE | LIABLE | MIDDLE | CENTER | DIDDLE |
| BOAT | SHIP | FLOAT | MINE | DRILL | LINE |
| BOG | SWAMP | FOG | NEW | MODERN | BLUE |
| BROW | FOREHEAD | SOW | NOOK | CORNER | BOOK |
| BREAK | SMASH | RAKE | PAIL | BUCKET | FAIL |
| BUG | INSECT | RUG | PARASOL | UMBRELLA | AEROSOL |
| BUNNY | RABBIT | FUNNY | PISTOL | REVOLVER | DISTAL |
| CAR | AUTOMOBILE | FAR | PLEDGE | PROMISE | HEDGE |
| CARESS | FONDLE | REGRESS | PORT | HARBOR | SORT |
| CASTLE | PALACE | HASSLE | PROOF | EVIDENCE | SLEUTH |
| CELLAR | BASEMENT | TELLER | QUEST | SEARCH | ZEST |
| CHURCH | TEMPLE | LURCH | QUIT | SILENT | RIOT |
| CLAW | TALON | SAW | REST | REPOSE | BEST |
| CLEAR | FREE | BEER | ROCK | STONE | LOCK |
| CLOCK | TIMEPIECE | ROCK | SCOOP | LADLE | LOOP |
| COACHED | INSTRUCTED | POACHED | SHORE | BEACH | LORE |
| COST | EXPENSE | LOST | SHREWD | ASTUTE | LEWD |
| DAWN | DAYBREAK | LAWN | SINGER | VOCALIST | RINGER |
| DEED | ACT | REED | SIZE | MAGNITUDE | RISE |
| DESCRIBED | RECOUNTED | PRESCRIBED | SKIN | FLESH | BIN |
| DINNER | SUPPER | WINNER | SOURCE | ORIGIN | FORCE |
| DOCTOR | PHYSICIAN | PROCTOR | SPEED | VELOCITY | SEED |
| DOG | CANINE | FOG | STEED | STALLION | FEED |
| DRAPES | CURTAIN | GRAPES | STORE | SHOP | CHORE |
| DUTY | OBLIGATION | BOOTY | STREET | ROAD | GREET |
| ENTICE | ALLURE | CONCISE | STYLE | FASHION | TILE |
| EXAM | TEST | CLAM | SWAMP | MARSH | ROMP |
| FARCE | MOCKERY | PARSE | TALE | STORY | WHALE |
| FEELING | SENSATION | REELING | TALES | STORIES | BAILS |
| FLOWER | BLOSSOM | SOUR | TALK | LECTURE | BALK |
| GEM | JEWEL | HEM | TAVERN | SALOON | CAVERN |
| GLOSS | SHEEN | TOSS | THOUGHT | IDEA | SOUGHT |
| GORED | PIERCED | BORED | TOIL | DRUDGE | SOIL |
| GRIME | DIRT | CRIME | TRADE | BARTER | CHARTER |
| HAMMER | MALLET | YAMMER | TRIAL | HEARING | PILE |
| HATCHET | TOMAHAWK | RATCHET | TROUBLE | DIFFICULTY | DOUBLE |
| HOME | HOUSE | COMB | TRUTH | VERACITY | COUTH |
| HORSE | STEED | COURSE | TUB | BUCKET | HUB |
| INN | HOTEL | PIN | VEST | WAISTCOAT | PEST |
| INTERIOR | INSIDE | INFERIOR | VILLAGE | TOWN | PILLAGE |
| ISSUE | EDITION | TISSUE | WOOD | LUMBER | COULD |
| JAIL | PRISON | BAIL | WRITER | AUTHOR | TIGHTER |
| JINX | HEX | LYNX | YARD | FIELD | CARD |
| JUDGE | MAGISTRATE | FUDGE | | | |

APPENDIX B
Means and Standard Deviations of the Lag + Latencies (in Seconds),
Hit Rates, and False Alarm Rates Across Subjects for Each Judgment

