

Parallel versus serial processing and individual differences in high-speed search in human memory

JAMES T. TOWNSEND and MARIO FIFIC
Indiana University, Bloomington, Indiana

Many mental tasks that involve operations on a number of items take place within a few hundred milliseconds. In such tasks, whether the items are processed simultaneously (in parallel) or sequentially (serially) has long been of interest to psychologists. Although certain types of parallel and serial models have been ruled out, it has proven extremely difficult to entirely separate reasonable serial and limited-capacity parallel models on the basis of typical data. Recent advances in theory-driven methodology now permit strong tests of serial versus parallel processing in such tasks, in ways that bypass the capacity issue and that are distribution and parameter free. We employ new methodologies to assess serial versus parallel processing and find strong evidence for pure serial or pure parallel processing, with some striking apparent differences across individuals and interstimulus conditions.

The short-term memory study published by Saul Sternberg in 1966 has led to numerous empirical, applied, and theoretical investigations of search in human short-term memory. Nevertheless, a survey of the literature suggests that even now, the case is not closed with regard to the mental architecture and processes associated with Sternberg's task. Here we will briefly review the issues involved and then introduce a new and powerful methodology based on metatheoretical developments and a novel experimental strategy. We propose that this new approach can help settle critical questions in applied, basic memory, and perceptual research.

In the Sternberg (1966) paradigm, a small set of items is quickly memorized, and this is followed by the presentation of a target item. The experimental participant indicates, usually with one of two buttonpresses, whether the target item was present in (a "yes" response) or absent from (a "no" response) the memory set. The primary measure of interest in such studies is response time (RT); response errors are generally less than 10%, and error rates at this level are thought to be largely unconnected

with the major aspects of processing in such an experiment (Swanson & Briggs, 1969).¹ In Sternberg's original study and in most subsequent experiments, the primary interest has been in whether comparisons of target and memory set items take place as *serial* (one at a time) or *parallel* (simultaneous) processing, and the mean RTs enable the researcher to make this determination.

Although Sternberg (1975) argued forcefully that processing in such short-term memory search is indeed serial, not all studies and approaches have led to the same conclusion (Ashby, Tein, & Balakrishnan, 1993; McElree & Doshier, 1989; Ratcliff, 1978). It is also apparent that within the common paradigms that are used to assess serial versus parallel processing, models of either kind could be mathematically prescribed to mimic the opposite kind of processing (Townsend, 1969, 1971, 1990; Townsend & Ashby, 1983). Such findings have portended thorny challenges concerning our ability to empirically pry these notions apart. In addition, a related challenge has been to test entire categories of models—in this case, parallel versus serial, rather than, say, specific models based on particular probability distributions on processing times. Therefore, early on, Townsend and colleagues maintained that more powerful methods were required in order to settle the case for broad classes of parallel and serial models (i.e., rather than those restricted to special cases; e.g., Townsend, 1974; Townsend & Schweickert, 1985).

We now turn to a more detailed introduction of the characteristics of the basic model within the classic search design. Serial models assume that within the short-term memory search task, each item in memory is compared with the target in sequence, so that a comparison on an item must be completed before the next comparison can begin. Parallel models assume, in contrast, that all items are compared with the target simultaneously, although

The results of this research were first presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL on November 17, 2001. The first author is indebted to the Hanse Institute for Advanced Study for a 10-month fellowship that enabled part of this work; to a grant from the NIMH (RO1 MH57717-04A1), which supported the research; and to the Indiana University Cognitive Science Summer Program Research Fellowship. The second author received support from the same NIMH grant. Useful comments and insightful criticism was provided by the anonymous reviewers. We benefited from discussions with Saul Sternberg, Richard Schweickert, Michael Wenger, and Åse Innes-Ker. We also appreciate astute critiques from the editor and anonymous reviewers. Correspondence concerning this article should be addressed to J. T. Townsend, Department of Psychology, 1101 East 10th St., Bloomington, IN 47405-7007, or to the second author (same address) (e-mail: jtowndsen@indiana.edu, mfific@indiana.edu).

comparisons on different items can be finished at different times. It is generally taken for granted that the comparison times and therefore the RTs are probabilistic in nature and thus must be described by their frequency functions (e.g., cumulative distribution functions) or various statistics such as the mean and variance.

On target-present trials, it may be possible for the comparisons to cease once the target has been located in the memory set, permitting the “yes” response to be implemented before all the items in the memory set have been compared with the target. This is known as *self-terminating* processing. On target-absent trials, however, all memory items must be compared before the participant can be certain that the target was not present; only then can a participant make a correct “no” response. Such processing is referred to as *exhaustive* processing. Target-absent trials will be the focus of the present investigation, in which we employ recently designed theoretical and methodological tools to attempt to settle the parallel–serial issue.

Sternberg (1966) varied the number of items (digits in this case, but the design has since been carried out over a huge array of stimulus types) was varied from $n = 1$ to $n = 6$ (i.e., less than the usual upper bound on short-term memory). Sternberg’s main finding was that mean RTs gathered under conditions of very low error rates were essentially an increasing linear function of n , with a slope of about 40 msec and an intercept of around 400 msec.

Sternberg (1966) showed how such a finding could result from a serial model, with each item consuming the same duration, on average, for its comparison. However, he was also able to falsify an important class of parallel models. That class is now referred to as *unlimited-capacity, independent* parallel processing (Townsend & Ashby, 1983). The “unlimited-capacity” designation indicates that the speed of any individual item, reflected in its processing time distribution, does not change and, in particular, does not exhibit a slow-down (e.g., by shifting its cumulative distribution function or mean upward). The “independent” designation indicates that the joint distribution function for the completion times of the various item comparisons is composed of the product of the individual item distributions; that is, they are stochastically independent. Employing certain mathematical results from statistics, Sternberg (1966) showed that even the fastest growing RT curve possible, associated with such parallel models, had to be negatively accelerated and could not keep pace with the linear slope of the obtained data. Subsequently, Townsend and Ashby (1983) proved that every independent, unlimited-capacity model had to be negatively accelerated—not just the class of models producing the fastest growing RT curves.

