Attention Gating in Short-Term Visual Memory

Adam Reeves
Northeastern University

George Sperling
New York University

Subjects first detected a target embedded in a stream of letters presented at the left of fixation and then, as quickly as possible, shifted their attention to a stream of numerals at the right of fixation. They attempted to report, in order, the four earliest occurring numerals after the target. Numerals appeared at rates of 4.6, 6.9, 9.2, and 13.4/s. Scaling analyses were made of (a) item scores, $P_i(r)$, the probability of a numeral from stimulus position i appearing in response position r, $r = (1, 2, 3, 4)$, and (b) order scores, P_{iBj}, the probability that a numeral from stimulus position i appeared earlier in the response than one from stimulus position j. For all subjects, targets, and numeral rates, the relative position of numerals in the response sequence showed clustering, disorder, and folding. Reported numerals tended to cluster around a stimulus position 400 ms after the target. The numerals were reported in an apparently haphazard order—at high numeral rates, inverted iBj pairs were as frequent as correct pairs. The actual order of report resulted from a mixture of correctly ordered numerals with numerals ordered in the direction opposite to their order of presentation (folding around the cluster center). These results are quantitatively described by a strength theory of order (precedence) and are efficiently predicted by a computational attention gating model (AGM). The AGM makes quantitatively correct predictions of over 500 values of $P_i(r)$, P_{iBj} in 12 conditions with two attention and three to six detection parameters estimated for each subject. The AGM may be derived from a more general attention model that assumes (a) after detection of the target an attention gate opens briefly (with a bell-shaped time course) to allow numerals to enter a visual short-term memory, and (b) subsequent order of report depends on both item strength (how wide the gate was open during the numeral's entry) and on order information (item strength times cumulative strength of prior numerals).

When an observer receives information from two or more distinct sources at once and is unable to process all of them, the observer may allocate processing capacity first to one source and then to another. We term such a transfer of processing capacity a *shift of attention,* although we do not imply that conscious awareness of the shift must occur. A classical example concerns a listener at a cocktail party who attempts to listen simultaneously to two different conversations. If the listener is unable to process both conversations at once, the listener may pay attention first to one conversation and then shift attention to the other (Broadbent, 1958; Cherry, 1953). Our present research concerns an observer's ability to shift focal attention (Kahneman, 1973) between two sources of visual input.

In studying visual attention, we used the *RSVP attention shift paradigm* (Sperling & Reeves, 1976, 1978, 1980). In this procedure the subject shifts attention between one source that pro-

duces a stream of letters and a second that produces a stream of numerals. In each stream, characters fall one on top of the next in what is known as *rapid serial visual presentation* (RSVP; Potter & Levy, 1969; Sperling, 1970). The stream of letters is located left of fixation, and the stream of numerals right of fixation. The subject shifts attention between streams without eye movement (Sperling & Reeves, 1980).

The RSVP attention shift paradigm combines RSVP with the requirements to first maintain attention away from the point of fixation, and then to move attention without making an eye movement. We first evaluate these attentional requirements, and then the suitability of RSVP for studying attention, before presenting our new findings[1] on the dynamics of attention shifts and their effect on the contents of visual short-term memory (VSTM; Scarborough, 1972).

Directing Attention to the Periphery

It is well known that observers can fixate on one location and direct attention peripherally. Early writers relied on introspection (Helmholtz, 1909; James, 1890; Wundt, 1912, p. 120). Current research uses two indicators of attention: accuracy of target detection (e.g., Beck & Ambler, 1973; Grindley & Townsend, 1968; Posner, Snyder, & Davidson, 1980; Remington, 1980; Shaw & Shaw, 1977) and speed of detection (e.g., Jonides, 1981, 1983; Posner, 1980; Posner, Nissen, & Ogden, 1978; see reviews by

[1] Preliminary statements of the theory were presented in Reeves and Sperling (1983, 1984) and Sperling and Reeves (1977, 1983).

Sperling, 1984; Sperling & Dosher, in press). These authors have argued that attention can be successfully directed and maintained peripherally.

Dynamics of Attention Shifts

Many studies have attempted to demonstrate the dynamics of a shift of attention to a peripheral target. In the most common procedure, attention is directed to the peripheral target by a preparatory cue. Performance improvements with prepared stimuli are taken to reveal an effect of selective attention. To improve performance on a visual target, the preparatory cue must occur earlier; the shortest facilitatory cue-to-target delay has been used to infer the time taken to shift attention (e.g., Eriksen & Collins, 1969; Eriksen & Hoffman, 1972).

Cued-target experiments have apparently shown that attention shifts may be independent of eye movements. Using a cost–benefit paradigm (reviewed in Posner, 1980), Shulman, Remington, and McLean (1979) found that a valid cue presented 150 ms or more before a target speeded reaction time more than did an invalid cue and argued that attention could shift over the visual field without eye movements. Tsal (1983) found that the benefit of a preparatory cue asymptoted at different cue-to-target delays depending on eccentricity and concluded that attention shifts to the near periphery at a rate of 8 deg/s. Again, Tsal's subjects did not move their eyes.

In an extension to the cued-target paradigm, eye movements are actually made, or prepared. Remington (1980) varied the time from the cue to the target and found that the (small) improvement in hit rate consequent on the cue occurred independently of the time course of saccadic suppression, consistent with the idea that attention and eye shifts are independent. Klein (1980) demonstrated that attention shifts (indexed by reaction times) could be independent not only of overt eye movements but also of oculomotor preparation.

The cued-target studies appear to measure the dynamics of attention shifts that occur without corresponding shifts of the eyes. However, there are two problems in deriving estimates of attention shifts from experiments that use single, briefly flashed targets. One potentially solvable problem is that the dependent measures used in the preceding studies are either reaction time or accuracy; neither can be uniquely related to processing efficiency unless some measure of the speed–accuracy tradeoff is available (see Sperling, 1984, and Weichselgartner, Sperling, & Reeves, 1985a, 1985b, for detailed discussions). A subtler but more serious problem is that the cued-target procedure cannot disentangle (a) the time course of the attention shift from (b) the persistence of the target in the visual display or in visual memory. Attention may improve processing of the target at any time from target onset to the final disappearance of the target representation. Only if it is assumed, usually incorrectly, that visual persistence is very brief is it obvious at which instant attention is exerting its effect. This means that control over target persistence in memory is necessary in order to infer attentional dynamics from attention shift experiments.

Rapid Serial Visual Presentation

One way to control visual persistence is provided by RSVP, in which a stream of stimuli succeed each other at the same spatial location, each overwriting the immediately preceding one (Potter & Levy, 1969; Sperling, Budiansky, Spivak, & Johnson, 1971). Because it controls the time during which information is visually available, RSVP, when combined with a shift of attention, provides a superior way of disentangling attention from visual memory. (In addition, both accuracy and latency are measured on each trial.) We call the combined method the *RSVP attention shift paradigm* (Sperling & Reeves, 1980). In that study, we used the procedure to estimate the time taken to shift attention (the attention reaction time; ART). In the present article, we analyze the same data (from Reeves, 1977) to study the effect of shifting attention *to* a stream of stimuli (the numerals) on the observer's visual memory *of* them.

Earlier work with RSVP had shown that memory for order of events in this paradigm can be remarkably poor. Norman (1967) observed that in a list of numerals presented in RSVP (with interleaved masking fields),
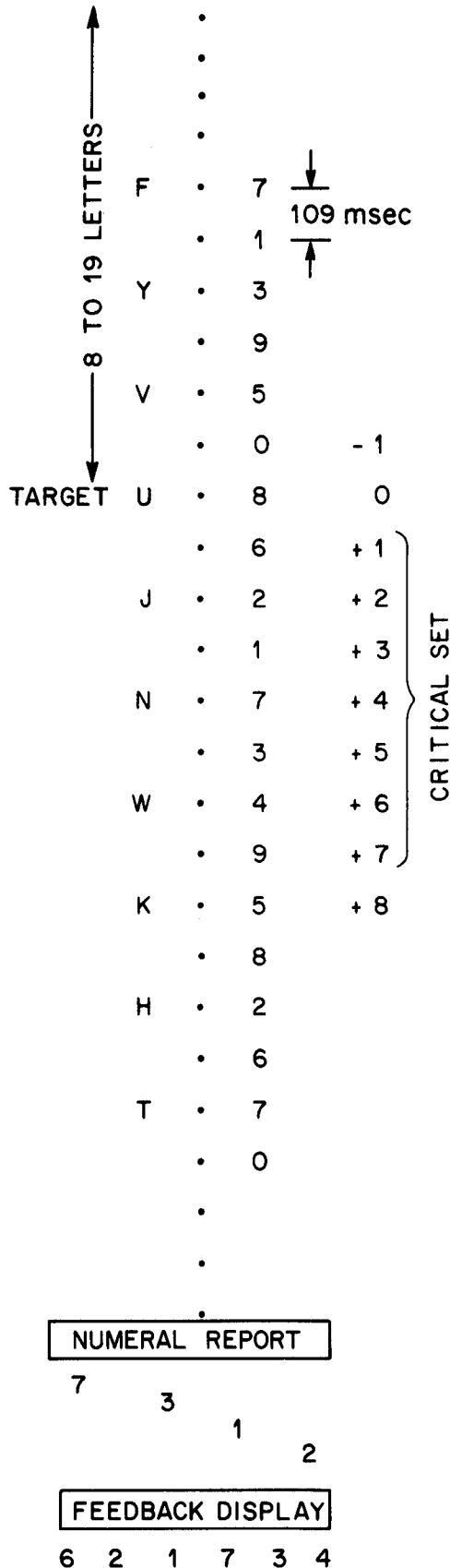
> Subjects report that they can clearly recall the last two items they were shown, but they have absolutely no idea of the order in which they were presented. This observation is often noted (but never studied) in the few memory experiments which have used such rapid presentation rates. (p. 295)

Scarborough and Sternberg (1967) and Sternberg and Scarborough (1969) found that when subjects monitored a stream of numerals, presented at 13.3 numerals/s, they could nearly always tell whether a target digit had been presented in the stream but were at chance in telling which digit followed the target. The authors noted their subjects' poor acquisition of order information but failed to offer a detailed description, theory, or explanation of their subjects' difficulties. In a somewhat similar experiment, Lawrence (1971) presented a stream of words in which the target word was capitalized. At fast presentation rates of 16 to 20 words/s, 30%–40% of the targets were reported incorrectly. Of these errors, a full 82% consisted of the word immediately following the target word (excluding those trials in which the target was the first word or was among the last four words in the stream). This is a systematic distortion of order information, quite like some of the effects we report here. But Lawrence's data are not sufficiently rich to enable us to determine the overall deficit—how the presented list might be represented in the subject's memory.

In the current work, we find that an attention shift to a stream of numerals, presented in RSVP mode, produces not a total loss, but rather a systematic distortion of order. The nature of the systematic disorder, an order illusion, provides the main theme of the study. Our explanation is, briefly, that the perceived order of rapidly presented items in short-term visual memory is determined primarily by the amount of attention they receive at the time of input.

Overview

We first describe the attention shift procedure (Sperling & Reeves, 1980) and then present results (item and order scores) for various targets and for four different numeral rates. The chief result, in harmony with Sternberg and Scarborough (1969), is that although item information is good, overall order information is poor. The Results section provides a detailed analysis of pair-

```
  ↑
  |
  |
ω  |        •
ER |        •
T  |        •
T  |        •
E  |        •
L  F    •   7   ↓
   |        •       109 msec
ω  |        •   1   ─
9  Y    •   3       ↑
1  |        •
O  |        •   9
T  |        •
8  V    •   5
   |        •
   |        •   0      - 1
   ↓
TARGET U   •   8       0

           •   6      + 1 ⎫
      J    •   2      + 2 ⎪
           •   1      + 3 ⎪  ⊢
      N    •   7      + 4 ⎬  ω  ω
           •   3      + 5 ⎪  ⊢  ⊢
      W    •   4      + 6 ⎪  ⊢  ⊢
           •   9      + 7 ⎭  ⊂
      K    •   5      + 8

           •   8
      H    •   2
           •   6
      T    •   7
           •   0
           •
           •
           •
┌─────────────────────┐
│  NUMERAL  REPORT     │
└─────────────────────┘
   7
        3
             1
                  2
┌─────────────────────┐
│  FEEDBACK  DISPLAY   │
└─────────────────────┘
   6   2   1   7   3   4
```

wise order scores and shows that order errors are not the result of random responding but rather of a systematic, but incorrect, perceived order. In the Discussion section, we first show that the systematic misorderings do not stem from guessing or from forgetting but instead accurately reflect the contents of visual short-term memory for the numerals. We next develop a strength model that provides a simplified description of all possible pairwise combinations of presented numerals in terms of an underlying scale of precedence values, V_i, different for each position i in the numeral stream and each numeral rate. We then develop a powerful descriptive tool, an *attention gating model* (AGM), which accounts for the precedences in terms of one underlying attention gating function. Finally, we derive the AGM from a general attention gating model (GAGM; Reeves & Sperling, 1984; Sperling & Reeves, 1983) in which item and order information are represented in a psychologically and physiologically plausible fashion.

Method

Task

In the RSVP attention shift paradigm, a subject is instructed to maintain steady eye fixation upon a dot shown on a display screen. (Subjects indeed do maintain fixation; Sperling & Reeves, 1980). Computer-generated streams of characters appear on each side of the fixation dot. To the left of the fixation dot, a steady stream of letters appears, and to the right, a steady stream of numerals. The letters appear one after another in the same location at a rate of 4.6 letters/s. The numerals appear one after another in the second location at various rates. The subjects' task is to monitor the letter stream at left of fixation for a target (a letter C, a letter U, or an outline square), and on detecting the target, to report the first four numerals that he or she can from the numeral stream appearing to right of fixation. Figure 1 illustrates a trial in which the target was the letter U, letters were presented at 4.6 letters/s, and numerals were presented at twice that rate.

Subjects

Three graduate students with normal or corrected-to-normal vision served as subjects. AR (the first author) and GL were thoroughly practiced in this type of experiment; AK was naive. Each subject received at least 10 hours of practice. AR and GL were then run for 40 hours, and AK for 30 hours, in sessions of 1 to 1½ hours on different days.

Stimulus Parameters

Stimuli were presented on a high quality Digital Equipment Corporation VT-11 graphics display unit controlled by a PDP-15 computer, and were

Figure 1. Procedure for a typical trial. (Letter stream, fixation dot, and numeral stream are schematically illustrated. Consecutive rows from top to bottom illustrate consecutive superimposed stimuli. Letters appear at a rate of 4.6/s to the left of fixation, and numerals at a rate of 9.2/s to the right. The critical set of numeral positions [from which the subject is required to select at least the first element of his response] is shown at right. In this example, the subject reports "7, 3, 1, 2" [NUMERAL REPORT]. The data on a trial are the positions of the response items in the critical set, in this case, (4, 5, 3, 2). After his response, the subject is shown the first six elements of the critical set [FEEDBACK DISPLAY]. A perfect report would have consisted of the first four elements of the critical set, correctly ordered.)

viewed binocularly. The display oscilloscope had a fast white phosphor (P4) with a decay time of less than 1 ms. We developed a special set of distinctive characters (Figure 2) that were highly legible at the viewing distance of about 0.68 m. The characters were (a) the numerals 0 through 9, (b) the targets—letters C and U and an outline square (Sq), and (c) the remaining (background) letters. Characters were 1.45 deg (1.72 cm) high and between 0.1 and 1.85 deg wide. They were presented for 3.2 ms at sufficient intensity to appear quite bright (see Sperling & Reeves, 1980, for details). Interstimulus intervals were blank.

The letters were presented at a fixed rate of 4.6 letters/s for subjects AR and GL, and 3.7 for AK. These rates were chosen so that subjects reported they had to pay "full attention" to the letter stream in order to be able to reliably—98% of the time or better—detect the target. In pilot work we found that at faster letter rates, subjects could not always detect the target.

Because the target set never varied, one might have expected accurate performance at even faster letter rates (based on the results of the "consistent mapping" conditions of Schneider & Shiffrin, 1977, and Shiffrin & Schneider, 1977). Unlike their detection task, however, the attention shift procedure does not allow the subject to recover easily from implicit false alarms because several letters, which may include the target, will pass before attention returns to the letter stream. In a detection paradigm, attention remains continuously on the target stream and recovery from an implicit false detection is possible when a much more conspicuous target appears subsequently. The demand for both high accuracy and high certainty forced our subjects to use a very high criterion for identifying the target.