| | | Interruption Lag (in Seconds) | | | | | | | | | | | |
|-------------------|---------------|-------------------------------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | | 0.128 | | 0.357 | | 0.500 | | 0.786 | | 1.200 | | 3.000 | |
| | | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Item Judgments | | | | | | | | | | | | | |
| Serial Position 1 | Lag + Latency | 0.408 | .020 | 0.585 | .022 | 0.717 | .019 | 0.994 | .020 | 1.405 | .019 | 3.194 | .018 |
| | Hits | 0.536 | .117 | 0.685 | .125 | 0.846 | .041 | 0.810 | .052 | 0.803 | .089 | 0.914 | .072 |
| Serial Position 2 | Lag + Latency | 0.397 | .022 | 0.585 | .025 | 0.716 | .021 | 0.990 | .012 | 1.409 | .014 | 3.200 | .011 |
| | Hits | 0.562 | .227 | 0.702 | .101 | 0.774 | .080 | 0.750 | .020 | 0.682 | .086 | 0.731 | .049 |
| Serial Position 3 | Lag + Latency | 0.408 | .014 | 0.580 | .019 | 0.712 | .017 | 0.991 | .021 | 1.404 | .015 | 3.195 | .016 |
| | Hits | 0.520 | .089 | 0.729 | .102 | 0.805 | .085 | 0.817 | .082 | 0.891 | .051 | 0.833 | .080 |
| Serial Position 4 | Lag + Latency | 0.399 | .019 | 0.578 | .026 | 0.703 | .016 | 0.987 | .013 | 1.401 | .014 | 3.195 | .020 |
| | Hits | 0.585 | .180 | 0.770 | .078 | 0.840 | .089 | 0.891 | .034 | 0.897 | .036 | 0.919 | .035 |
| Serial Position 5 | Lag + Latency | 0.367 | .022 | 0.565 | .020 | 0.693 | .015 | 0.985 | .017 | 1.399 | .015 | 3.192 | .021 |
| | Hits | 0.847 | .128 | 0.934 | .059 | 0.988 | .024 | 0.960 | .044 | 0.993 | .013 | 0.987 | .016 |
| Negatives | Lag + Latency | 0.406 | .014 | 0.590 | .014 | 0.717 | .013 | 0.994 | .009 | 1.416 | .008 | 3.205 | .015 |
| | False Alarms | 0.448 | .203 | 0.191 | .082 | 0.169 | .056 | 0.123 | .056 | 0.102 | .086 | 0.083 | .006 |
| Synonym Judgments | | | | | | | | | | | | | |
| Serial Position 1 | Lag + Latency | 0.406 | .016 | 0.592 | .018 | 0.725 | .018 | 0.996 | .012 | 1.415 | .017 | 3.202 | .017 |
| | Hits | 0.629 | .142 | 0.635 | .064 | 0.678 | .034 | 0.688 | .155 | 0.753 | .101 | 0.736 | .084 |
| Serial Position 2 | Lag + Latency | 0.412 | .024 | 0.591 | .020 | 0.726 | .019 | 1.011 | .021 | 1.412 | .012 | 3.204 | .017 |
| | Hits | 0.616 | .159 | 0.513 | .148 | 0.669 | .137 | 0.634 | .120 | 0.669 | .061 | 0.694 | .097 |
| Serial Position 3 | Lag + Latency | 0.404 | .016 | 0.595 | .019 | 0.725 | .016 | 1.002 | .014 | 1.410 | .014 | 3.201 | .012 |
| | Hits | 0.594 | .161 | 0.581 | .147 | 0.676 | .091 | 0.649 | .037 | 0.686 | .058 | 0.779 | .070 |
| Serial Position 4 | Lag + Latency | 0.406 | .017 | 0.599 | .016 | 0.726 | .019 | 1.002 | .011 | 1.411 | .014 | 3.198 | .017 |
| | Hits | 0.586 | .158 | 0.660 | .107 | 0.657 | .084 | 0.704 | .082 | 0.758 | .075 | 0.815 | .079 |
| Serial Position 5 | Lag + Latency | 0.396 | .019 | 0.582 | .018 | 0.712 | .017 | 0.988 | .008 | 1.400 | .013 | 3.197 | .015 |
| | Hits | 0.749 | .143 | 0.793 | .040 | 0.869 | .042 | 0.880 | .039 | 0.883 | .034 | 0.929 | .036 |
| Negatives | Lag + Latency | 0.404 | .019 | 0.594 | .015 | 0.728 | .013 | 1.009 | .014 | 1.423 | .008 | 3.209 | .016 |
| | False Alarms | 0.638 | .188 | 0.475 | .136 | 0.322 | .113 | 0.254 | .104 | 0.192 | .119 | 0.117 | .093 |
| Rhyme Judgments | | | | | | | | | | | | | |
| Serial Position 1 | Lag + Latency | 0.404 | .025 | 0.589 | .022 | 0.724 | .017 | 0.994 | .012 | 1.409 | .019 | 3.200 | .017 |
| | Hits | 0.664 | .210 | 0.507 | .146 | 0.604 | .106 | 0.603 | .112 | 0.667 | .094 | 0.702 | .108 |
| Serial Position 2 | Lag + Latency | 0.409 | .025 | 0.586 | .019 | 0.722 | .017 | 1.005 | .016 | 1.414 | .017 | 3.201 | .018 |
| | Hits | 0.656 | .198 | 0.538 | .173 | 0.542 | .066 | 0.632 | .076 | 0.617 | .150 | 0.643 | .095 |
| Serial Position 3 | Lag + Latency | 0.403 | .022 | 0.592 | .020 | 0.720 | .015 | 0.997 | .012 | 1.409 | .018 | 3.201 | .020 |
| | Hits | 0.671 | .182 | 0.591 | .140 | 0.557 | .095 | 0.616 | .061 | 0.673 | .096 | 0.647 | .104 |
| Serial Position 4 | Lag + Latency | 0.400 | .022 | 0.589 | .022 | 0.720 | .016 | 0.998 | .020 | 1.415 | .021 | 3.200 | .020 |
| | Hits | 0.690 | .214 | 0.499 | .106 | 0.596 | .116 | 0.685 | .095 | 0.672 | .095 | 0.766 | .095 |
| Serial Position 5 | Lag + Latency | 0.386 | .029 | 0.574 | .023 | 0.700 | .015 | 0.987 | .021 | 1.402 | .017 | 3.199 | .027 |
| | Hits | 0.849 | .087 | 0.857 | .083 | 0.931 | .045 | 0.936 | .025 | 0.936 | .019 | 0.932 | .056 |
| Negatives | Lag + Latency | 0.399 | .025 | 0.592 | .016 | 0.725 | .013 | 1.003 | .007 | 1.419 | .009 | 3.205 | .015 |
| | False Alarms | 0.648 | .197 | 0.440 | .147 | 0.261 | .144 | 0.223 | .084 | 0.144 | .072 | 0.107 | .065 |

(Manuscript received August 4, 1994;
revision accepted for publication February 22, 1995.)