It was later demonstrated, however, that a different type of parallel model could predict linear mean RT curves such as those found in the data (Atkinson, Holmgren, & Juola, 1969; Murdock, 1971; Townsend, 1971; Townsend & Ashby, 1983). This kind of model is called *limited capacity*, since it allows for the processing rates

of individual items to decrease as the memory set size n increases. Many investigators find such models to be quite plausible, given that limitations in human cognitive, perceptual, and motor performance abound in the cognitive and human performance literature.

On the other hand, if the RT curves were completely flat or almost so, parallel processing would be supported. Serial models have to increase their individual item processing speed proportionately to n and thus are usually unrealistic theoretical competitors in the face of such results. Although self-terminating standard types of parallel models (e.g., unlimited capacity and independent) naturally predict flat RT functions with one target present (Van Zandt & Townsend 1993), exhaustive parallel processing by such systems cannot produce a nonincreasing RT function of n . It requires supercapacity parallel models to make such predictions; that is, the average rates on individual channels must actually increase their processing speed as the overall load (here, the set size) increases. Such curves are empirically rare but do occur (albeit in visual search or search involving visual features; see, e.g., Egeth, Jonides, & Wall, 1972; Nielsen & Smith, 1973; Schneider & Shiffrin, 1977).

Since Sternberg’s (1966) original study, quantitative methods have been devised that permit stronger tests of parallel versus serial processing for a broad set of perceptual, cognitive, and performance tasks (Townsend, 1990). Several of these methods are not afflicted with the dilemma posed by limited-capacity parallel models. One especially promising methodology grew out of another innovation by Sternberg—the *additive factors method* (Sternberg, 1969). This method tests whether processing is serial by manipulating experimental variables that selectively affect (assumption of *selective influence*) the different stages (e.g., item comparisons) of processing. If mean RTs are additive in the experimental variables, then the RTs are compatible with serial processing. If mean RTs are interactive across the factor manipulation, then Sternberg’s original method either concludes that one of the factors affected both stages, or that two separate serial stages simply do not exist.

A useful statistic to capture an interaction between two factors with two levels is the mean interaction contrast, $M_{IC} = RT_{ll} - RT_{lh} - (RT_{hl} - RT_{hh})$ (e.g., Sternberg, 1969). In this formula, the left subscript designates the level of one factor (e.g., the factor affecting the item’s comparison time in first position of presentation, in a spatial or temporal sense; h = high, indicating fast RTs, and l = low, indicating slow RTs) and the second subscript designates the level of the other factor (e.g., the factor affecting the processing time of the item in the second position). Notice that the mean interaction contrast is basically a difference across the first factor, of two differences across the second factor. If the mean interaction is 0, as predicted by serial processing with exhaustive processing, then $M_{IC} = 0$.

A limitation of the original additive factors method was that it did not encompass parallel or more complex

kinds of systems. That is, within the original methodology, if a nonzero interaction was detected, the prescribed inference was either that selective influence was failing or that, simply speaking, two separate and serial stages did not exist. Thus, there was no provision for how architectures other than serial might behave if the items or channels were selectively influenced by experimental factors.

The scope of factorial methodologies was gradually expanded to include parallel and more complex architectures, assuming selective influence on the various subsystems, items, or processing stages (Schweickert, 1978; Schweickert & Townsend, 1989; Townsend & Ashby, 1983; Townsend & Schweickert, 1985). If processing obeys an exhaustive stopping rule (due, for instance, to no target's being present in a search task), then parallel models predict that $M_{IC} < 0$ (see, e.g., Townsend & Ashby, 1983, p. 373) rather than $M_{IC} = 0$, as in the case of serial processing.

We have recently developed a curve that captures factorial interactions at a considerably more sensitive level than is the case with M_{IC} . For instance, whereas M_{IC} is a relatively coarse point (i.e., single number) measure, the new measure delivers an entire curve theoretically and as many points as are used across time to estimate the frequency functions for each setting of the factors (Townsend & Nozawa, 1995). Basically, if $F_{ij}(t)$ represents the cumulative distribution function for factor settings i, j , then let $S_{ij}(t) = 1 - F_{ij}(t)$, a mathematical curve known in information processing (and statistics) as the *survivor function* (Townsend & Ashby, 1983, p. 27). Then we define the *survivor interaction contrast function* as $S_{IC}(t) = S_{ll}(t) - S_{lh}(t) - [S_{hl}(t) - S_{hh}(t)]$. Observe that the same logical sequence of terms relative to the factor levels is used as in the case of M_{IC} , but now there is an approximation of an entire curve, given by the appropriate values of the survivor functions at the time bins used for sampling RTs.

In addition, the new interaction contrast statistical function $S_{IC}(t)$ permits discrimination of processing models that make the same prediction when only M_{IC} is used (Townsend & Nozawa, 1995). For instance, there are presently two major types of parallel processing. The first, which we could call "regular" or "separate decisions" parallel processing, assumes that each channel possesses its own detection stage (e.g., a criterion indicating presence of a feature). After one or more channels have a criterial quantity of information (activation, etc.), subsequent logical gates may apply (e.g., OR, AND, XOR, etc.). (For instance, in an OR design, the first channel or item completed determines the correct response, whereas in an AND design, both channels or items must be completed to ensure a correct response.) In contrast, the second major type of parallel processing, called "coactive," assumes that before any decision is made, the various channels pool their combined activation or information into a common resultant channel, after which a criterion can be reached and a response made. Townsend and Nozawa (1995) showed that regular and coactive parallel models make the same mean interaction contrast prediction for

OR designs. However, they proved that the two parallel types of predictions differ when the more powerful survivor interaction contrast assays are employed.

Furthermore, the qualitative form of the $S_{IC}(t)$ function is distribution free (i.e., any probability distribution, such as gamma, Weibull, ex-Gaussian, etc., must predict the same qualitative form). These distribution-free characteristics render the $S_{IC}(t)$ functions, like their predecessor, the M_{IC} statistic, considerably more powerful than model predictions that may depend on specific distributional assumptions.