Numerals were presented at rates neither so fast as to produce "blurring" (about 20/s) nor so slow as to allow the subject to implicitly name each numeral as it appeared (about 3/s; Landauer, 1962; Pierce & Karlin, 1957; Sperling, 1963). Numeral rates in various conditions were 4.6, 6.9, 9.2, and 13.4/s for AR and GL and 5.6, 6.9, and 9.2/s for AK. Intervals between numerals were adjusted by up to 3 ms during the trial to ensure that letter and numeral streams were in synchrony. The center to center separation of the letter and numeral streams was fixed at 1.87 deg.

Procedure

Subjects self-initiated each trial, which consisted of the stimulus sequence, a response, and 2 s of a feedback display. A sample test sequence
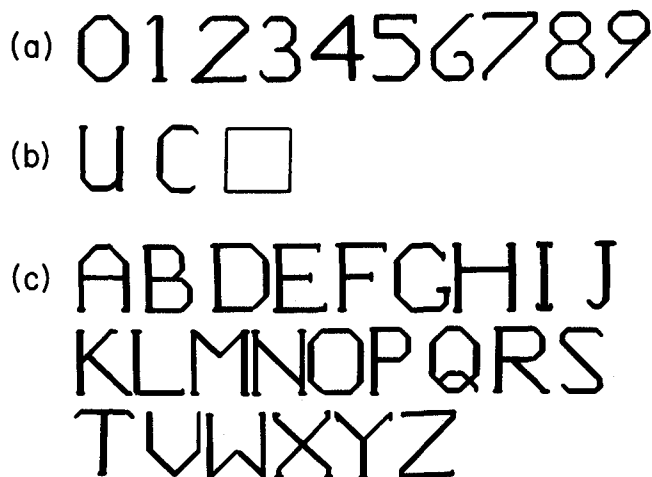
is shown in Figure 1. On each trial a new stream of 24 letters was obtained by randomly permuting the alphabet, excluding the letters C and U. The target was chosen at random from the letter C, the letter U, and the square (Sq). The target was presented at a randomly chosen position strictly between Positions 7 and 20 of the letter stream; the serial position of the target was randomized so that the subject could not anticipate when it would occur.

To describe positions in the numeral stream, we designate the position of the numeral that occurs simultaneously with the target as Position 0. On each trial, a new stream of numerals was constructed by choosing numerals at random, with the restrictions that (a) at least 6 different numerals occurred between any numeral and its next appearance and (b) that a sequence of 10 all-different numerals began at Position 0 in the fast conditions (rates of 6.9 numerals/s or faster) and at Position −1 (the position immediately before the target) in the slower conditions.

On the basis of pilot work, we found the seven earliest, consecutive positions in the numeral stream from which subjects were likely to report numerals. These seven positions are called the *critical set*. Owing to the way the data were collected, only numeral reports from the critical set were available for subsequent analysis. The critical set began at Position −1 in the slowest conditions (4.6 and 5.6 numerals/s) because in pilot work it was found that this was the earliest position from which subjects reported numerals. In faster conditions, only later numerals were reported, so the critical set was progressively delayed. For subjects AR and GL, the critical set began at Positions 0, 1, and 2 for numeral rates of 6.9, 9.2, and 13.4/s, respectively. For subject AK, the critical set began at Positions −1, 1, and 2 for numeral rates of 5.6, 6.9, and 9.2/s. As long as the subject reports numerals from inside the critical set rather than from before or after it, we can uniquely identify the position of the reported numeral in the numeral stream. We argue that this is indeed the case (see Discussion).

Two procedural matters require further comment. Although this experiment deals with attention shifting, subjects were not explicitly instructed to shift attention from the letters to the numerals; they reported that they were forced to do so by the task. Subjects reported that reliable detection of the target at the left of fixation required "full attention" to the letter stream and that the report of the numerals from the numeral stream at the right of fixation required them to "shift attention" from left to right. That is, while our subjects are searching for the target, they have little awareness of the numeral stream. A similar observation was made by Wolford and Morrison (1980), who showed that when their subjects directed attention peripherally (analogous to our subjects' target search), they subsequently were unable to recognize words that had been presented centrally (analogous to our numeral stream).

The choice of a report length of four is a compromise. The longer the required report, the more information it yields, but the greater is the danger of information loss at stages subsequent to the perceptual memory we are attempting to study. With reports of length of four, we were able to show (see Discussion) that subjects virtually never had to fill out a response with a randomly chosen numeral. Thus, four is a conservative choice of report length.

Feedback

After subjects typed their responses, they were shown a feedback display for 2 s, consisting of the first six numerals in the critical set arranged from top to bottom of the display screen. Subjects were instructed to use the feedback to improve performance, both by aiming for earlier presented numerals (those higher on the screen in feedback) and by aiming to improve the accuracy of their report order. Subjects were also instructed to release a reaction time key when they detected the target; these data are discussed in Sperling and Reeves (1980) but are not relevant for the present analyses. Subjects were told to give priority to the numeral report if a conflict should occur between the reaction time task and the numeral report task. However, after initial practice, no such conflict was reported.
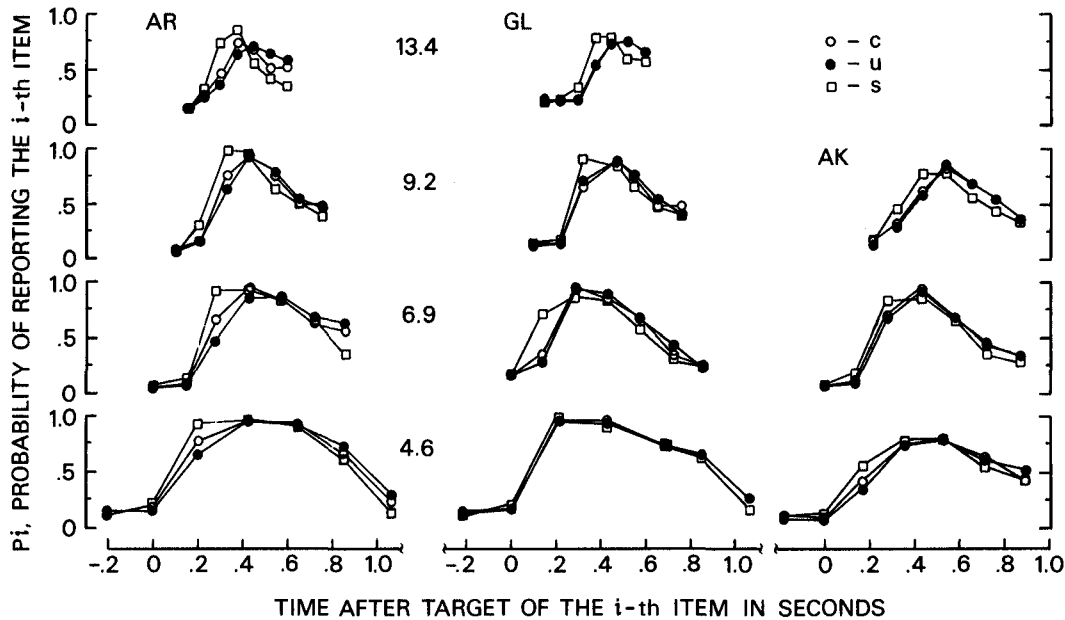
(a) 0 1 2 3 4 5 6 7 8 9

(b) U C □

(c) A B D E F G H I J
K L M N O P Q R S
T V W X Y Z

Figure 2. Photographs of the stimulus characters: (a) numerals; (b) targets (U, C, Sq); and (c) the background letters. (Presented stimuli were 1.45 deg high and 0.1 to 1.85 deg wide.)

Figure 3. Composite item scores for the three subjects, AR, GL, and AK. (The probability P_i of a numeral from Stimulus Position i appearing anywhere in the response is plotted against t_i, the time in seconds from onset of the target to onset of the ith numeral. Panels within a column of the figure represent data from the same subject; panels within a row represent the same numeral rate, as indicated. Targets U, C, and Sq are indicated by filled circles, open circles, and open squares, respectively.)

Design

The conditions reported here were run in a randomized order, counterbalanced over 16 sessions of 300 trials each for AR and GL, and 12 such sessions for AK. Numeral rate was blocked, being changed every 100 trials. Numeral rates were blocked to allow strategies to change between rates if needed to optimize performance. Targets were not blocked but were equally likely on each trial. Targets were randomized to ensure comparability of measures of latency over all other conditions (Sperling & Reeves, 1980). There were about 400 trials in each condition of target and numeral rate, for AR and GL, and about 300 for AK (Reeves, 1977, Table 1, gives exact values). Other conditions, in which the subject reported only one numeral, were also run (Reeves, 1977; Sperling & Reeves, 1980). Subjects averaged four to five trials per min, and rested briefly after every 50 trials.

Results

Item Scores: P_i, $P_i(r)$

The reaction time results are reported elsewhere (Sperling & Reeves, 1980). For the results presented here, motor reaction times were used only to exclude trials: If the subject reacted less than 0.17 s, or more than 1.7 s, after the target was presented, we assumed that the target had not been correctly detected. Overall, 1.6% of trials were excluded thereby.

In analyzing the numeral reports, we are concerned not with the identity of the particular numerals that were reported but only with the positions in the critical set (of numeral stream positions) from whence they came. We collapse over numeral identity in the analysis because all the numerals were reported about equally often (with the exception of numeral 1, which was

reported slightly less often than the others), and because there was no tendency for any particular numeral to be reported earlier in the response than the others (Reeves, 1977).

We are concerned with two ways of scoring numeral position, which we denote for convenience as *item* and *order* scores. The item score for a particular position in the critical set indicates whether numerals from that position appear in the response or not. The order score for a position in the critical set indicates where in the response numerals from that position appear relative to numerals from other positions.

Item Scores: P_i and Clustering

Let P_i be the proportion of trials in which the subject reports (anywhere in the response) a numeral from Position i of the critical set of numeral positions. Let t_i be the time of presentation of the numeral in Position i, measured from the onset of the target. Figure 3 shows P_i as a function of t_i for the full set of targets and rates. The bell-shaped curves of Figure 3 show that subjects most often report numerals from a cluster of positions centered about 400 ms after the target, with a range from 200 to 600 ms (or more, at slow numeral rates). Numeral reports in response to Sq targets occur from slightly earlier positions than for C and U. (The Sq curves are slightly to left of C and U curves in Figure 3.)

Item Scores for Each Response Position, $P_i(r)$

Let $P_i(r)$ be the proportion of trials in which the subject reports a numeral from Stimulus Position i of the numeral stream in
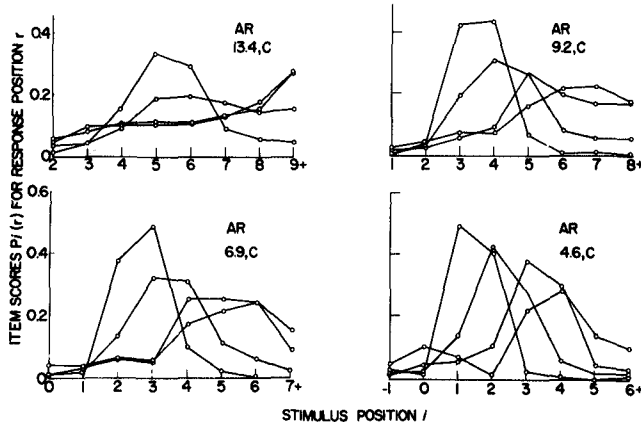
Figure 4. Item scores for each individual response position. (The probability $P_i(r)$ of a numeral from Stimulus Position i appearing in Response Position r is plotted against i. The curve parameter is r. For the leftmost curve, $r = 1$; for the rightmost curve, $r = 4$. Data are for subject AR, target C, at numeral rates of 13.4/s, 9.2/s, 6.9/s, and 4.6/s.

Position r of his response. Because the subject must report a numeral in each response position on each trial,

$$\sum_{i=1}^{10} P_i(r) = 1.0, \quad r = 1, \ldots, 4.$$

The $P_i(r)$ define the item scores for each Response Position r. Figure 4 illustrates $P_i(r)$ versus Stimulus Position i for subject AR, target C, and all four numeral rates. Reports from outside the critical set are marked by a plus sign on the abscissa at the right-hand edge of each plot. Although such reports might have originated from numerals presented before the critical set (e.g., in Position −2), the low probability of report from the first two positions in the critical set makes this highly unlikely and justifies lumping these data in a position after the critical set.

At the fastest numeral rate, the $P_i(1)$ versus i curve is bell-shaped, peaking at Position 5, but the bell-shaped character is gradually lost with later response curves, which become depressed in the center because the numerals in central stimulus positions tend to be reported in earlier response positions. At the slowest rate, the $P_i(r)$ versus i curves are clearly segregated. Although normally one cannot determine order information from item recall, the segregated curves indicate good retention of order.

The pattern of results shown in Figure 4 is typical for other subjects and targets. Whereas the effect of the target was principally to cause the lateral positioning of all the $P_i(r)$ curves to vary together, the effect of numeral rate was more complex. The $P_i(r)$ curves overlap considerably at fast numeral rates but separate at slower rates.

Order Scores: P_{co}, $D(x)$, P_{iBj}

The previous sections on item scores dealt with response items taken one at a time. To analyze report order, we consider response items two at a time, that is, pairs of items. The analysis proceeds from coarse to fine. We consider first the aggregate of all pairs in a response; second, pairs of items separated by a particular distance; and third, pairs containing items from any two stimulus

positions i, j. To begin, we need to define several quantities. Let $i < j$ (read "i earlier than j") denote that Position i is an earlier position in the critical set than is Position j. Let $iBFj$ (read "i before j") mean that, on a particular trial, the subject reports the numeral from Position i before reporting the numeral from Position j. The definition of $iBFj$ refers only to the report order; the actual order of i and j in the numeral stream is irrelevant.

For data analysis, we let $n(E)$ denote the number of trials in which event E occurred. Then $n(iBFj)$ is the number of trials in which the subject reported a numeral from Stimulus Position i before one from Stimulus Position j, with (i, j) in the critical set of numeral positions. We use $n(iBFj)$ in defining several pairwise order scores, which we abbreviate *order scores* because we will not consider higher measures of order than pairs.

Correctly Ordered Response Pairs, P_{co}

A "correctly ordered response pair" means $i < j$ and $iBFj$. That is, the numeral in Position i is both presented before ($i < j$) and reported before ($iBFj$) the numeral in Position j. The proportion of correctly ordered response pairs P_{co} (the subscript *co* stands for correctly ordered) in the total number of response pairs $n(i)$ is

$$P_{co} = \sum_{i,j} n(iBFj \text{ and } i < j)/n(i),$$

where

$$n(i) = \sum_{i,j} n(iBFj).$$

Figure 5 shows P_{co} for each condition and each subject. P_{co} increases from near the chance level of 0.5 at the highest numeral rate (13.4/s) to about 0.8 at the slowest rate. The most surprising result is that P_{co} was only 0.51 at the highest numeral rate, averaging over the two subjects tested at this rate and targets. That is, at 13.4/s, subjects were just about as likely to report a numeral pair in the wrong order as in the correct order!

Order Scores as a Function of Temporal Separation

Disorder, $D(x)$. To discover more about response order than merely that it is weakly related to stimulus order, we consider order scores separately for each pair i, j of stimulus positions.
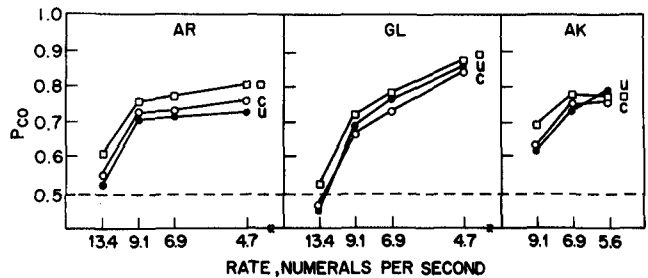

Figure 5. P_{co}, the proportion of correctly ordered response pairs, as a function of numeral rate for each of the three subjects, AR, GL, and AK. (Targets U, C, and Sq are indicated by filled circles, open circles, and open squares, respectively. The horizontal line at $P_{co} = 0.5$ indicates the chance level.)

Diffusion models of processing order information (e.g., Estes, 1972) predict that as positions become further apart, the opportunity for confusing items (numerals) in those positions becomes smaller. The separation between Stimulus Positions i and j, denoted x, is $x = |i - j|$, the absolute value of $i - j$. Disorder $D(x)$ is defined as twice the ratio of the number of inconsistently ordered i, j pairs to the total number of i, j pairs at separation x. Writing $iBFx + i$ for $iBFj$, where $j = i + x$, we have

$$D(x) = \frac{2 \sum_{i=1}^{7-x} \min\{n[iBF(x + i)], n[(x + i)BFi]\}}{\sum_{i=1}^{7-x} \{n[iBF(x + i)] + n[(x + i)BFi]\}}. \quad (1)$$

When there is total confusion in order, so that $n(iBFj) = n(jBFi)$ for all $|i - j| = x$, then $D(x) = 1$. When items are completely consistently ordered in the response, then $D(x) = 0$. In this formulation, consistency rather than correctness of order is counted. In this way the subject is not penalized for any consistent order illusion.

Disorder as a function of stimulus separation. Values of $D(x)$ are plotted as a function of x in Figure 6 for each subject, with numeral rate as the curve parameter. The values have been averaged over targets because the target effect was slight. These data show that $D(x)$ drops as x increases from 1 to 2, levels off, and then rises quite sharply for $x = 5$ and 6 at the slower rate for two of the three subjects. All subjects show U-shaped segments in $D(x)$ for most conditions. These U-shaped curves mean that the order in the response of numerals from well-separated stimulus positions is confused more often than the order of numerals from positions only two apart.

The results of Figure 6 disallow any model that predicts a monotonic decline of $D(x)$ with increasing x. For example, Estes's (1972) "control element" model assumes that (a) loss of order is primary, loss of item is derivative; (b) rate of loss of order is greater the smaller the time intervals between successive items; and (c) transpositions are most likely between adjacent items. Our theory (see Discussion) assumes (a) and predicts (b) but, like our data, disputes (c). Estes (1972) used visual presentation of letters, one on top of the other, as in the present experiments, at a presentation rate approximately equal to the slowest used here and with explicit instructions to subjects to vocally rehearse successive letters. (Rehearsal would have been impossible in the present experiments because subjects were occupied in the detection task until target occurrence.) The conclusion is that differences in procedure between the two experiments caused the to-be-reported items to be stored with different order properties.

Order Scores for Individual Positions, P_{iBj}

Definition. To analyze the order data still further, we computed order scores for individual pairs of positions (i, j) rather than averaging over positions x apart as in computing $D(x)$. We defined a new event, iBj (read "i before j"), which means that on any one trial the subject either (a) reports the numeral from Position i before reporting the numeral from Position j or (b) reports the numeral from Position i and does not report the numeral from Position j at all. Type (a) events alone are included in $n(iBFj)$, which is a simple, convenient measure but which, unfortunately, from a theoretical point of view, is too dependent

on the length of the response (four, in these experiments). If the response were not truncated after four numerals but continued, as for an "ideal subject" with perfect recall, all of the observed type (b) events would become type (a) events, because eventually the numeral in Position j would be reported. The order score (iBj), which includes both types of events, is more appropriate for the strength model developed below in the Discussion section because response limitations are not central to the model. We should note here, however, that when the analysis of order scores is undertaken solely in terms of $iBFj$, that is type (a) events alone, the empirical conclusions drawn here are not altered (Reeves, 1977).

Pairwise order scores are defined as follows. Let P_{iBj} be the proportion of trials on which numerals from Position i are re-
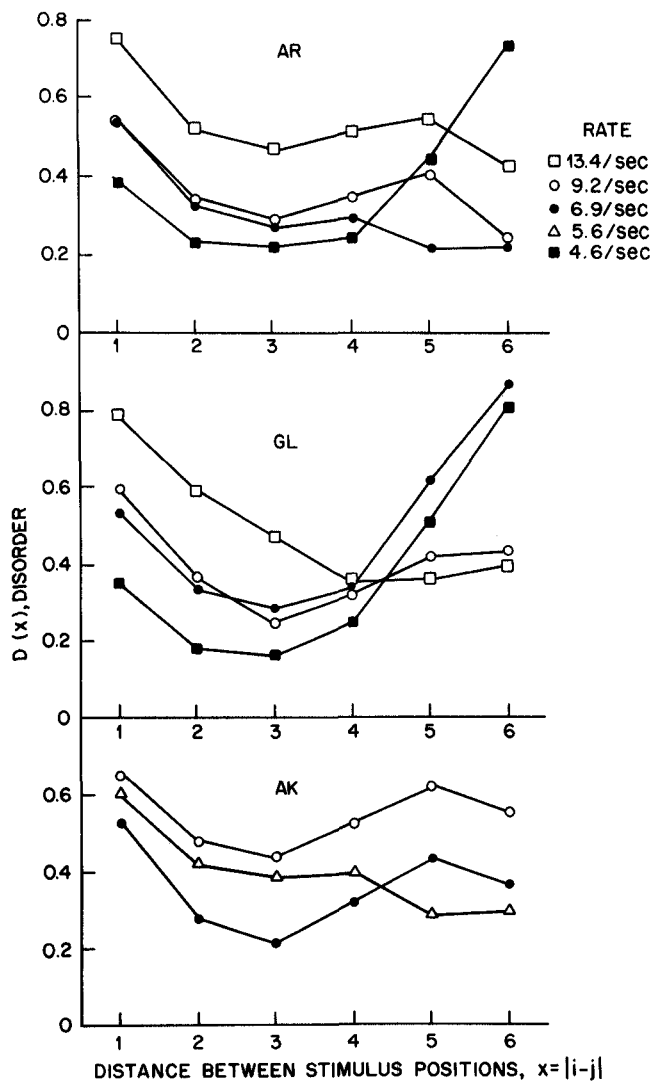

Figure 6. Disorder $D(x)$ as a function of x, the separation between Stimulus Positions i and j. $D(x) = 1$ if there is total disorder, $P_{iBFj} = 0.5$ for all $|i - j| = x$. $D(x) = 0$ for complete order, $P_{iBFj} = 1$ or $P_{iBFj} = 0$ for all $|i - j| = x$. Different symbols identify different numerals rates as indicated on the figure. Data for each subject are averaged over stimulus positions, response positions, and target identity.
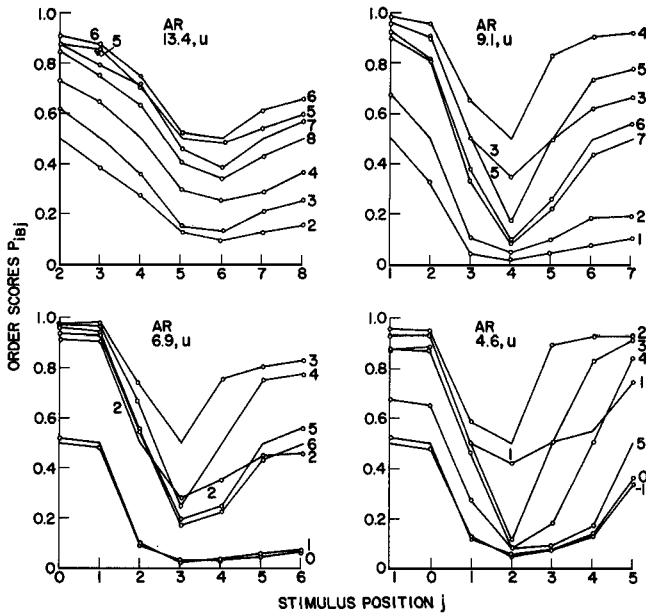
Figure 7. The proportion P_{iBj} of trials on which numerals from Stimulus Position i are reported earlier in the response than those from Position j, as a function of j, with the curve parameter i. (Data are for subject AR, target U; the four panels are for four different numeral rates, as marked.)

ported before those from Position j (irrespective of the order of i and j in the stimulus). Then,

$$P_{iBj} = \frac{n(iBj)}{n(iBj) + n(jBi)} . \quad (2)$$

For every condition, there are $\binom{7}{2} = 21$ independent i, j pairs in the seven positions for which data are collected. (Note that $P_{jBi} = 1 - P_{iBj}$ and, by convention, $P_{iBi} = 0.5$.)

Data. Typical results for P_{iBj} are plotted in Figure 7 for subject AR and target U. Each panel shows data for a different numeral rate, indicated in the figure. The abscissa gives Stimulus Position j, and the curve parameter is Stimulus Position i for (i, j) in the critical set of numeral positions. The curves are U-shaped and are roughly parallel to each other. To understand this data format, consider the uppermost curve of the top left panel in Figure 7, P_{6Bj}. That P_{6Bj} lies above the other curves means that numerals from Stimulus Position 6 tend to be reported before numerals from other positions, whether or not those other positions were earlier (2 through 5) or later (7 and 8) than Position 6. Data for other numeral rates shown in Figure 7 are similar, although there are one or two curve crossings in panels that represent slower numeral rates. The different ranges of stimulus positions graphed in the various panels reflect the adjustment of the critical set of stimulus positions to include the earliest reported positions. Similar results for a second subject (GL) and another of the three targets (C) are shown in Figure 8, and for the third subject (AK) and third target (Sq), in Figure 9.

Precedence and folding. When the P_{iBj} curves are laminar (i.e., when they do not cross each other), the order in which they lie above each other defines an order of precedence. (Laminarity is equivalent to monotonicity in multidimensional scaling.) Thus,

in Figure 7, upper left, the order of precedence is (6, 5, 7, 8, 4, 3, 2), which is quite different from the order of presentation (2, 3, 4, 5, 6, 7, 8), and this difference is reflected in the low P_{co} and in the high $D(x)$ scores in this condition (13.4/s). Results with other subject–target-rate combinations also were generally laminar (Reeves, 1977), with a few violations at slow numeral rates. Laminarity and violations of laminarity are considered later in connection with the various attention models.

The regularity of the P_{iBj} curves shows that P_{co} is near chance not because the subject responds in a haphazard order but because he responds systematically in an order that is different from the stimulus order. At slower numeral rates, P_{co} rises above chance (Figure 5), and the precedence order begins to approximate the presentation order more closely (e.g., Figures 7, 8, and 9, lower right panels). Precedence orders such as (6, 5, 7, 8, 4, 3, 2) show *folding,* that is, a tendency to report a central position first, then the nearest postcentral and precentral positions, followed by more distant positions with pre- and postcentral positions mixed together haphazardly. The central position (here, 6) may be termed the *folding point.* Precedence orders in every condition were folded, although at slow numeral rates the folding was less symmetrical than at high rates.

Phenomenological Reports

Subjects were surprised by their inability to report the order of the numerals correctly. They claimed to report the numerals in the order in which they "saw" them and expected that order to be veridical. Subjects also claimed that they could not and did not implicitly name the numerals as they appeared on the screen. Rather, if they named a numeral, it was only while entering the response on the teletypewriter at the end of the trial. This is not surprising, inasmuch as the rate of implicit naming of numerals (about 3/s, Landauer, 1962; Sperling, 1963) is slower
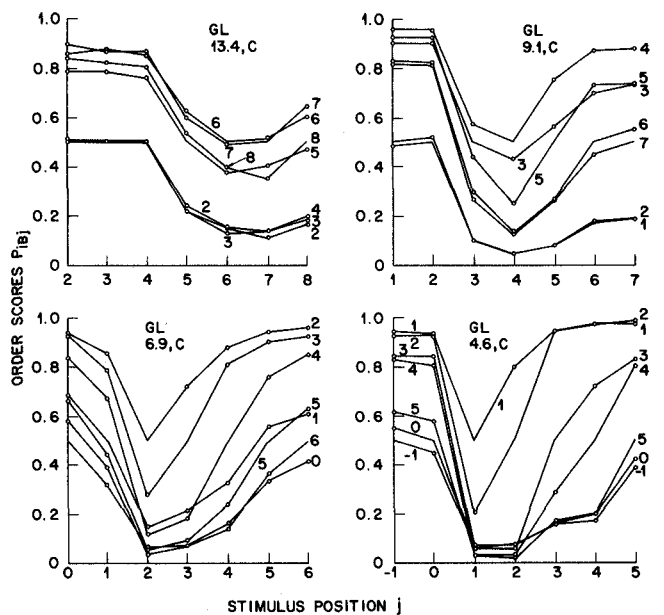

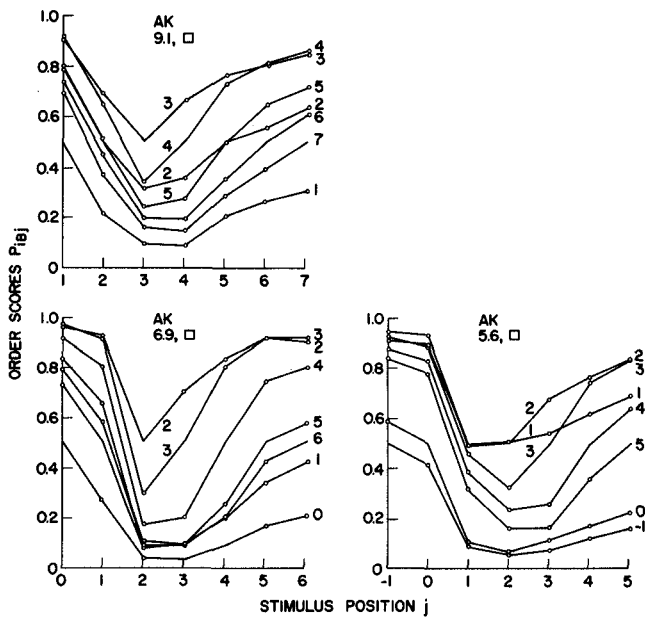Figure 8. Subject GL, target C, as for Figure 7.

Figure 9. Subject AK, target Sq, as for Figure 7.

than the presentation rates used here. These introspections suggest that the to-be-reported numerals were held in visual short-term memory (VSTM) until the time to respond and imply that the reordering (relative to the stimulus order) of the numerals in the subject's visual memory occurred prior to the subject's awareness of the numerals. Further evidence for visual recall of rapidly presented digits was obtained by Kaufman (1977). She studied recognition memory, using RSVP both with numerals as in the present experiments, and with nonidentifiable line patterns unique to each trial (and therefore presumably representable only in VSTM). She found no essential differences in performance between the numerals and the nonverbalizable line patterns, suggesting that the numerals, too, must have been stored in visual memory.

Discussion

Precedence

Our results show that in the attention shift procedure, identity information is good but order information is poor. Poor order information in RSVP has often been reported (e.g., Scarborough & Sternberg, 1967) but it has never been adequately described or explained. One of the most important new findings of this study is that report order can appear to be randomly related to presentation order (P_{co} near chance) but be systematically related to a *precedence order.* In later sections, we propose that the actual order of the numerals in visual short-term memory (following the attention shift) is determined not so much by presentation order but by the amount of attention allocated to each position in the numeral stream. We ultimately present an attentional theory to explain the data but must first develop a succinct, comprehensive description of the data upon which to base the theory.

To begin, we dispose of two essentially trivial explanations of the near-chance P_{co} values and U-shaped P_{iBj} data: (a) random guessing of items and (b) forgetting the order of items. Second,

we establish that all of the main results are consistent with the simple, strength description of order, which we term *precedence.* Third, we derive precedence from an attention gating model. Finally, we show that the attention gating model can be derived from an even more fundamental model (the generalized attention gating model).

The Item Guessing Hypothesis

Suppose the subject remembers less than the required four numerals and completes his report with random guesses from the remaining numerals. Because actually recalled numerals tend to come from the center positions of the critical set (Figure 3), whereas guessed numerals would tend to be those that had in fact been presented at the start or end of the critical set, numerals from the center positions (with high P_i) would tend to be typed before those from extreme positions (with lower P_i)—the result actually shown in Figure 7.