The upshot is that now it becomes theoretically feasible, under the assumption of selective influence, to rigorously investigate parallel, serial, and even considerably more complex systems (e.g., Schweickert & Giorgini, 1999) in ways not troubled by limited-capacity concerns, since the number of subprocesses (e.g., item comparisons) need not be varied. Recall that in the standard approach, the rate of ascent of the mean RT curves as functions of number of items, n , was used to test serial versus parallel processing. Unfortunately, as noted earlier, limited-capacity parallel models could mimic serial models (e.g., Townsend, 1969, 1971).

Even the stopping rule can be simultaneously (i.e., from the same blocks of trials) assessed with this methodology. Recall that the stopping rule refers to when the system allows itself to stop, as a function of the accumulated information. For instance, in the standard Sternberg paradigm, on trials when no targets appear (i.e., all the stimuli are distractors), a correct "no" response can be made only after exhaustive processing on all the distractors has taken place. This situation is equivalent to an AND condition on the distractors. On the other hand, when a single target is present among $n - 1$ distractors, in principle the system can cease when the target is processed, thus implying "self-terminating" processing.

If all the items on a trial are targets, a special kind of self-termination called "first termination" (Colonius & Vorberg, 1994), "minimum processing time processing" (Townsend & Nozawa, 1995), or simply a "race" (Raab, 1962) is possible. The latter type of paradigm constitutes an OR condition. Of course, in any of these cases, the system might nevertheless employ the exhaustive processing rule wherein all items are processed regardless of when the target is found and processed. Sternberg (1966, 1975) advocated the latter possibility even in circumstances where self-termination would be possible. The fact that the stopping rule is also assessable within the same blocks of trials that are employed to determine architecture and capacity is a marked asset (Townsend & Nozawa, 1995). Since we will focus on exhaustive processing, guaranteed by observing behavior on target-absent trials, we can confine the parallel and serial predictions for $S_{IC}(t)$ to that type of stopping rule.

The predictions relevant to this study are shown in Figure 1. They are based on exponential distributions, but the general qualitative form of the functions does not depend on particular distributions. Note that whereas ex-

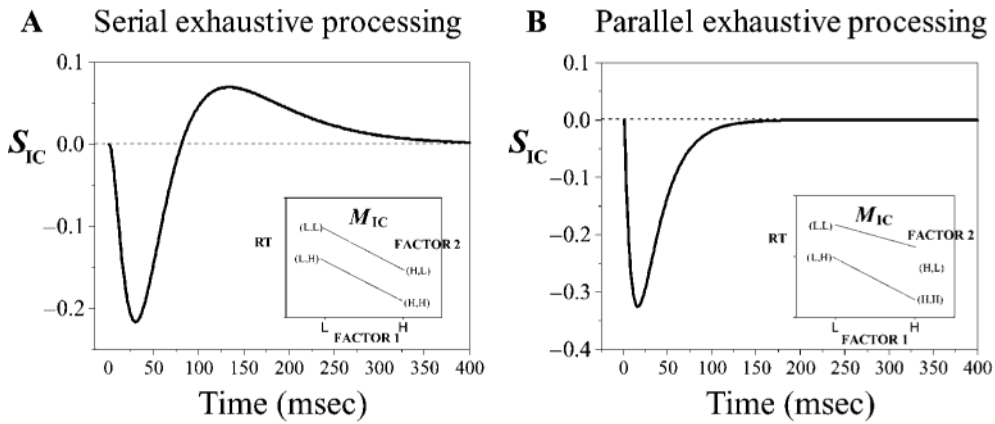


Figure 1. Predictions of survivor interaction contrast functions (S_{IC}) for simple serial exhaustive (A) and parallel exhaustive (B) processing models. Corresponding mean interaction contrasts (M_{IC}) for each model are presented in the lower right corner of each figure.

haustive serial processing predicts $M_{IC} = 0$, contrastingly, in the case of $S_{IC}(t)$, we learn that exhaustive serial processing reveals an interesting S-shaped curve going from negative to positive. It is important to observe that the exact form of the curve is predicted; thus, an S-curve beginning positive and then going negative or some other function would not be compatible with exhaustive serial processing. Note that the predictions of the serial and parallel models were developed before any intuitions about differential behaviors were forthcoming. This is one of the major powers of mathematical theorization.

Nevertheless, an intuition can be gained about the serial survivor contrast curves by the following analysis: To simplify the discussion, assume that the two serial stages with times T_1 and T_2 are governed by the same probability distributions; that is, T_1 and T_2 are random variables with the same underlying probability distribution (i.e., their probability density functions, the cumulative distributions, and the survivor functions are identical). This means that they are identical random variables, although on any one trial, of course, they will typically have different values. In fact, we assume that T_1 and T_2 have the same distributions under both low and high conditions. Letting the symbol “=” signify that two random variables are identical, we then see that $T_{1,i} = T_{2,i}$, where i represents $i = 1$ for the low condition or $i = h$ for the high condition (this assumption just simplifies the presentation and is not at all necessary). Hence, to simplify the notation even further, we can set $T_{1,i} = T_{2,i} = T_i$, dropping the stage index but retaining the factor level index i . However, we still need to write the overall serial RT(i, j) as $T_i + T'_j$ (note the prime on the second T) even if the two levels are the same ($i=j$), because each time is a sample from its own distribution.

Hence, for the low–low condition, the response time random variable $RT(l,l) = T_1 + T'_1$. Likewise, $RT(l,h) = RT(h,l) = T_1 + T'_h$, and finally, $RT(h,h) = T_h + T'_h$. We know, of course, that the mean, or expected value, of

$M_{IC} = RT(l,l) - RT(l,h) - [RT(h,l) - RT(h,h)]$ is zero. Now, it turns out that for short times, the sum of the two stage times is determined more by the longest of the two times than by the shortest. Since $RT(l,l)$ and $RT(l,h)$ both contain a random time that tends to be quite slow, whereas both components of $RT(h,h)$ tend to be fast, the survivor functions (or the cumulative distribution functions for that matter) of $RT(l,l)$ and $RT(l,h)$ are closer than are $RT(l,h)$ and $RT(h,h)$ (see Figure 2A). Just the reverse happens for long durations. The longest times are highly probable only if both the times in the sum tend to be long—that is, if both are distributed as T_l . If either one or two of the times tend to be shorter, exceedingly long times are less likely, meaning that their two survivor functions are closer together.