To refute the guessing hypothesis, we note that the proportion of numerals guessed in the central positions (third, fourth, and fifth) of the critical set was less than 0.07 for all experimental conditions. (This low proportion of guessing follows from the very high observed P_i in these positions, as noted in Reeves, 1977.) Because these central positions also show a near chance P_{co} and U-shaped P_{iBj} functions at high numeral rates, guessing, if it occurs at all, cannot have accounted for the main (center position) findings. Indeed, in a subsidiary experiment, similar U-shaped P_{iBj} curves were obtained at similarly fast numeral rates when the set of numerals was enlarged with five additional letters. Because subjects virtually never reported characters from the terminal positions of this extended set (which they would have if they had guessed items at random), we have direct evidence that virtually no random guessing of items occurs.

The Forgetting Hypothesis

Subjects might forget either the numerals or their positions before reporting them on the teletypewriter, although with only a four-numeral memory load any forgetting is rather unlikely. Numerals whose identities are forgotten would be replaced by guessed numerals, and we have seen above that guessing cannot account for the main findings. Forgetting the order of the numerals might occur if an initially correct recall of their order was disrupted later on, perhaps by the requirement to report so many of them. However, we now show that forgetting of order cannot have been critical.

Consider $P_i(r)$, the proportion of trials in which the subject reports a numeral from Position i of the critical set in Position r ($r = 1, 2, 3, 4$) of the response (as in Figure 4; the full set of $P_i(r)$ matrices is in Reeves, 1977). If the order of the numerals were forgotten, $P_i(1)$ should have a broader distribution over i (more variance) in the present (Recall-4) experiment than in an experiment in which the subject is asked to recall only the first numeral that he can (the Recall-1 experiment of Reeves, 1977). This is because if the subject is asked to report four numerals and forgets their order, he will occasionally interchange the first numeral with one from a later position in the critical set and so broaden $P_i(1)$. Such an interchange cannot occur in the Recall-1 experiment. Nevertheless, the P_i distributions in Recall-1 are
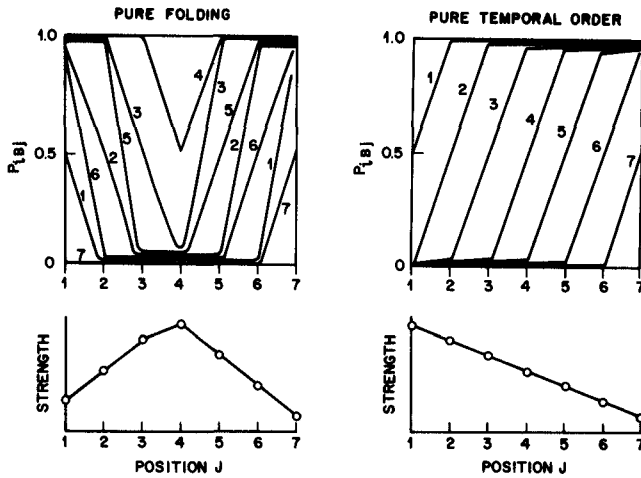
Figure 10. Comparison of pure folding and pure temporal order in ideal responses. (Upper panels show examples of laminar P_{iBj} versus j graphs; lower panels show the strengths derived from these P_{iBj} values as a function of Stimulus Position j. See Figure 7 and text.)

indistinguishable from the $P_i(1)$ distributions obtained in the present Recall-4 experiment (Reeves, 1977). Therefore, the requirement to report four numerals did not produce any significant overall scrambling of report order.

We also found in the Recall-1 experiment that at high numeral rates subjects were virtually unable to estimate the relative position in the critical set—early, middle, or late—of the single numeral they actually recalled, even after extensive practice with feedback (Reeves, 1977). The subjects did not know the position of the single numeral they had to recall. It is therefore very unlikely that they first knew the order and then forgot it. A further analysis of the Recall-4 data showed that response pairs involving Response Position 1 were not qualitatively different (e.g., in the laminarity of P_{iBj}) from other response pairs. Although this does not rule out the possibility of some forgetting of order in the later response positions, forgetting cannot account for the main features of the order data because these are observed in data from the very first response position.

Precedence for Positions

Proof of the feasibility of a precedence scale. Having discounted the item guessing and order forgetting hypotheses, we account for the paradox of a near-chance proportion of correctly ordered response pairs (P_{co}), and yet a systematic ordering of the positional order scores (P_{iBj}), by postulating an internal precedence scale on which numeral positions are ordered.

It was asserted in the Results section that observed precedence orders such as (6, 5, 7, 8, 4, 3, 2) show folding around a central position (here, 6). In this section, we provide proof that such one-dimensional precedence scales do in fact describe the data. Figure 10 shows two examples of theoretical data: Figure 10 (left) shows data that would be expected if the only systematic effects were pure folding (e.g., 4, 3, 5, 2, 6, 1, 7), in which overall order information P_{co} is at chance; Figure 10 (right) illustrates pure temporal order (e.g., 1, 2, 3, 4, 5, 6, 7), in which P_{co} is 1.0.

These examples show that either chance or perfect overall order information (P_{co}) is compatible with the existence of a precedence scale. The data actually obtained tend to follow the pattern for pure folding shown in Figure 10. However, the question is not, for the moment, the extent of order information but rather whether the data can be adequately described by a single precedence scale in the first place. (We will conclude that precedence scales provide a very good description of the data, except for some discrepancies at slower numeral rates.)

Let V_i be a one-dimensional scale of precedence on which each Position i of the critical set has a precedence value V_i. (Although all numeral stream positions may have precedences, we consider here only the seven critical set positions for which data are available. In the model developed below, a combined precedence V_x is assigned to the remaining positions.) If this scale exists, the 21 independent P_{iBj} values in each condition should be predictable from the seven V_i values (6 independent values) for that condition; indeed, P_{iBj} should be determined just by V_i and V_j, the component precedences. The laminar structure of the observed P_{iBj} suggested that P_{iBj} might be related to V_i and V_j by a difference equation:

$$P_{iBj} = H[V_i - V_j] = H[\Delta V], \tag{3}$$

where the monotonic increasing function H maps the real-valued differences between precedences into the interval [0, 1]. Because $P_{iBj} = 1 - P_{jBi}$ by definition, $H(0) = 0.5$ and $H(\Delta V)$ is antisymmetric about 0.

The quadruple condition: A consequence of precedence. Equation 3, without further specification of the functions V_i and H, implies the following quadruple condition on the probabilities (Block & Marschak, 1960):

$$P_{iBj} < P_{kBl} <=> P_{iBk} < P_{jBl} \tag{4}$$

The quadruple condition follows from Equation 3 because if $P_{iBj} < P_{kBl}$, then $V_i - V_j < V_k - V_l$ and, by rearrangement, $P_{iBk} < P_{jBl}$.

Significant violations of the quadruple condition would require rejection of Equation 3. The quadruple condition is tested by inspecting the relevant sets of four positions in each experimental condition. There are two pairs of responses chosen from seven stimuli, and thus 210 quadruples to check in each condition. A violation of the quadruple condition occurs when the observed inequality on the right side of the equation is in the opposite direction of the expected inequality. The number and size of the observed violations were small in every condition, and so Equation 3 is not strongly falsified. However, between 4% and 7% of the quadruples in each experimental condition were violations, slightly more than the 2%–3% expected by chance (see Appendix for details).

A Strength Model of Precedence

Assumptions and Predictions

For the moment, we ignore the few residual violations of the quad condition and proceed to a specific strength model of precedence, which specifies the function H of Equation 3 as a Normal distribution and permits estimation of the values of V_i.

The strength model assumes that each position in the critical set has a precedence equal to V_i plus a random error term $\epsilon_{i,T}$ that varies from trial to trial, T, yielding an instantaneous strength of precedence, $s_{i,T} = V_i + \epsilon_{i,T}$. The $\epsilon_{i,T}$ are assumed to be independent random samples from a Normal distribution centered at zero with unit variance. (Lowercase letters indicate instantaneous values; uppercase letters indicate average values. Because v_i is assumed not to vary from trial to trial, $V_i = v_i$.)

On any particular trial T, the model produces a "response" R_T of exactly four items; the response is scored just like data. The response R_T is produced by selecting the numerals in the four positions (say, $i, j, k,$ and l) with the four highest strengths so that $R_T = (i, j, k, l)$ if and only if

$$V_i + \epsilon_{i,T} > V_j + \epsilon_{j,T} > V_k + \epsilon_{k,T} > V_l + \epsilon_{l,T} > V_m + \epsilon_{m,T},$$

(5)

for all $m \neq i, j, k, l$. The response R_T varies from trial to trial because each precedence value V_i is combined with a random error, $\epsilon_{i,T}$.
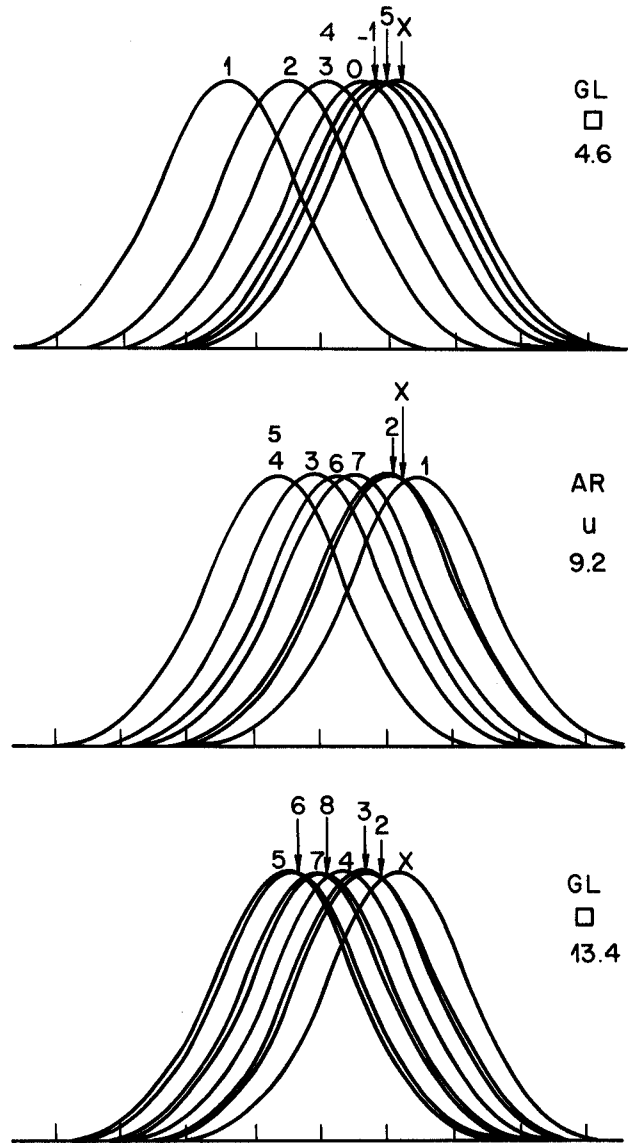
To illustrate the strength model, strength distributions $s_{i,T}$ for three conditions (good order, bad order, and a typical intermediate condition) have been plotted in Figure 11. Strength values increase to the left so that the strongest numeral position—from which numerals are most likely to be reported and are most likely to be written first in the response—is represented by the leftmost distribution. The $s_{x,T}$ distribution (marked by an x in Figure 11) represents all reports from numeral positions outside the critical set and provides our best estimate of the strengths of the residual (poorly attended) numeral positions. The distribution $s_{x,T}$ has the character of a composite noise distribution, although it should be noted that in principle there are no 100% noise-determined responses.

Figure 11, top panel, shows strengths for subject GL, target Sq, at the slowest rate; this condition produced the best experimental correct order ($P_{co} = 0.86$). The middle panel of Figure 11 shows precedences for subject AR, target U, rate 9.2/s, for which P_{co} is lower (0.72). The bottom panel shows precedences for subject GL, target Sq, rate 13.4/s, for which P_{co} was near chance (0.54). At the fastest numeral rates, the strength functions crowd together near the residual distribution $s_{x,T}$, and order information is minimal. At slower rates the functions separate so that order becomes more consistent. Because the order of strengths (left–right order of distributions) is not veridical, more consistent order does not necessarily imply more accurate order (and so P_{co} does not rise to 1.0).

A Monte Carlo simulation (outline). A computer was programmed to simulate the model described in Figure 11. Eight independent Normally distributed values $\epsilon_{i,T}$ were generated for each trial T, added to the V_i, and the four items having the largest values were chosen as the response. In each condition, some 2,000 trials were run in this way to build up a large enough trial set for the item and order scores to be stable in the second decimal position. We then proceed as follows. First, given an artificially generated data set, we demonstrate that the generating V_i parameters can be recovered. We show first an extremely quick, almost correct Thurstone Case 5 method for recovering V_i from data. Second, the Case 5 computation is used to estimate V_i for all subjects and conditions. Third, we then demonstrate a much

more tedious, rigorous parameter recovery using iterative Monte Carlo simulation. Fourth, the estimates of V_i derived from order scores are used also to predict item scores. Last, the order and item predictions of the strength model are evaluated.

1. Thurstone Case 5 approximation to V_i. Adding Normally

INSTANTANEOUS PRECEDENCE , $V_i + \epsilon_{i,T}$

Figure 11. Strength model of precedence. (Instantaneous precedences [strengths]) $s_{i,T}$ of Stimulus Position i on trial T are represented on the abscissa; the ordinate represents probability; the labels above the curves indicate the stimulus positions. Mean values of precedence V_i fall directly under the peaks of the curves. The distributions derive from adding to V_i a random term $\epsilon_{i,T}$. Top graph = Distribution of $s_{i,T}$ for subject GL, target Sq, 4.6/s, the condition with the highest probability of correctly ordered response pairs P_{co} in the experiment; middle graph = subject AR, target U, rate 9.2/s, a typical condition; bottom graph = subject GL, target Sq, rate 13.4/s, the condition with the lowest P_{co}.)
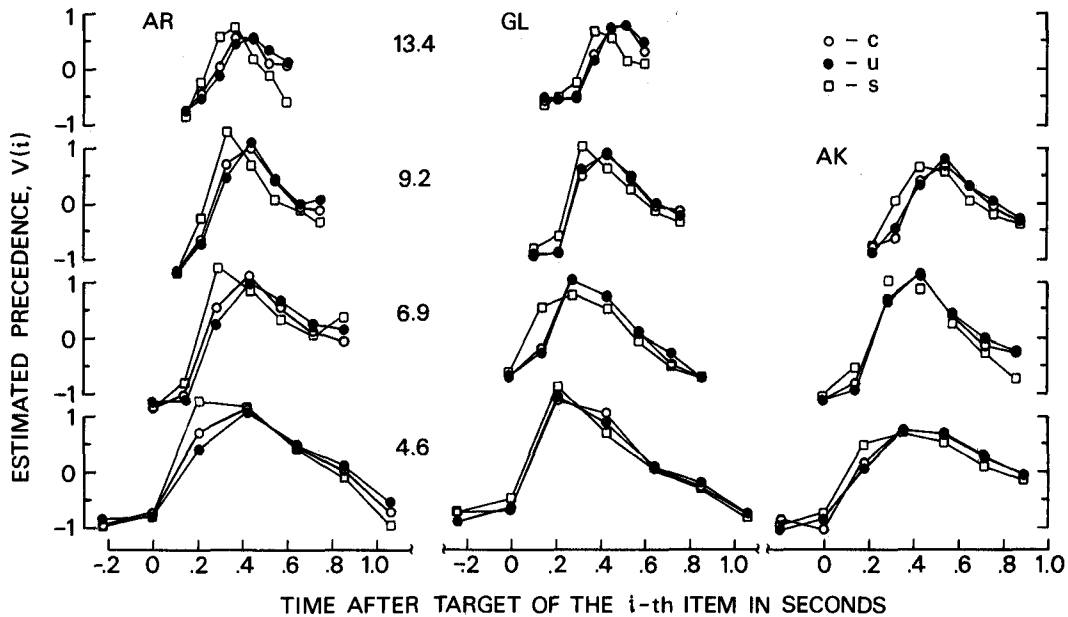
Figure 12. Estimated precedences $V(i)$ as a function of the time in seconds from onset of the target to onset of the ith numeral. (Panels within a column represent the same subject; panels within a row represent the same numeral rate, as indicated. Targets U, C, and Sq are indicated by closed circles, open circles, and open squares, respectively.)

distributed noise $\epsilon_{i,T}$ in the process model (Equation 5) is equivalent to choosing H (in Equation 3) as a cumulative Normal distribution function with unit variance, centered at zero (with $P_{iBi} = H(0) = 0.5$). Estimates of V_i are then given by

$$V(i) = \frac{1}{7} \sum_{j=1}^{7} H^{-1}(P_{iBj}), \tag{6}$$

where H^{-1} is the inverse cumulative Normal distribution function. (We use V_i for theoretical model parameters and $V(i)$ for estimations based on data.)

Equation 6 yields unbiased estimates of the V_i under the assumptions of Thurstone's Case 5 (Bock & Jones, 1968, p. 122; Thurstone, 1927), in which all responses are assumed to be independent (sampled with replacement). In our procedure, responses are all different (sampled without replacement) and, hence, not independent. Although this violates the assumptions of the Case 5 analysis, we show below that the resulting $V(i)$ are very nearly optimal ($r > .98$, all conditions) and a useful starting point for more complicated estimation procedures.

2. Estimated precedences, V(i). In Figure 12, precedences $V(i; sj, rt, tg)$ estimated by Equation 6 from subjects' data are plotted as a function of p_i (given in the legend), with a separate curve for each condition of Subject × Numeral Rate × Target. [The explicit dependence of V on subject (sj), numeral rate (rt), and target (tg) is omitted when it is clear from the context, and precedence values are written simply as $V(i)$.] Obviously, the $V(i)$ are highly regular and consistent between subjects. (The stimulus positions for the various panels are indicated in Figures 7–9 on the abscissa.) That the maximum of $V(i)$ occurs at the same horizontal position in the various panels is a consequence of the temporal abscissa scale.

The precedence functions $V(i; sj, rt, tg)$ of Figure 12 clearly have inverted-U shapes in all instances, although at 13.4/s, the right-hand falloff is truncated by the shortness of the critical set. Except for statistical fluctuation, these $V(i)$ describe the order in which subjects make their responses, writing the item from the highest valued position first, and so on. The inverted-U $V(i)$ functions reflect the property of folding in the sequence of responses emitted by the subject. When the maximum of $V(i)$ occurs at Position m, the corresponding responses tend to be folded around m, with m being written first, then $m + 1$, $m + 2$, and $m + 3$, interleaved with $m - 1$, $m - 2$, and $m - 3$, and so forth.

If $V(i)$ were a monotonic decreasing function of i, the responses would reflect pure temporal order. Ideal relations of $V(i)$ to P_{iBj} with folding and with pure temporal order were illustrated previously in Figure 10.

The data of Figure 12, which are based on the relative order of two response items, are remarkably similar to the data of Figure 3, which are based on whether or not an individual item occurred in the response. This will be important for the models developed below, but first we ask whether the Case 5 model reasonably describes these data. A simple initial check is to insert the estimated $V(i)$ back into Equation 3 directly and compute predicted $\hat{P}_{iBj}$ scores for comparison with the data P_{iBj}. These scores agree well: Mean absolute differences $|P_{iBj} - \hat{P}_{iBj}|$ are 0.06 or less in all conditions.

3. Monte Carlo simulation with Thurstone Case 5 parameters. As noted above, the simple Thurstone Case 5 model is not strictly correct because it ignores the problem of independence. On each trial there are only $10 \cdot 9 \cdot 8 \cdot 7 = 5,040$ possible responses, not the 10^4 that there would be if the numerals had been written independently of one another in each response position. Although independence over trials can be assumed, independence of the pairwise comparisons within a trial cannot, contrary to the as-
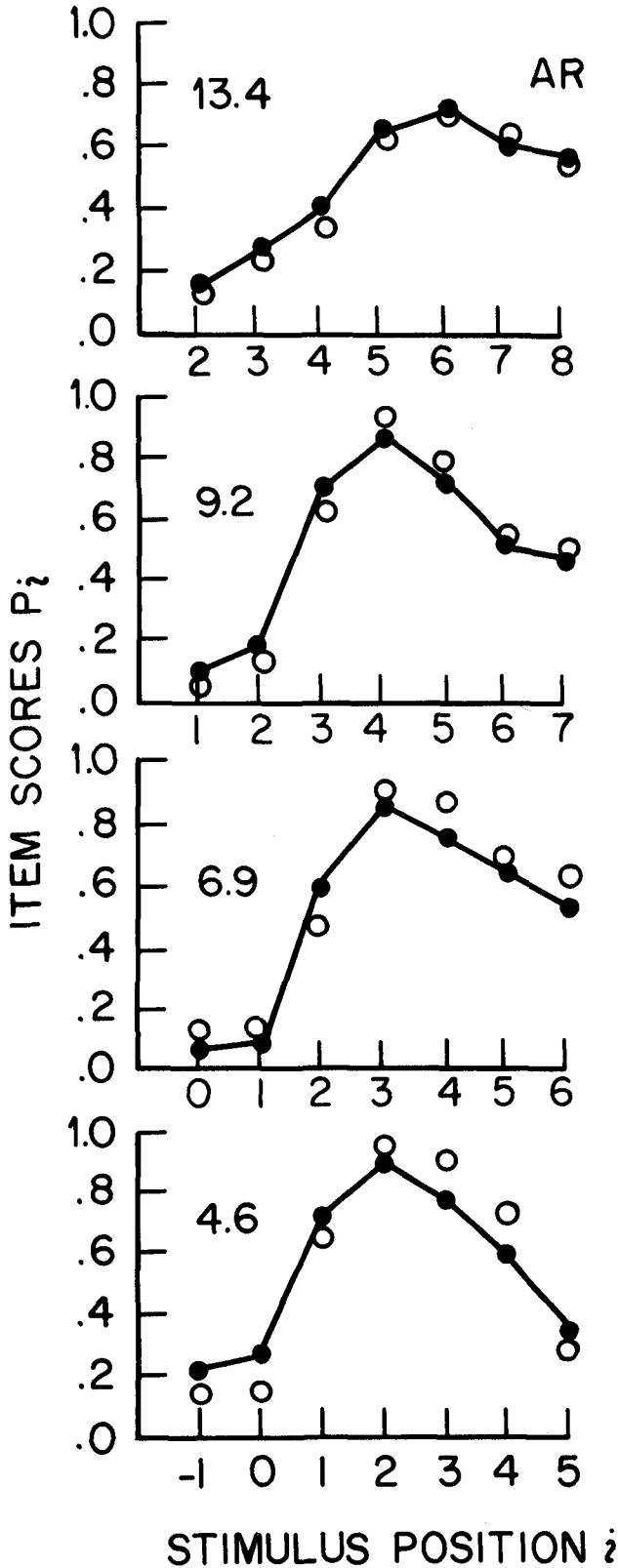
Figure 13. Item scores: the probability of the numeral from Stimulus Position *i* appearing anywhere in the response, P_i, plotted against Stimulus Position *i*. (Data for subject AR, target U, are indicated by closed symbols. Open symbols are scores predicted by the strength model. Data are from Figure 3, col. 1, target U.)

sumptions of Thurstone's Case 5. To make the assumptions of the analysis strictly congruent with the experimental procedure requires, unfortunately, a substantially more complicated analysis based on the Monte Carlo data generation outlined above; Case 5 merely provides the starting point.

The $\hat{V}_i$ parameters used to generate Monte Carlo trials were estimated from the Case 5 solution (Equation 6) by setting $\hat{V}_i = V(i)$. Of the seven $\hat{V}_i$, six are independent, and these are sufficient to predict the P_{iBj}. Because a numeral from outside the critical set is sometimes reported, with probability

$$q = 1 - \frac{1}{4} \sum_{i=1}^{7} P_i,$$

a simple precedence $\hat{V}_x = H^{-1}(q)$ is assigned to represent these combined, outside positions. It is assumed for convenience that the corresponding error term $\epsilon_{x,T}$ also has unit variance, so $s_{x,T} = \hat{V}_x + \epsilon_{x,T}$. The 2,000 Monte Carlo trials in each condition were analyzed with the same methods as the data to obtain seven predicted item scores $\hat{P}_i$ and 21 predicted order scores $\hat{P}_{iBj}$ for comparison with the data.[2] Each condition was tested separately.

In some cases, the differences between the observed and predicted data were used to drive an iterative optimization of the Monte Carlo derived $\hat{V}_i$ parameters. Although small improvements did result thereby, they were not sufficiently significant to warrant the increased complexity of estimation. There were failures of prediction (see below), but they were not caused by nonoptimal parameters (see Appendix).

4. Fit to item scores. The model parameters $\hat{V}_i$ were estimated solely from the ordered pairs of responses. Yet, the predicted item scores, $\hat{P}_i$, fit the data (P_i) quite well, accounting for 94% of the variance of the data or better. Here, percentage variance equals $100[1 - \Sigma (P_i - \hat{P}_i)^2/\Sigma (P_i - EP_i)^2]$, where EP_i is the mean of P_i. The predictions and data for one representative set of conditions are shown in Figure 13. However, the $\hat{P}_i$ slightly overestimated the P_i for small i and underestimated P_i for large i. This small but systematic error increased at slower numeral rates. The discrepancies were statistically significant, inasmuch as more than 15% of the discrepancies were larger than the 95% confidence interval around each $\hat{P}_i$, that is, outside the interval I_{95},

$$I_{95} = \hat{P}_i \pm 2 \sqrt{\frac{\hat{P}_i(1 - \hat{P}_i)}{N - 1}},$$

where N is the number of trials.

The fit to the item scores for each position, $P_i(r)$, was less good. Although mean absolute deviations $|P_i(r) - \hat{P}_i(r)|$ between predicted and obtained scores averaged just 6% in the various conditions of Subject × Rate × Target, the percentage of variance accounted for averaged only 63% ($r = .80$). The model fit the shape of the first response $P_i(1)$ reasonably well, accounting for 75% of the variance, but underestimated the observed differences between response positions and so did not position the peaks of the $P_i(r)$ functions late enough for $r = 3$ and $r = 4$.

---

[2] In the Monte Carlo tests of the models, it is immaterial which method of scoring is used so long as the scoring method is the same in the model and the data.

5. Fit to order scores. The $\hat{P}_{iBj}$ matrices predicted by the model were similar to the data P_{iBj} matrices, as illustrated in Figure 7 (data) and Figure 14 (model). The fits were very close at the fastest rate (upper left panel) but not as good at slower rates (particularly the lower left panel), as discussed below. The percentage of variance accounted for is

$$\%VAR = 100 \left\{ 1 - \frac{\sum (P_{iBj} - \hat{P}_{iBj})^2}{\sum (P_{iBj} - EP)^2} \right\}$$

(where EP, the mean of P_{iBj}, is 0.5). The $\%VAR$ averaged 94%. The mean absolute deviations between predicted and data order scores rose slightly from 0.04 at the fastest rate to 0.07 at the slowest, averaged over subjects and targets.

A consequence of the strength model is that item scores and precedences should be closely related: the higher the precedence, the more likely a report. The success of the strength model in predicting item and order scores from precedences confirms this. At a more descriptive level, both in the data and in the model, the correlations between $V(i)$ and $H^{-1}(P_i)$, both in z scores, were

Figure 14. Order scores, P_{iBj}, predicted by the strength model, for subject AR, target U, numeral rates 13.4/s (top) and 4.6/s (bottom).

greater than 0.96 at the fastest numeral rates and 0.93 at the slowest numeral rate.

Differences Between the Model and Data

Nonoptimality of the strength model. The present strength model is not the optimal precedence model satisfying Equation 3, because (a) it makes the strong and improbable assumption of equal-variance Normal distributions for the trial-to-trial variation in s_i and (b) the parameters—for technical reasons—are not completely optimal. However, the small extent to which the model deviated from the P_{iBj} data makes it unlikely that any powerful improvement in the model can be tested with the present data. Nevertheless, there are two, possibly related, areas of difficulty: (a) the prediction of $P_i(r)$ from parameters derived from P_{iBj}, as discussed before, and (b) the data's small but consistent violations of the quad condition and corresponding violations of laminarity at slow numeral rates.

Laminarity violations. An example of a laminarity violation occurs in Figure 7, rate 4.6, where the P_{1Bj} curve runs flat through the right-hand side of the figure, crossing the P_{3Bj} and P_{4Bj} curves. Similar violations of laminarity were observed at some numeral rates for all subjects (e.g., rates 9.1 and 6.9 in Figure 8 and rates 9.1 and 5.6 in Figure 9). In fact, these kinds of violations are the only obvious violations of Equation 1 in Figures 7, 8, and 9 and are typical of the remaining (undisplayed) data. These violations of laminarity cannot be accounted for by any model satisfying Equation 3, and the question naturally arises: What causes them?

We consider two possible causes of laminarity violations: A change in the precedence scale v_i either (a) within each trial or (b) between trials. A change within each trial could occur if the basis of the ordering changed during the course of a trial so that, for example, late positions were all given similar precedences at the start of the response, when early positions tended to be processed, but were given different precedences at the end, when their temporal positions began to exert a greater effect. We did not succeed in simulating the laminarity violations with reasonable choices of early- and late-in-the-trial V_i parameters.

Alternatively, if order scores were derived from two (or more) sets of trials with different precedence scales, laminarity might fail. Laminarity (Equation 3) could hold in each data set individually but not in the combined data because, when averaging sets of trials, it is not the P_{iBj} values from each set that are averaged but the $n(iBj)$ components of these values. If precedences peaked earlier in one set of trials than in another, orderings of early positions would be dominated by the earlier set. We were able to simulate violations of laminarity with mixtures of x- and y-type trials in which Vx_i peaked about one position earlier than Vy_i. That is, the values Vx_i and Vy_i each had a similar form as that shown in Figure 12, and $Vy_i \approx Vx_{i-1}$. Although each set of predictions $\hat{P}_{x,iBj}$ and $\hat{P}_{y,iBj}$ was laminar individually, the $\hat{P}_{x+y,iBj}$ (taken from the combined set of trials) were not laminar and showed the violations in laminarity typical of the data in Figures 7, 8, and 9.

An earlier peak in Vx_i than Vy_i would imply shorter attentional reaction times (ARTs, in the terminology of Sperling & Reeves, 1980) in x-type trials. If shorter ARTs were correlated with shorter motor reaction times (MRTs), an independent check on the hypothesis of two sets of trials could be made by conditioning trial type on MRT. However, ARTs and MRTs are only weakly cor-
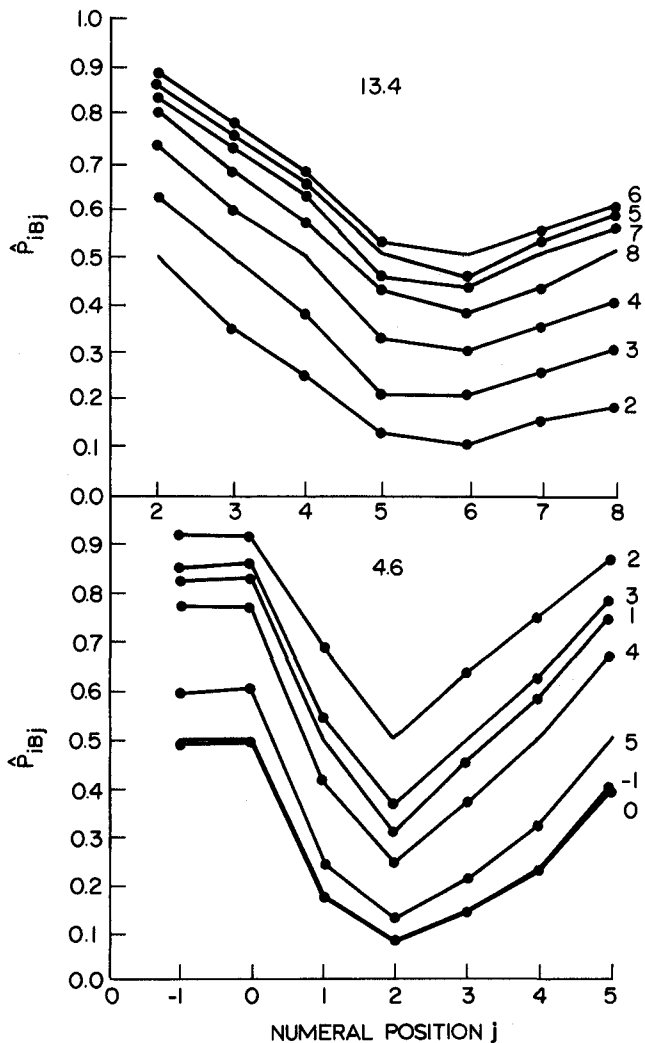
related in these data (Sperling & Reeves, 1980), and they are therefore not suitable for a test. Thus, while we have been able to demonstrate that reasonable mixtures of trials could produce the kinds of violations of laminarity we observed, we do not have independent corroborative evidence for this proposed mechanism. Incorporation of ART variability into the model might also solve the $P_i(r)/P_{iBj}$ discrepancy, but inasmuch as such models are too complicated to consider here, these improvements are deferred.