Our theory also predicted that the negative region must be equal in area to the positive region (Townsend & Nozawa, 1995). This prediction follows from the fact that the integral (i.e., summed area) of a survivor function on time equals the mean processing time. Then, the integral of the double difference that defines the survivor contrast must equal the mean contrast, and hence the negatively signed area must equal the positive signed area. The predicted form of the $S_{IC}(t)$ function is a prime example of where mathematically oriented theoretical work can derive interesting, unexpected, or even seemingly bizarre predictions (from the associated theorems) in the absence of a priori intuition. We had no prior intuition—in fact, we were rather surprised that the survivor function contrast was not predicted by our theory to be identically zero (i.e., zero at every point of the curve).

In contrast to the serial effects, the exhaustive parallel prediction is for an entirely negative $S_{IC}(t)$. Although this prediction too was originally developed in the absence of intuition, it is readily apparent that in exhaustive parallel processing, the overall RT is given by the maximum of the individual item times (an AND gate following separable parallel channel processing produces a “slowest

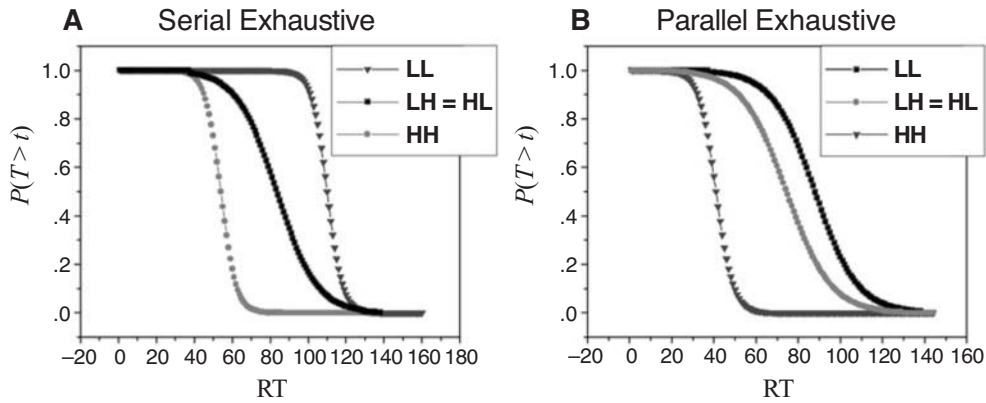


Figure 2. Two demonstrations of survivor functions generated from the serial (A) and parallel (B) exhaustive independent processing models, under selective influence. Each survivor function is based on a joint exponential density distribution of two processing channels where either of them could be in either of two states (low or high). The S shape of the serial exhaustive survivor interaction contrast function S_{IC} (A) is calculated by taking the difference between four survivor functions (on the left); it is obvious that the difference $HH - LH$ is negative and bigger than the difference $HL - LL$ for some short RTs. This relationship tends to reverse for longer RTs. In contrast, the parallel exhaustive S_{IC} difference stays negative for all RTs (B).

horse” result). Thus, the RT is determined by the longest item time. Since on an lh (or hl) trial, the longest time tends to be closer to the longest time on an ll trial, the difference $S_{ll}(t) - S_{lh}(t)$ tends to be smaller than the difference $S_{ll}(t) - S_{hh}(t)$, because the longest time on an hh trial is still determined by a fast-finishing item (see Figure 2B). This fact can aid us in our statistical analyses, as we shall find.

The number of studies employing the expanded methodology at least for M_{IC} has grown over the past two decades. It has been applied successfully to paradigms where a single target or target item is presented first and a multi-item set is presented second—for instance, visually (Egeth & Dagenbach, 1991; Townsend & Nozawa, 1995). The new statistical curve, $S_{IC}(t)$, has begun to be applied (Wenger & Townsend, 2001).

To our knowledge, no factorial investigation to determine parallel versus serial processing has been applied, even using M_{IC} , to the original progenitor, Sternberg’s short-term memory scanning task. The reason is probably that it is much easier to manipulate experimental variables in a visual search task (e.g., brightness, noise level, etc.) than in an internal memory search task. We introduce a method that employs phonemic dissimilarity among items as a method of influencing processing speed of the target-to-memory-item comparisons.

It has been suggested that the interstimulus interval (ISI) that separates the presentation of the memory set and the presentation of the target item might affect processing strategy (parallel vs. serial) (Atkinson, Herrman, & Westcourt, 1974; Forrin & Cunningham, 1973). Short ISIs usually produce a recency effect—that is, faster processing of last presented items. Some researchers have suggested that recency could be due to a processing mechanism different from that providing for the main compar-

ison process: last items are stored in a visual or iconic-like memory store that could be accessed by a fast parallel or direct access mechanism. Older (i.e., earlier) items are, according to this scenario, stored in the short-term memory store and are likely, according to this scenario, to be serially processed. In his original study, Sternberg (1966) used a longer exposure time (1,200 msec) for each item and a longer ISI (2,000 msec); therefore, under this interpretation, perhaps the rapid parallel stage is absent and the short-term store is accessed in a serial fashion. We ran two conditions, one with the ISI equal to 700 msec and the other with the ISI equal to 2,000 msec, expecting that the former condition might be more likely to lead to parallel processing than the latter.