Conclusion

A strength model with equal variance, Normally distributed error at each position accounts for about 94% of the variance in P_{iBj} scores and 98% of the overall item scores P_i, with precisely the same parameters for both sets of predictions. The fit to individual item scores $P_i(r)$ was not as good, averaging 63% VAR. These fits are near the limit of precedence models because both discrepancies—the inconsistencies between item and order scores ($P_i(r)$, P_{iBj}) and the nonlaminarity (quad violations) of order scores at slow rates—are due to violations of Equation 3. The probable cause of both limitations is a significant trial-to-trial variation in attention reaction time. All of the alternative explanations that occurred to us could be excluded (see Appendix).

An Attention Gating Model of Attention Shifts (AGM)

We now assume that the strength model of precedence is essentially correct. The Monte Carlo simulations have demonstrated the utility of the six independent precedences in each condition for predicting the $P_i(r)$ and P_{iBj} scores. Here, we provide an attention gating model (AGM) to derive the values of the precedences in each condition and, thereby, the $P_i(r)$ and P_{iBj} scores.

AGM Assumptions

Overview. The attention gating model assumes that all stimulus items—numerals in the present experiment—are represented *peripherally,* but only some survive to be represented *centrally.* The strength of the central representation ultimately determines precedence. Specifically, the model assumes that the central availability of each numeral at each moment in time, t, is determined jointly by two factors: an attention function that controls an input gate and a stimulus availability function that determines the items that approach the gate. The first factor is called the *attention gating function* $a(t - \tau)$. *Gating* in the model corresponds to *attention* in the real world. Thus, we assume that before the moment τ when attention is shifted to the numeral stream, $a(t - \tau) = 0$; afterwards, $a(t - \tau) > 0$. The second determinant of central availability of a numeral is the *peripheral availability* of the numeral in Position i, which is governed by a persistence function, $b_i(t)$. Central availability $c_i(t)$ of a numeral i is determined by the peripheral input from i during the time the gate is open, that is, by the product $c_i(t) = a(t - \tau)b_i(t)$. The precedence $v_i(t)$ at time t of numeral i is the cumulative central availability of numeral i, which is the integral of $c_i(t)$ from the time t_i of occurrence of i until t:

$$v_i(t) = \int_{t_i}^{t} c_i(t')dt' = \int_{t_i}^{t} a(t' - \tau)b_i(t')dt'. \quad (7)$$

The AGM assumes there is no central forgetting during a trial. Precedence v_i is perturbed by internal noise (error, $\epsilon_{i,T}$) to produce the predicted strength $s_{i,T}$ of item i on trial T, as described in Equation 5.

Persistence. In the versions of the model detailed here a numeral is assumed to persist from its onset until the onset of the next numeral (see Appendix); thus, $b_i(t)$ is a rectangular pulse: $b_i(t) = 1$, for $t_i < t < t_{i+1}$, otherwise $b_i(t) = 0$ (see Figure 15, top).

Attention gating function, $a(t - \tau)$. The attention function is assumed to be time invariant, independent of the target that triggered it, and independent of numeral rate. After some search, we chose for $a(t - \tau)$ a Gamma function that represents an input impulse filtered sequentially through two exponential stages.[3] The proposed $a(t - \tau)$ is described by Equation 8 and illustrated in Figure 15:

$$a(t - \tau) = \begin{cases} \dfrac{(t - \tau)}{\sigma\alpha^2} e^{-(t-\tau)/\alpha} & t \geq \tau \\ 0 & t < \tau, \end{cases} \quad (8)$$

where τ is the trigger time at which the attention gate begins to open, α is the time constant of the underlying exponential filters, and σ scales $a(t - \tau)$ relative to $\epsilon_{i,T}$, the internal noise. The model with these particular choices of functions for $a(t - \tau)$, $b_i(t)$ can provide a reasonable fit to the data of the three subjects (see below).

Figure 15 (bottom) illustrates the computation carried out by the attention gate. The horizontal axis shows the onset times t_i of numerals at each of the four rates. The hatched rectangle on the time line indicates the peripheral availability, and the hatched area under the attention gating function indicates the precedence v_5 of the fifth numeral position at the rate of 13.4 numerals/s. The left-hand portion of Figure 15 (top) shows a block diagram representation of the precedence computation; the right-hand portion of Figure 15 (top) illustrates the combination of precedence with noise to yield strength s_i (Equation 7).

Noise, $\epsilon_{i,T}$. The precedence v_i of item i on each trial T in the model is subject to random Gaussian noise, $\epsilon_{i,T}$, with zero mean and unit variance; $\epsilon_{i,T}$ varies independently from item to item and trial to trial. In order to maintain compatibility between the AGM and the earlier strength model, the scale factor σ that is needed to relate strength s_i to noise $\epsilon_{i,T}$ is incorporated into v_i— via $a(t - \tau)$—rather than into $\epsilon_{i,T}$. The sum $s_{i,T} = v_i + \epsilon_{i,T}$ is the predicted strength of item i on trial T. To compare predictions to data, we consider the expected values over T, which are denoted by capital letters as follows: $s_i = S_i$, $E(\epsilon_{i,T}) = 0$, $E(v_i) = V_i$, and we have, simply,

$$S_i = V_i. \quad (9)$$

---

[3] Two exponential (RC) stages were chosen because (a) the resultant transfer function has an approximately appropriate shape and (b) it seems plausible—insofar as an RC stage represents one stage of neural processing—that at least two stages might be involved between the command to open a gate and the process of actually doing so.
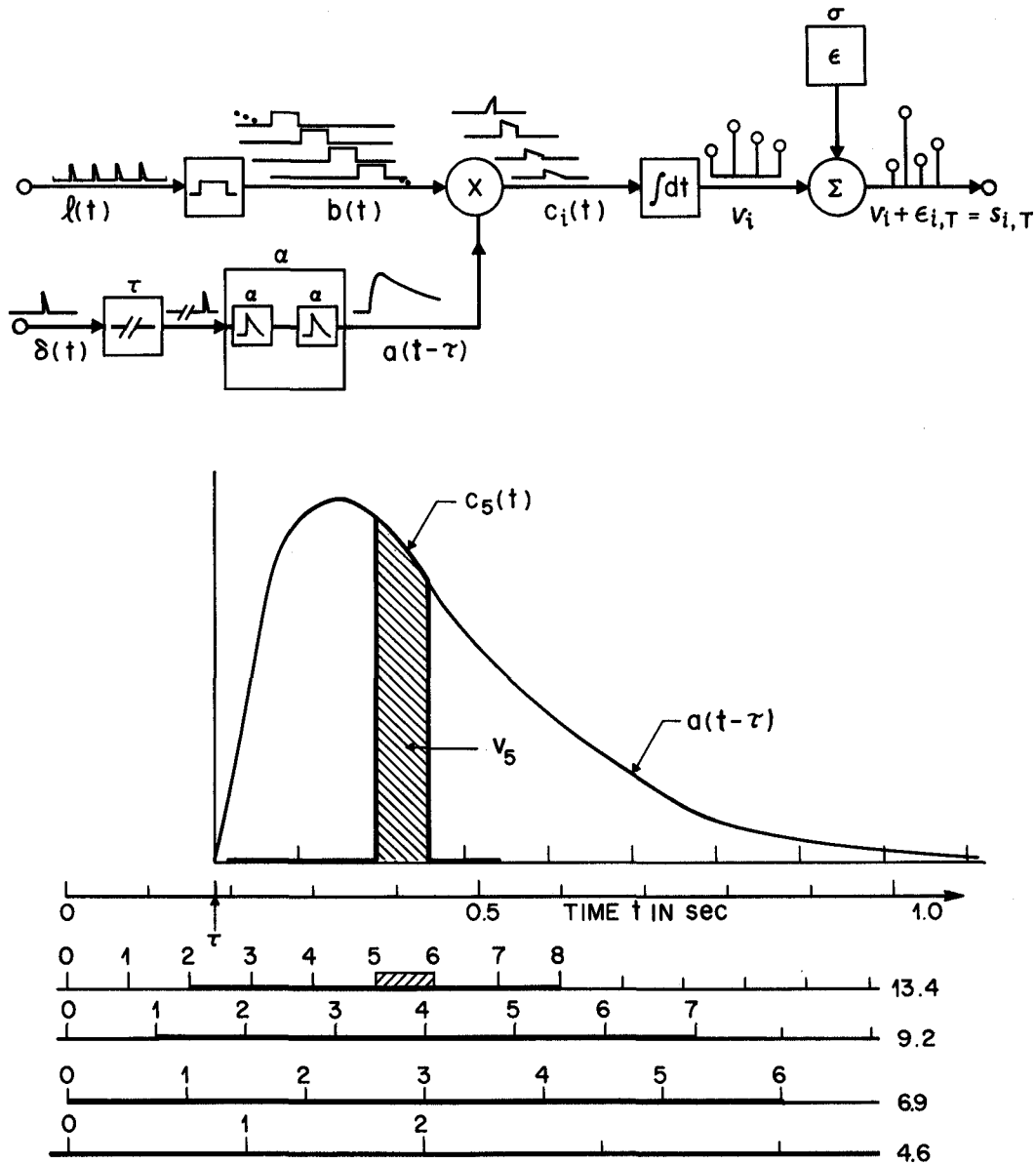
Figure 15. Attention gating model (AGM). Block diagram (top): Boxes represent components with function indicated; arrows indicate the direction of signal flow; graphs indicate outputs of the components for the inputs indicated at extreme left. Stimulus input $l(t)$ is represented by the luminance as a function of time of successive numerals; $b_i(t)$ represents persistence function (the tracing represents $b_i(t)$ for successive numerals i); $\delta(t)$ represents occurrence of target at time $t = 0$; τ is fixed delay (dependent on target) of the attention "trigger" pulse; α represents the time constants of the filters that shape the attention gating function $a(t - \tau)$; $\times$ indicates a gate that multiplies its inputs; $c_i(t)$ indicates central availability of numeral i; $\int dt$ represents the integration of information about numeral i (Equation 7); v_i represents the precedence of numeral i; ϵ represents error or noise, a Gaussian random variable with unit variance over trials T that is added independently to each v_i; the scale of $\epsilon_{i,T}$ relative to strength is determined by σ, which for computational convenience is incorporated into the definition of $a(t - \tau)$ in Equation 8; Σ represents a component that sums its inputs; $v_i + \epsilon_{i,T} = s_{iT}$ represents the model's output on the trial. Graphs (bottom): The attention gating function $a(t - \tau)$ as a function of time t. One second of $a(t - \tau)$ after onset of the target at $t = 0$ is shown, with parameters for subject AR. The attention gate begins to open at $t = \tau$. The lower abscissi show numeral onset times (t_i, Equation 7) at different presentation rates (13.4, 9.2, 6.9, and 4.6 numerals/s). The abscissi are thickened to indicate the critical sets of numeral positions, which extend back to −214 ms at the 4.6 rate (not shown). The peripheral availability $b_5(t)$ of the 5th numeral position (at the 13.4 rate) is indicated by the cross-hatched rectangle on the abscissa labeled 13.4 under the assumption that peripheral availability $b_i(t)$ is 1.0 during the interval [t_i, t_{i+1}] and 0 otherwise. Central availability of the fifth stimulus item $c_5(t)$ is indicated by the heavily drawn trapezoidal function; precendence of this item v_5 is indicated by the cross-hatched area under $c_5(t)$.

Table 1

Parameters and the Percentage of Variance in Item Scores, P_i (1), and Order Scores, P_{iBj} That Is Accounted for by the Attention Gating Model

	Subject		
Parameter[a]	AR	GL	AK
α	133	161	183
τ_u	294	245	255
τ_c	265	237	271
τ_{Sq}	222	208	219
σ	7.92	9.03	8.25
$\tau_{9.2}$ [b]	0	+25	+94[c]
$\tau_{6.9}$	0	-35	-22
$\tau_{4.6}$	0	-55	(-72)[d]
% VAR P_i (1)	81	86	88
% VAR P_{iBj}	84	90	89

[a] Parameters are α = time constant of attention gate; τ_{target} = detection delay for targets u, c, Sq; σ = noise scaling; τ_{rate} = gate delay due to numeral rate. All parameter values are in ms except σ which is in (ms × intensity)$^{-1}$.

[b] $\tau_{13.4} + \tau_{9.2} + \tau_{6.9} + \tau_{4.6} = 0$; thus, there are only three independent τ_{rate}.

[c] Only three numeral rates were run for subject AK: 9.2/s, 6.9/s, and 5.6/s.

[d] This parameter is not independent (see note b); numeral rate was 5.6/s for AK.

Estimating Parameters of the Attention Gating Model (AGM)

The functional forms—Gamma function for $a(t - \tau)$, rectangular persistence function for $b(t)$, Gaussian for $\epsilon_{i,T}$—were assumed and not formally estimated. To estimate parameters of the AGM, we assume that all of the parameters remain invariant between conditions except the delay, τ, which varies with target difficulty. Best fits (least squares) were obtained with the search program STEPIT (Chandler, 1965) with five free parameters for each subject: three detection parameters, τ_c, τ_u, τ_{sq}, and two attention parameters, α and σ.

Initially, AGM parameters $(\tau; \alpha, \sigma)$ were optimized for the overall mean-square prediction of the $V(i; sj, rt, tg)$ that had been estimated from the Case 5 solution for each condition. Given that $V_i = S_i$ (Equation 9), this step is quite straightforward. The resulting $(\tau; \alpha, \sigma)$ were then used to predict precedences—$\hat{V}(i)$—in the AGM, and thence (by means of the Monte Carlo procedure), the 21 independent P_{iBj} and the 7 $P_i(1)$ scores for each condition. The values of $(\tau; \alpha, \sigma)$ were then varied in small steps, the entire procedure being repeated, until a best fit to P_{iBj} and $P_i(1)$ was found. The same total weight was given to each type of score because the model predicts both. However, we chose to ignore $P_i(r)$, $r = 2, 3, 4$ in fitting because we already know that these $P_i(r)$ cannot be reconciled exactly with the P_{iBj} in any precedence model with zero variance $(\tau_T \equiv \tau)$ ARTs, and we wished to compare the AGM with the Case 5 strength model (which was based exclusively on P_{iBj}).

The five-parameter attention gating model predicts data for 3 targets × 4 rates, 12 conditions in all. The data are $P_i(r)$ (7 × 4 independent values per condition) and P_{iBj} (21 independent values

per condition), giving a total of 588 predictions. The resulting fits were reasonable for AR, with the percentage of variance accounted for averaging 84% for the P_{iBj} and 83% for the $P_i(1)$. Parameter values are shown in Table 1. The target τ values (265, 294, 222 ms) correlate well with the ARTs for these targets (386, 423, 329 ms), which are about 120 ms longer than their respective τ values. The maximum of the attention gating function, which tends to determine the ART, occurs α ms ($\alpha = 133$, 161, and 183 ms, for AR, GL, and AK) after the gate opens at $t = \tau$.

The AGM predictions are illustrated in Figures 16, $P_i(r)$; 17, P_{iBj}; and 18, ART, in which data are on the left and predicted values are on the right of each panel. The data in Figures 16 and 17 are for target Sq, with the rate identified in each panel; the fits for targets C and U were not worse. The 5-parameter AGM predictions bear a remarkable likeness to the data. The relatively worse quantitative fit of the AGM to the $P_i(r)$ scores in Figure 16 for $r = 2, 3, 4$ is expected. The same kind of $P_i(r)/P_{iBj}$ inconsistency was exhibited by the $V(i)$ estimated from the Case 5 solution separately for each condition, as noted before. (The attention gating model with only 5 parameters for all conditions cannot fit better than the 72-parameter strength model, which has 6 separate parameters for each of the 12 conditions.)

The AGM predictions for subjects GL and AK were not as good as the predictions for AR, averaging only 48% of the variance. To fit their data, it was necessary to modify the assumptions governing τ. For AR, τ was chosen to be independent of numeral rate, which corresponds to assuming that the attention shift (triggered by the target) is not influenced by the numeral stream. However, the block design allowed the subjects to vary strategy between conditions with different numeral rates, and it may have been possible for a subject to shift attention earlier when the numeral rate was slower. Indeed, motor reaction times (MRTs) did slow slightly at slower numeral rates (Sperling & Reeves, 1980), suggesting such an ART/MRT tradeoff strategy. Therefore, we allowed rate to influence τ. Because rate and target do not interact in their effects on attention reaction time (Sperling & Reeves, 1980), for target j and rate k we let $\tau_{j,k} = \tau_j + \tau_k$. Here τ_j reflects delay due to target detection (as before) and τ_k represents the additional delay (+ or −) attributable to numeral rate;

$$\sum_{k=1}^{4} \tau_k = 0.$$

The STEPIT program was then used to optimize eight parameters, as before, for predicting $P_i(1)$ and P_{iBj}; α, τ_c, τ_u, τ_{sq}, τ_{75}, τ_{109}, τ_{144}, and σ. (In this procedure, $\tau_{214} = -\tau_{75} -\tau_{109} -\tau_{144}$, so there are only three rate parameters for subjects AR and GL and two rate parameters for AK, who ran one less condition of rate.)

Results showed a mild improvement in the already good fit for AR and a considerable improvement of fit for GL and AK. The percentage of variance accounted for averaged 86 and 89 for GL and AK, respectively. Parameters are given in Table 1. The τ_c, τ_u, τ_{sq} values correlated highly with the respective mean ARTs; the τ values are about 135 ms less than the mean ARTs for GL and 173 less for AK, commensurate with their α values. Model predictions are illustrated for GL, target U, in Figures 19, $P_i(r)$; 20, P_{iBj}; and 21, ART. Fits for the other targets, and for AK, were similar or better.

Because α values were similar for the three subjects, the gating function shown in Figure 15 for AR is illustrative for GL and AK. The effect of the shift in τ at slow rates is shown best in the comparison of the ARTs in Figure 21 (GL) with those in Figure 18 (AR) in which no shift occurred.

Many variations in the assumptions of the AGM were explored. In particular, variations in assumptions concerning visual persistence $b_i(t)$ did not result in significant improvements. Partly this is because a change in one parameter or assumption can be compensated by a corresponding change in another parameter or assumption. For this reason we have opted for the simplest possible computational model. All in all, the five-parameter attention gating model for AR and the seven- and eight-parameter AGMs for AK and GL each provide accurate predictions of many
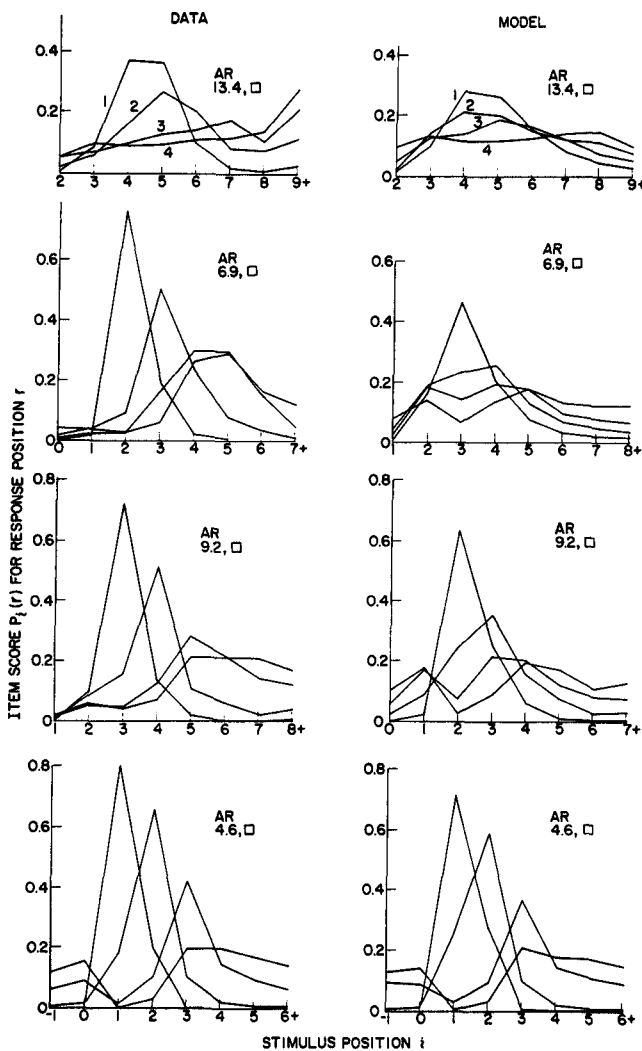

Figure 17. Observed and AGM-predicted order scores P_{iBj} for subject AR, target Sq. (Numeral rate is indicated in each panel. Stimulus Position j is given as abscissa; Stimulus Position i is the indicated curve parameter. Left panels: data; right panels: attention gating model with parameters given in Figure 16.)

hundreds of diverse data points—a very impressive data reduction.

Generalized Attention Gating Model (GAGM)

Attention and the Attention Gating Model

The AGM predictions show that a simple attention gating function $a(t - \tau)$ can provide a good description of the data. The moment at $t = \tau$ at which $a(t - \tau)$ departs from zero may be

Figure 16. Observed and AGM-predicted item scores $P_i(r)$ for subject AR, target Sq. (Numeral rate is indicated in each panel. Stimulus Position i is given on the abscissa; the target occurs at $i = 0$. The curve parameters, indicated on top panel only, are response positions, r. Left panels: data; right panels: attention gating model with two attention parameters, $\alpha = 133$ ms, $\sigma = 7.92$, and one detection parameter, $\tau_{Sq} = 222$ ms.)
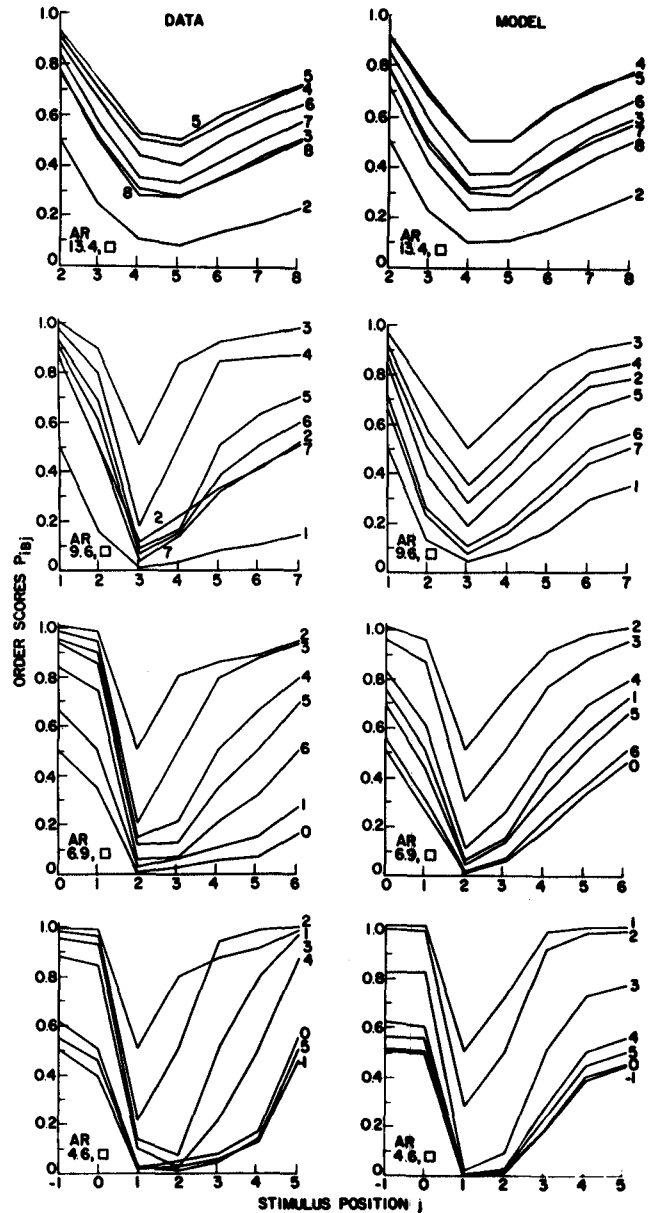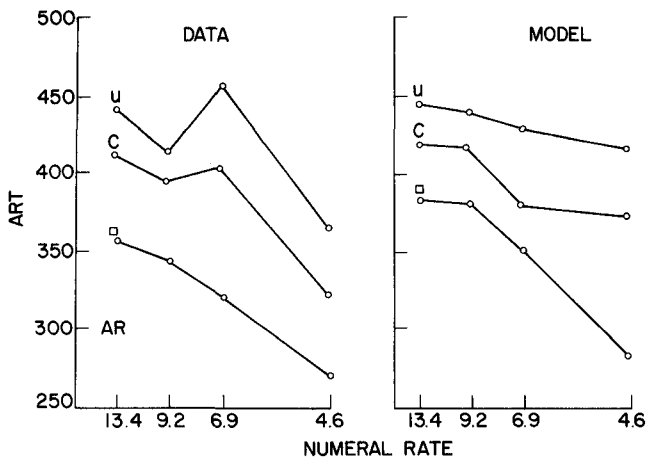
Figure 18. Observed and AGM-predicted attention reaction times (ARTs) for subject AR, all three targets, as a function of numeral rate. (Left panel: data; right panel: attention gating model with parameters as given in Figure 16. The parameters τ_u, τ_c, τ_{Sq} determine the heights of the curves; their shape is determined by the attention gating function $a(t - \tau)$.)

identified with the time at which attention begins to shift from the target to the numeral stream. Further, in the AGM, it would be natural to identify the shape of $a(t - \tau)$ (which ultimately determines the presence and ordering of the numerals in the response) with the time course of attention. Unfortunately, the time course of $a(t - \tau)$ bears no simple relationship to any conceivable time course of attention.

To illustrate the impossibility of identifying $a(t - \tau)$ directly with attention, consider a hypothetical attention-shift experiment using a recall procedure that somehow prevents use of verbal short-term memory during the stimulus presentation, but in which the numerals to be reported occur at a very slow rate (e.g., 3/s). At slow rates, four numerals should be easily reportable with nearly perfect order from visual memory. In terms of an attention shift, once attention is fixed on the numeral stream, it remains there continuously. Attention is high and constant for the relatively long period of time needed to "grab" four numerals. In the AGM, the corresponding attention gating $a(t - \tau)$ function has to slope sharply downward to the right to represent good recall order. To be output earlier, a stimulus item needs higher strength, that is, more area under $a(t - \tau)$ than its successor. Thus, in the AGM, steady attention is represented by a downward sloping $a(t - \tau)$ function rather than by a constant $a(t - \tau)$ function.

A second problem with the AGM is that it yields a unidimensional description of item strength, whereas item strength, at least at slow stimulus rates, is inherently two-dimensional. Consider two consecutive items A, B. Assume both are strong in memory (in the sense that a large amount is known about each item and the subject is sure that each has occurred), but B is even stronger than A. In terms of item (identity) strength, $S_i(B) > S_i(A)$. On the other hand, the subject might also be sure that A occurred before B; in terms of order strength S_o, the strength ordering is reversed: $S_o(A) > S_o(B)$. To reflect these two kinds of knowledge, precedence strength—which determines the

sequence of responses—should be at least a two-dimensional quantity, a combination of item and order strength components.

Ideally, an attention model would reflect these aspects of attention as we understand it and not merely serve as a computational device. Constant, steady attention would be represented by a constant-valued $a'(t - \tau)$ function (or its equivalent) in the model. Item information and order information would be represented individually. The model would show how these kinds of information are computed and how they combine to determine the selection and ordering of response items. Here, we propose such a model, a generalized attention gating model (GAGM).

Just as the strength model could be derived from the attention gating model, the attention gating model is derivable as a special case of the generalized attention gating model. In going from AGM to GAGM, however, there is no further saving in parameters nor are there any new predictions for the present set of data. Of course, in an expanded data set, parameter savings and new predictions might be possible. The reason for proposing the generalized attentional model at this point is to demonstrate that the attention gating model can be derived from a full, logically consistent attention theory that satisfies the reasonableness and the dual-process conditions. The GAGM theory is outlined below. Primes (') are used to distinguish GAGM components from their counterparts in the AGM.
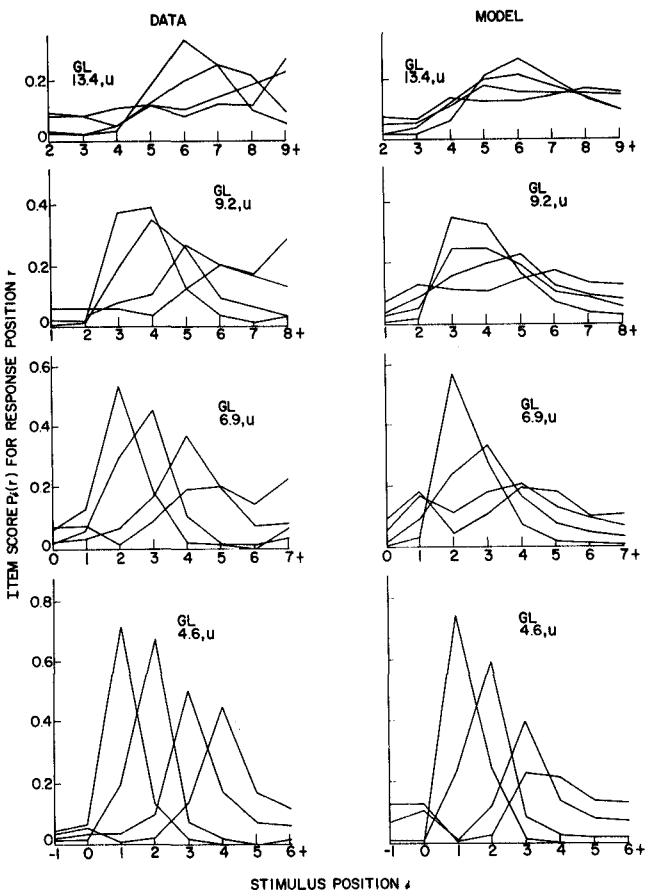

Figure 19. Observed and AGM-predicted item scores $P_i(r)$ for subject GL, target U. (Parameters of the AGM model are given in Table 1; see Figure 16.)