The experimental variation of distractor–probe phonemic similarity is by itself not a new manipulation (Dick & Hochhaus, 1975; Huesmann & Wocher, 1976). However, to our knowledge, the present strategy is novel in that it has not been employed heretofore in parallel–serial factorial methodologies. The introduction of distractors with varying similarity to the target, while straightforward to analyze in target-absent data, is more methodologically complex on target-present trials, due to possible interactions of the similarity manipulation with the present target. Hence, we confine the present analyses and tests to the target-absent data.

We note prior to the Method section that we have continued the convention (e.g., Townsend & Ashby, 1983) of indicating the level of a factor as “high” which speeds up the subprocess and as “low” when (relatively speaking) it slows down the subprocess. This convention implies that high dissimilarity will be designated as a high level of the factor and low dissimilarity (i.e., high similarity) will be designated as a low level of the factor.

METHOD

Participants

Five participants (2 female and 3 male) were paid for their participation.

Materials

The experiment was carried out in Belgrade, using stimuli crafted from Serbian linguistic features. In the Serbian language, one letter represents one sound, so grapheme-to-phoneme correspondence is straightforward. Stimuli were pseudowords in consonant-vowel-consonant (CVC) form.² The middle letter (voice) was “A” in all cases. Two sets of mutually confusable consonants were used: fricatives (F,S,V) and semivocals (L,M,N). Letters from each separate group were used to design one pseudoword. For example, six possible combinations were formed from one group of letters (e.g., for semivocals, MAL, LAM, NAL, LAN, MAN, and NAM; for fricatives, FAS, SAF, SAV, VAS, FAV, and VAF). A total of 12 stimuli were constructed from both groups of letters. In order to produce factorially combined dissimilarity/item conditions (item being designated by position of presentation), target items and memory-set items were drawn from both sets. Examples of high–high, high–low, low–high, and low–low are shown in Table 1. Each pseudoword appeared with the same probability in each position of the memory set and as a target.

Design and Procedure

There were two types of trials, one half in which the target was present in the memory set and one half from which it was absent, randomly arranged. In the use of factorial methodology, the factors were manipulated for a constant set size, $n = 2$. Participants signified their answers as “yes” with one index finger or “no” with the other. On target-absent trials, neither, either one, or both of the items in a memorized set (distractors) could be of high or low dissimilarity to the target. On the basis of unrelated literature on acoustic similarity of auditory linguistic features, we expected high-dissimilar distractors to be rejected faster than low-dissimilar distractors, and this was verified in pilot experiments. The target-absent data exhibit particularly clear evidence concerning processing strategies, so we focus on those. As noted, we also manipulated the ISI between memorized set and a target (700 and 2,000 msec).

Each trial consisted of a fixation point and warning low-pitch tone for 1 sec, successive presentation of two items in the memory

set for 1,200 msec, an ISI, and a target. The ISI period started with a fixation point and a second warning high-pitch tone that lasted for 700 msec. Onset of this second warning signal was activated so that its end coincided with the end of the ISI period. Thus, for short ISI = 700 msec, the ISI started with both fixation point and second warning tone, and for ISI = 2,000 msec, the warning tone was activated at 1,300 msec from the beginning of the ISI period. Participants were run for 44 blocks of 128 trials each. Each block was divided into 6 subblocks of 20 trials (except the last one, which had 28 trials). The participants were requested to achieve very high accuracy, and usually only 1 block was completed on a particular test day. Thus, each mean RT in a specific ISI condition and particular factorial combination represented 300–400 trials per participant (depending on duration of participation). Brief rest periods were allowed every 24 trials.

RESULTS AND DISCUSSION

Averaged percentages of errors over all participants for all target-absent factorial conditions (hh, hl, lh, ll) was less than 1% (0.81%). Out of 5 participants \times 4 factorial conditions, errors greater than 2% were found in only three factorial conditions, all from Participant 5, with 5.97% being the largest. No evidence of speed–accuracy tradeoff was found. These figures are well within the usual standards for concentration on RT.

Our major focus is on the set of $S_{IC}(t)$ curves, owing to their powerful qualitative features, which are distribution free. Nevertheless, traditional analyses of variance (ANOVAs) are useful too; they can affirm or disconfirm interaction in the predicted direction. Recall that the integral of the $S_{IC}(t)$ curve equals M_{IC} . Since exhaustive parallelism predicts an $S_{IC}(t)$ that is always negative, we can expect a negative interaction on the means. Similarly, the integral of a serial $S_{IC}(t)$ function must be 0, the same as the standard result. It will be helpful therefore, to discuss the results for the M_{IC} and $S_{IC}(t)$ together, for individual participants. However, it is worth observing that if we had averaged individual data, the ANOVA results would have completely obscured the individual differences and the conclusion would have been additivity and therefore seriality, as verified by express analyses. As we shall see, this is a far different conclusion from what we found when we assessed individual patterns.

The S_{IC} curves appear in Figure 3A, and the M_{IC} graphs in Figure 3B. Table 2 exhibits the ANOVA results. We used the general linear model to test F statistics for the following fixed factors: position of an item in a set (first, second), phonemic dissimilarity of item to the target (high, low), and order of blocks of trials. Recall that each participant completed around 40 blocks, approximately 20 for each ISI level. The variable order of blocks was included in the analysis in order to account for any variability that comes from learning effects during the experiment. Analysis was performed for each participant on the separate ISIs. The S_{IC} points represent the data; the continuous curves represent theoretical predictions also discussed below. To calculate the respective survivor functions, a bin size of 10 msec was used. The re-

Table 1

Several Examples of Target-Absent (“No”) Experimental Trials for Each Factorial Combination of Item Position (First and Second) and Item-to-Target Dissimilarity (Low and High)

Factorial Combination of Position and Phonemic Dissimilarity	Memorized Set	Target
LL	MAL, NAM	NAL
	FAS, SAV	VAF
HL	NAM, SAV	VAS
	FAS, NAM	LAM
LH	NAL, FAS	NAM
	VAS, MAL	FAV
HH	VAS, FAV	NAL
	MAN, NAL	FAS

Note—The memorized set consists of two items, both pseudowords. In Condition LL, both first and second pseudowords in the memorized set possessed low dissimilarity to the target (H, high dissimilarity). Each factorial condition could be formed from different phonemic groups (nasals and fricatives). Duration between the memorized set and the target was manipulated (ISI = 700 or 2,000 msec) in all experimental conditions.

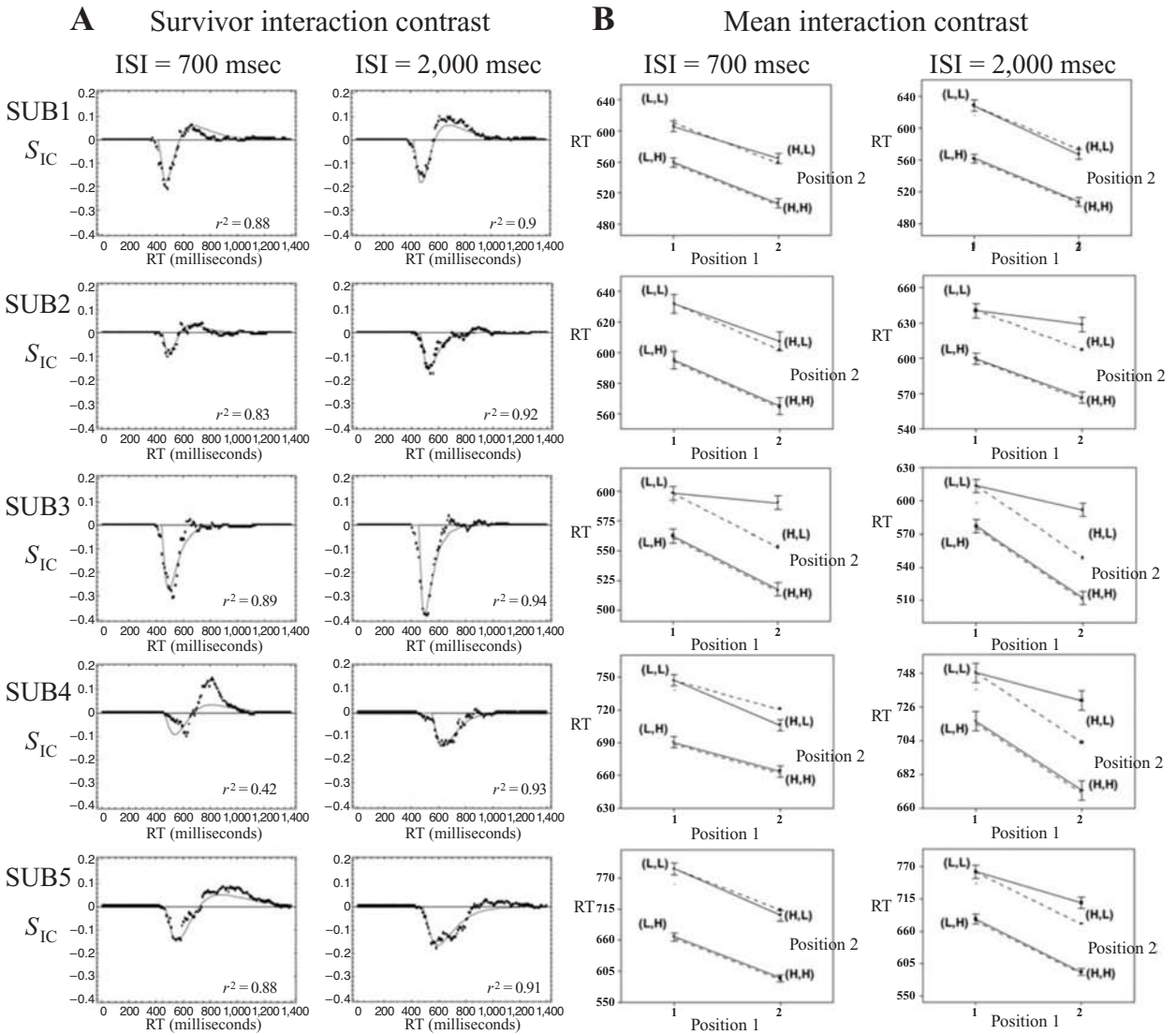


Figure 3. (A) Data obtained for each participant are presented across rows. Survivor interaction contrasts are presented in the first two columns, for each level of ISI (700 and 2,000 msec). Each obtained survivor interaction contrast (dotted line) is presented together with its corresponding model fit (solid line). The r^2 statistic is shown in lower right corner for each model. (B) Corresponding mean interaction contrasts are presented in the last two columns for each level of ISI (700 and 2,000 msec).

sults differ across participants, revealing some apparent individual differences; ISI effects were present as well.

Participant 1's S_{IC} curves as well as the ANOVA strikingly suggest seriality in both ISI conditions. That is, the two S_{IC} curves follow the negative-to-positive S shape, whereas the M_{IC} values are not close to significance at any reasonable level.

Participant 3 revealed overwhelmingly negative functions for both ISIs. Thus, Participant 3 evidenced parallelism in both ISI conditions. This conclusion is backed up by the ANOVA, as can be seen in Table 2, with significant M_{IC} results in the case of short as well as long ISIs.

Contrastingly, Participants 2, 4, and 5 manifest a transition from approximately equal positive and negative areas for the short ISI in the prototypical serial fashion to preponderantly negative areas for the long ISI, again suggesting exhaustive parallelism in the latter condition. In fact, Participants 2 and 5 revealed very large p values in the short ISI condition, and significance at the $\alpha = .05$ level or better in the long ISI condition.

Participant 4's short-ISI curve appears somewhat more positive than negative while basically exhibiting the serial type of profile. The ANOVA suggests that the excess positive part of the curve may be a little worrisome, since p is near statistical significance at the .10 level. However, with

Table 2
General Linear Model F Statistics of Interaction
(Test of Parallelism \times Seriality),
Together With Mean Interaction Contrast (M_{IC})

Participant	df_1	df_2	F	p	M_{IC}	Power
Short ISI (700 msec)						
1	1	291	.83	.362	-11	0.15
2	1	545	.297	.586	-5	0.09
3	1	130	11.42	<.01	-37	0.92
4	1	310	2.53	.11	+15	0.35
5	1	351	.472	.492	-10	0.10
Long ISI (2,000 msec)						
1	1	295	0.317	.345	+7	0.09
2	1	606	4.13	<.05	-21	0.53
3	1	129	15.812	<.01	-43	0.98
4	1	303	8.26	<.01	-27	0.82
5	1	324	4.4	<.05	-36	0.56

the high power afforded by the large number of trials run, we are willing to tentatively accept the serial inference at this point in time, especially since the parallel prediction goes in the opposite direction. This individual's long-ISI curve is negative, indicating parallel processing, which is strongly supported by the ANOVA results. Therefore, we provisionally group Participant 4 together with Participants 2 and 5 in basically revealing a transition from serial to parallel in moving from the short-ISI to the long-ISI condition.

Finally, some very simple models (e.g., Townsend & Ashby, 1983, p. 50), based on exponential processing times at the individual item level, were fit to the individual data using the nonlinear regression package in Mathematica 3.0. To find the best least-square fits, we used the Levenberg-Marquardt search method, which is considered to be one of the standard methods for fitting nonlinear models. In addition, constant residual RT components were estimated (i.e., it was assumed that the additional time consumed by nonsearch processes was a constant for each individual and ISI condition). Note that the residual time parameter was fixed across factor levels.

Thus, for each participant and each ISI condition, there were three free parameters: one exponential rate parameter, whether serial or parallel, for the high and for the low factor conditions, plus the residual time component. The smooth curves in the S_{IC} graphs in Figure 3A depict these fits. The r^2 for Participant 4 in the short-ISI condition was fairly low, and we recall that this individual's positive area was a bit larger than expected on the basis of true serial processing. Nevertheless, given the extreme simplicity of these models, the fits appear remarkably close, in terms of the r^2 and the appearances as well. Furthermore, the fits of the appropriate models were always substantially better than those for the inappropriate models.

CONCLUSIONS

We employed new factorial methodology in order to assess processing architecture in a classical short-term memory search task. Our focus was on target-absent tri-

als, which naturally force exhaustive processing. According to the serial exhaustive prediction, we expected zero M_{IC} and negative-to-positive S-shaped S_{IC} curves. It can hardly be overemphasized that these strong predictions hold for the vast class of *all* serial models, not just the standard serial models that predict equal mean processing times on the various items (e.g., Sternberg, 1966; Townsend & Ashby, 1983) or those based on specific stochastic processes, such as the gamma or ex-Gaussian distributions, and so forth.

In dramatic contrast, parallel exhaustive models predict a negative M_{IC} and a completely negative S_{IC} function. Again, all parallel models with exhaustive processing must make this prediction. Indeed, we found that all *significant* M_{IC} values were negative and associated with the survivor contrasts decisively supporting parallelism. Where M_{IC} suggested additivity, the associated S_{IC} obeyed the profile of seriality as predicted.

Three participants (2, 4, and 5) revealed a striking trend from seriality to parallelism in moving from the short- to the long-ISI condition. One participant (1) strongly adhered to a serial profile in both ISI conditions; another participant (3) exhibited parallel-like RT evidence in both ISI conditions.

Recall that earlier we pointed out that some investigators have suggested that short ISIs may be more likely to encourage parallel processing, whereas long ISIs may result in serial processing (Atkinson et al., 1974; Forrin & Cunningham, 1973). If anything, the reverse was preponderant in our data (serial-to-parallel in moving from the short to the long ISI). Nonetheless, we are in no position to draw strong inferences concerning the form of individual differences or, for that matter, strong conclusions concerning the effects of the ISI. Our findings in this domain must be regarded as extremely tentative, if rather provocative.

A bedrock assumption underlying all the factorial methodologies is that of selective influence, as mentioned in the introduction. Although to test this assumption in an absolute fashion is not trivial, Schweickert and Giorgini (1999) and Townsend and Thomas (1994) have discussed its viability, with most evidence pointing to its support. In addition, it seems fair to question whether and how failures of selective influence could promote the fine-grained profiles predicted by the two opposed serial versus parallel architectures. Indeed, we view the new survivor contrast measures combined with the traditional mean interaction contrast, plus present and other recent applications (e.g., Townsend & Nozawa, 1995; Wenger & Townsend, 2001) as providing cogent support for the methodologies.

A related issue is whether a particular factorial manipulation might alter the very processing mechanisms that the researcher is investigating. Needless to say, this is a common obstacle in psychological domains, and in fact, to varying degrees, in any scientific program. More experimentation with and without the factorial strategies will at least be able to show whether typical aspects are similar—for instance, the approximation to linearity

with set size. At this point, we can emphasize only the fact that in this specific paradigm, we now appear to possess clear evidence for either serial or parallel processing, depending on the participant and experimental condition. And, the results are startlingly different from our findings so far with visual patterns such as facial features, where highly consistent evidence for parallelism has arisen (e.g., Wenger & Townsend, 2001).

Questions arise concerning memory set size. First, it is worth noting that even for our small set size value of two, we have found evidence of serial processing. We must confess that we will be surprised if we should discover more evidence for parallel processing with larger set sizes, but that remains to be seen. An even more important matter concerns the viability of our factorial methodology with large set sizes. In principle, of course, the methodology should generalize. At present, however, we must be cautious. We have obtained some data for a set size of four, using our distractor-similarity factorial manipulation. Some theoretical work in conjunction with beginning analyses of the data suggest that there may be several different approaches to performing tests, some of them requiring more assumptions than others. Some of the tests appear to rest on assumptions that are stronger than what the data allow, but others are more promising. We are presently laboring on these fronts.

Whereas Sternberg's original conclusions and subsequent research (e.g., Roberts & Sternberg, 1993) favored serial processing or at least were compatible with it (e.g., Ashby & Townsend, 1980; Snodgrass & Townsend, 1980, the latter in visual pattern matching), a number of studies have yielded evidence consistent with parallelism (Ratcliff, 1978; Townsend & Ashby, 1983). Our results suggest that different individuals, and even the same individuals in distinct conditions, may evince either parallel or serial processing; again, these results must be accompanied by caveats about the exact nature of individual differences in relation to ISI at present. Nevertheless, these results, plus the fact that averaging individual data in our study obscured the sizeable individual differences, supplement those of other recent studies calling into question the usual strategy of averaging across participants (Ashby, Maddox, & Lee, 1994; Estes, 1956).

Interesting research remains to be done: attempts to pin down when and why distinct strategies occur, including attempts to replicate our basic individual difference/ISI results. Another interesting issue is whether the parallel or serial strategy can be brought under voluntary control. More secure linkages with the sizeable literature on visual search would also be an appropriate goal. In addition, it would be helpful to attempt to provide an overall perspective that includes both RT and accuracy and relates to other approaches such as that of McElree and Doshier (1989). Finally, we look forward to allying our procedures with neurocognitive strategies (Chelazzi, 1999; Rahman & Sommer, 2002) and neural computational models (Luck, 1994) in order to provide even stronger tests and a physiologically sensible mathematical theory.

REFERENCES

- ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*, 144-151.
- ASHBY, F. G., TEIN, J.-Y., & BALAKRISHNAN, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, *37*, 526-555.
- ASHBY, F. G., & TOWNSEND, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, *21*, 93-123.
- ATKINSON, R. C., HERRMAN, D. J., & WESTCOURT, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 101-146). Hillsdale, NJ: Erlbaum.
- ATKINSON, R. C., HOLMGREN, J. E., & JUOLA, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, *6*, 321-326.
- CHELAZZI, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research/Psychologische Forschung*, *62*, 195-219.
- COLONIUS, H., & VÖRBERG, D. (1994). Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, *38*, 35-58.
- DICK, R. B., & HOCHHAUS, L. (1975). Memory search as a function of phonological context. *Bulletin of the Psychonomic Society*, *5*, 256-258.
- EGETH, H., & DAGENBACH, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 551-560.
- EGETH, H., JONIDES, J., & WALL, S. (1972). Parallel processing of multi-element displays. *Cognitive Psychology*, *3*, 674-698.
- ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*, 134-140.
- FORRIN, B., & CUNNINGHAM, K. (1973). Recognition time and serial position of probed item in short-term memory. *Journal of Experimental Psychology*, *99*, 272-279.
- HUESMANN, L. R., & WOOCHER, F. D. (1976). Probe similarity and recognition of set membership: A parallel-processing serial-feature-matching model. *Cognitive Psychology*, *8*, 124-162.
- LUCK, J. S. (1994). Cognitive and neural mechanisms of visual search. *Current Opinion in Neurobiology*, *4*, 183-188.
- MCELRREE, B., & DOSHER, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*, 346-373.
- MURDOCK, B. B., JR. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, *10*, 289-291.
- NIELSEN, G. D., & SMITH, E. E. (1973). Imaginal and verbal representations in short-term recognition of visual forms. *Journal of Experimental Psychology*, *101*, 375-378.
- RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574-590.
- RAHMAN, R. A., & SOMMER, W. (2002). *Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing*. Manuscript submitted for publication.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.
- ROBERTS, S., & STERNBERG, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611-653). Cambridge, MA: MIT Press.
- SCHNEIDER, W., & SHIFFRIN, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66.
- SCHWEICKERT, R. (1978). A critical path generalization of the additive factor methods analysis of a Stroop task. *Journal of Mathematical Psychology*, *26*, 105-139.
- SCHWEICKERT, R., & GIORGINI, M. (1999). Response time distributions: Some simple effects of factors selectively influencing mental processes. *Psychonomic Bulletin & Review*, *6*, 269-288.

- SCHWEICKERT, R., & TOWNSEND, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology*, **33**, 328-347.
- SNODGRASS, J. G., & TOWNSEND, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 330-354.
- STERNBERG, S. (1966). High-speed scanning in human memory. *Science*, **153**, 652-654.
- STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, **30**, 276-315.
- STERNBERG, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, **27**, 1-32.
- SWANSON, J. M., & BRIGGS, G. E. (1969). Information processing as a function of speed versus accuracy. *Journal of Experimental Psychology*, **81**, 223-229.
- TOWNSEND, J. T. (1969). *Mock parallel and serial models and experimental detection of these*. Paper presented at the Purdue Centennial Symposium on Information Processing.
- TOWNSEND, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, **10**, 161-163.
- TOWNSEND, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133-168). Hillsdale, NJ: Erlbaum.
- TOWNSEND, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, **1**, 46-54.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TOWNSEND, J. T., & NOZAWA, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, **39**, 321-359.
- TOWNSEND, J. T., & SCHWEICKERT, R. (1985). Interactive effects of factors prolonging processes in latent mental networks. In G. d'Ydewalle (Ed.), *Cognition, information processing, and motivation: XXIII International Congress of Psychology* (Vol. 3, pp. 255-276). Amsterdam: Elsevier, North-Holland.
- TOWNSEND, J. T., & THOMAS, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, **38**, 1-34.
- VAN ZANDT, T., & TOWNSEND, J. T. (1993). Self-terminating versus exhaustive processes in rapid visual and memory search: An evaluative review. *Perception & Psychophysics*, **53**, 563-580.
- WENGER, M. J., & TOWNSEND, J. T. (2001). Faces as Gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 229-284). Mahwah, NJ: Erlbaum.

NOTES

1. However, in certain approaches, accuracy has been varied on purpose, and we contend that the type of processing is the same for low and high accuracy (McElree & Doshier, 1989).
2. In fact four of the stimulus words have meaning in the Serbian language (LAN, NAM and SAV, and VAS) and are not considered to be pseudowords. However, possible systematic confounding effects were prevented by including these words in each experimental condition as members of a memorized set and as targets, with the same probability.

(Manuscript received February 3, 2003;
revision accepted for publication October 13, 2003.)