GAGM Assumptions

Variable attention gating function, a'(t; sj, rt, tg). There are two generalizations of the GAGM over the AGM: a potentially variable attention gating function and separate computations for item and order strength. The issue of the recall of numerals presented at slow rates is critical for an attention model because four such numerals, which can be recalled perfectly, occupy a longer time interval than is spanned by a fixed attention gating function. Either the attention function is stretchable, or recall of slowly presented numerals requires several successive acts of attention, or both may occur. An experiment to determine temporal parameters of successive attentional windows in the atten-
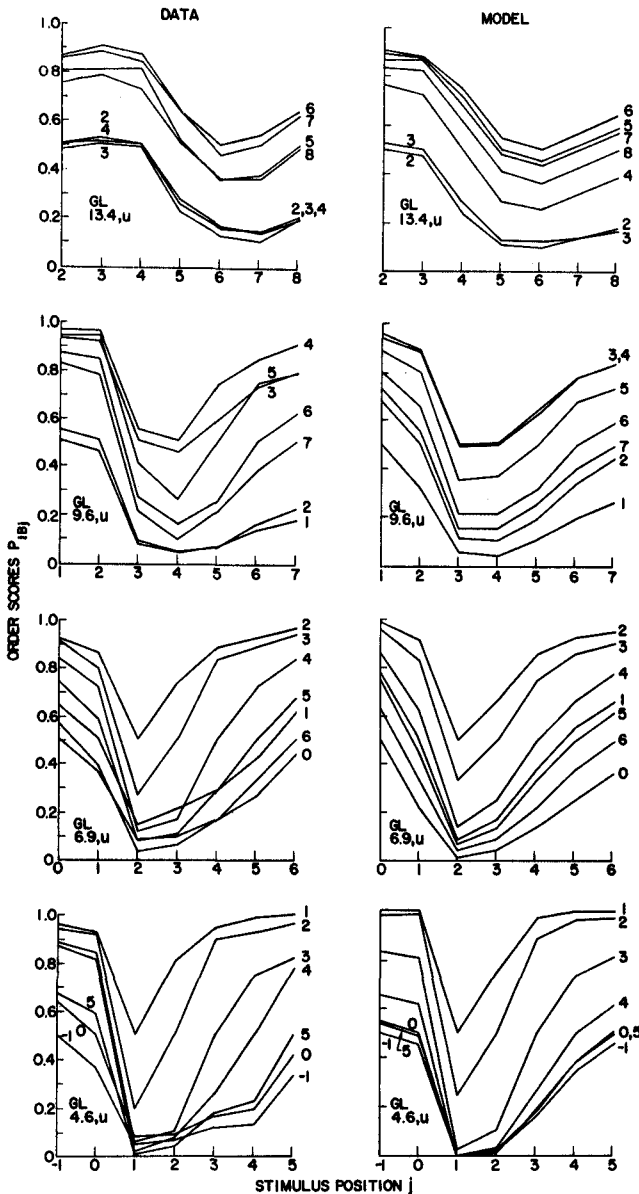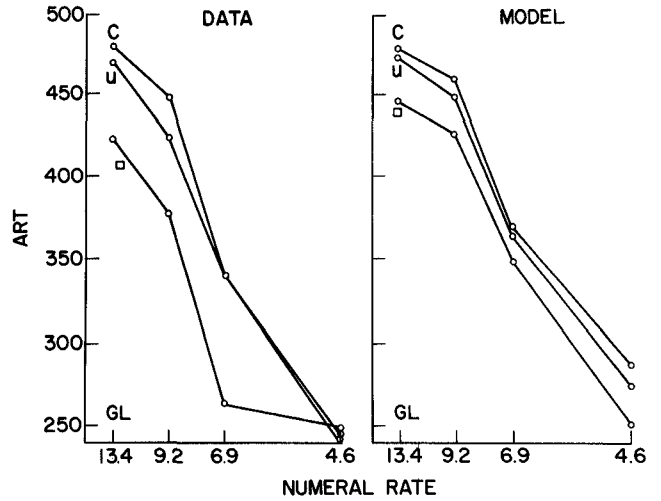

Figure 21. Observed and AGM-predicted attention reaction times (ARTs) for subject GL, all three targets, as a function of numeral rate. (AGM parameters are given in Table 1; see Figure 18.)

tion-shift paradigm (Weichselgartner, 1984) indicates that short sequences of digits at 3/s or 4/s could fall within a single attentional window or into successive windows. We explore here the limits of a single attentional window.

Variable attention gate. In the AGM, the attention gate operates as follows: At the moment τ of target detection, a signal (an impulse) is sent to the gate controller to open the gate. The gate does not have a perfect frequency response, so in response to the impulse, it opens and closes with a time course described by $a(t - \tau)$, the impulse response of the gate control system. The simplest generalization of $a(t - \tau)$ is to simply remove all constraints on its shape except that it be nonnegative. That is, $a'(t; sj, rt, tg)$ is allowed to vary with all factors, subject, rate, and target, instead of varying only with subject, as in the AGM. It will then be possible to determine empirically the optimal shape of $a'(t; sj, rt, tg)$ and to what extent the shape remains constant for all conditions. Alternatively, another generalization of $a(t - \tau)$ that we have explored is to allow the gate control signals to be pulses of longer duration than the impulses of the AGM (Figure 15, top). This generalization preserves the physical intuition of the AGM and is much easier to test because it involves only one new parameter. Control pulses of various widths generate a family of gating functions $a'(t, w)$ where w is the width parameter. In fact, for the available data, the addition of this pulse-width parameter did not improve predictions, so we will not further consider this particular constrained variable-attention-gate model.

Separate order and item channels. The second generalization in the GAGM is the addition of an explicit order channel to the item channel of the AGM. It is supposed that these two kinds of information, item and order, might be separately utilized in some experimental tasks. To generate predictions for the present experimental procedure, however, the GAGM combines item and order information by adding them together. Quite generally, the perceptual system seems to reach decisions by linearly adding the appropriately scaled evidence for and against a perceptual hypothesis. The addition of item and order information to de-
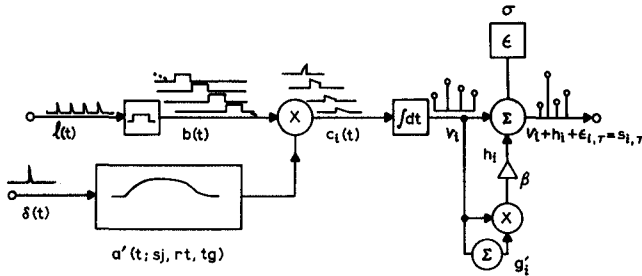

Figure 20. Observed and AGM-predicted order scores P_{iBj} for subject GL, target U. (Parameters of the AGM model are given in Table 1; see Figure 17.)

Figure 22. The generalized attention gating model (GAGM). The entire gated pathway from $l(t)$ to $s_{i,T}$ is the same as in the attention gating model of Figure 15, top; the modifications in the gating function and the order channel are as follows: First, the attention gating function $a'(t; sj, tg, rt)$ is a function of time with parameters determined by the subject sj, the numeral rate rt, and the target tg. Attention gating is triggered by target presentation; detection delay is incorporated into the parameter tg of a'. Second, the order channel h'_i (absent in AGM) combines two pathways: v'_i, strength of the current item, and g'_i, cumulative strength of earlier items. The Σ in the order channel indicates the sum of the first $i - 1$ strengths; the component β indicates multiplication by a negative constant $-\beta$. The final sum $s'_{i,T} = v'_i + h'_i + \epsilon'_{i,T}$ is the output on a trial. Given any set of S_i derived from data or an AGM, the GAGM can be backsolved for a'.

termine precedence is analogous to examples of the addition of perceptual evidence reviewed by Sperling, Pavel, Cohen, Landy, and Schwartz (1983) and by Dosher, Sperling, and Wurst (1984, 1986).

The GAGM with the generalized attention gating function and with an order channel is shown in Figure 22. The main gated-information path—input $l(t)$ to output $s_{i,T}$ of Figure 22— is identical to the path in Figure 15, top. Shown below the main path in Figure 22, the gate control path—$a'(t; sj, rt, tg)$—has been generalized, and the order path has been added to the AGM of Figure 15, top. We show how to choose $a'(t; sj, rt, tg)$ so that the generalized GAGM (Figure 22) becomes computationally equivalent to the particular AGM described above (Figure 15, top).

Order information, h'_i. Order information associated with an item i is assumed to have two properties: (a) It is defined relative to other items in the attentional window (as contrasted to being defined on an absolute time scale), and (b) it is assumed to be proportional to the amount of identity information already available about the item. (When very little is known about an item, e.g., whether or not it occurred in the critical set, then, in the present paradigm, it would be impossible to know precisely where in the critical set the item did occur.)

The assumptions about item and order information are paradigm specific. The items in this theory are numerals that are presented three times in the course of a single trial. An absolute time tag associated with each occurrence of an item would be a very inefficient way of performing the recall task. The relations between the items captured within an attentional window are what is important, and these relations are what the model attempts to capture.

In order to determine the implications of the assumptions about item and order information, they must be elaborated quantitatively. Let v'_i be the precedence strength of an item in

the GAGM, defined similarly to strength v_i in the AGM (Figure 15). A measure of relative order for item i is g'_i, the cumulative precedence strength of all the items up to but not including the present item

$$g'_i = \sum_{j=0}^{i-1} v'_j.$$

The order information h'_i (a value or tag associated with item i) is the product of the item's strength and the measure of its relative order

$$h'_i = \beta s'_i g'_i = \beta v'_i \sum_{j=0}^{i-1} v'_j \qquad (10)$$

where β is a negative constant. The strength s'_i on a trial includes error noise $\epsilon'_{i,T}$ thus $s'_i = v'_i + h'_i + \epsilon'_{i,T}$. The expected value of noise is zero, so, using capital letters to denote the expected value of the various quantities, $S'_i = V'_i + H'_i$. Substituting for h'_i in Equation 10 and factoring gives

$$S'_i = V'_i[1 + \beta G'_i] = V'_i[1 + \beta \sum_{j=0}^{i-1} V'_j]$$

$$= f_i(V'_1, V'_2, \cdots V'_i). \qquad (11)$$

Equivalence of the AGM and the GAGM. Equation 11 represents a set of n equations in which $i - th$ strength predicted by GAGM, S'_i, is computed from all prior precedences V'_j, $j = 0, i$. The normal way to use these equations would be to compute the V'_i from the inputs by a computation analogous to computation of the AGM strengths by Equation 7—the difference is in the gating function $a'(t - \tau)$. The n precedences $[V'_i]$ would then be used to compute the strength outputs $[S_i]$.

The set of equations (11) can also be solved in the reverse direction (see Appendix). That is, given a set of S'_i derived, for example, from data, Equations 11 can be solved for V'_i and $a'(t)$. The backsolution of Equation 11 shows how to choose the attention gating function $a'(t)$ to predict precisely any set of observable data, V'_i. Therefore, if there were complete freedom to choose $a'(t)$, the GAGM would be a perfect predictor of V'_i. Moreover, it could perfectly predict V'_i for any reasonable value of the order parameter β. Obviously, evaluating the GAGM requires additional constraints. As before, we have explored two constraints. First, the effects of target and numeral rate are captured simply by a time-shift τ in a'; that is, $a'(t; sj, rt, tg)$ reduces to $a'(t - \tau; sj)$, where τ is a function of rt and tg. Second, the constant β remains the same over different numeral rates and targets. Note that for $\beta = 0$, the order channel disappears and the GAGM essentially reduces to the AGM. Nonnegative solutions for $a'(t)$ are possible only when $0 > \beta > -0.5$, and solutions with β infinitesimally larger than -0.5 have the largest contribution of the order channel. Data predictions equivalent to the AGM were obtained with $a'(t - \tau; sj)$ and $\beta = -0.49$. The derived $a'(t - \tau; sj)$ of the GAGM look like stretched versions of the corresponding $a(t - \tau; sj)$ of the AGM. The order channel (h_i) contributes less than 12% of the variance of the predicted precedences. Whether the small role of the order channel represents an interesting fact about visual memory or a defect of the GAGM will have to be resolved by further experimentation.

The predictions of the AGM and GAGM are strengths (V_i, precedences) for items occurring in a particular position i of the stimulus stream. The AGM and GAGM strengths, derived for all conditions and subjects from only a few parameters, logically cannot be better than V_i derived from the strength model that were optimized separately for each subject and condition. The power of the AGM is its ability to recover the many, many V_i of the strength model with only a very small number of parameters. From our additional explorations of the GAGM, we can conclude the following. First, predictions equivalent to those of the AGM can be derived from a more logically motivated attention model, with no more free parameters than the AGM. Second, the freedom to vary the attention gate does not significantly improve fits to the existing data, as the very good predictions of the AGM have already suggested. Third, under the conditions of the present experiments, most precedence information is carried indirectly by the channel that codes identity; only a small fraction of order information is carried by an order channel. Fourth, proving that the course of attention is truly described by the logically defensible $a'(t - \tau)$ of the GAGM rather than by the computationally equivalent gating function $a(t - \tau)$ of the AGM, or by yet some other function, will require other sources of evidence.

While the GAGM is not required by the present data, it nevertheless is a significant improvement over the AGM. Its attention gating function could correspond to the introspectively observed and logically defensible time course of attention. The GAGM is general enough to apply to a wide range of conditions beyond those of the immediate experiment. The components are designed to reflect processes that can readily be embodied in neurons, and therefore it suggests a possible physiological basis for attention. For example, the order channel uses a shunting (gain-control) network to compare the present input to the aggregated recent inputs. Precisely this general principle has been proposed for lower level neural networks that detect luminance pulses or flicker; it is the temporal analogue of the spatial center-surround receptive field organization (Sperling & Sondhi, 1968); and it is widely used in higher level neural models of memory and control processes (e.g., Grossberg, 1978a, 1978b). Finally, for reasonable choices of its parameters, the general attention model reduces to the computationally efficient attention gating model.

The Attention Gating Model and Three Attention Experiments

Spatial shifts of visual attention. The RSVP attention shift paradigm was developed to answer the questions about how attention shifts between locations in the visual field. The answer it suggests is that initially an attention gate opens at one location (the expected location of the target). After detection, the first gate closes and a second gate opens at the location from which items are to be reported. There is nothing in the attention gating theory to exclude a "searchlight" theory of attention in which attention moves continuously from location to location, illuminating intermediate locations as it passes over them. However, the gating process observed here is much more suggestive of a faucet-gate theory, in which a gate is opened and closed first at one location and then at another, with no particular dependence on the distance between the locations or requirement to open at intermediate locations. Weichselgartner et al. (1985a, 1985b) used

the RSVP attention shift paradigm to investigate the effect of distance on the time for attention shifts. Subjects shifted attention from a peripheral letter stream (in which a target was embedded) to a centrally fixated numeral stream (from they had to report numerals). There was no effect of distance between streams on the latency of attention shifts, nor was there any effect of a visual obstacle placed between the locations, directly in the path that attention presumably had to cross. Their null results are evidence against a continuous searchlight process and in support of an attention theory in which a gate simply opens at a second location while closing at the first.

Partial reports and visual short-term memory. In a procedure designed to measure the decay of very short-term visual memory, Sperling (1960) presented subjects with a brief flash of a 3×4 letter array, and subsequently with a tone (chosen randomly on each trial to be of high, middle, or low frequency) that instructed them which row to report (top, middle, or bottom, respectively). When a tone occurred simultaneously with the end of the flash, subjects' partial reports were very accurate. As the tonal delay approached 0.5 s, partial report accuracy diminished to that of whole reports (reports of the entire array, made on control trials). Sperling interpreted the data as indicating that subjects momentarily had available to them a very short-term visual memory of almost the entire 12-letter array, and that this memory decayed within a fraction of a second. Dubbed "iconic memory" by Neisser (1967, p. 15), it has spawned extensive study (see reviews by Coltheart, 1980, and Long, 1980).

The original motivation for the partial report procedure was the observation that subjects could not retain more than about 4 or 5 items in a memory that persisted until their report could be made, even though they knew many more items had been presented. The partial report allowed them to "gate" the items from iconic memory into a limited-capacity longer-term memory according to the demand of the tone cue and thereby to demonstrate their very large, very short-term memory capacity. The time taken to open an attentional gate at the designated row in response to the tonal cue was confounded with the estimated iconic decay time because there was no independent measure of the attention gating process. The RSVP attention shift paradigm offers an independent measure. Indeed, Weichselgartner (1984) used the RSVP attention shift paradigm to measure attention gating in response to a tonal stimulus and found it to have a somewhat shorter latency than the gating responses to the embedded letter targets that were the objects of the present study. Now that independent measures of attention gating can be obtained, it becomes feasible to attempt to disentangle the iconic memory and attentional components in partial reports.

Temporal order judgments (TOJs). In the classical TOJ paradigm, a subject is presented with two stimuli, such as a brief tone burst and a brief light flash, and must say which occurred first—the "complication experiment" (Dunlap, 1910). In more recent conceptualizations of TOJs, the pair of stimuli are assumed to arrive on independent peripheral channels, which may be in the same or different sense modalities. Stimuli arriving on the two channels are judged by a memoryless central decision mechanism that responds with various amounts and types of uncertainty (depending on the decision model) to whichever channel delivers its input first. Memory does not contribute to the central decision, because the information fed by each input channel to

the decision mechanism consists of only a single time value—the arrival time. Attention influences temporal order judgments (the law of prior entry; Stone, 1926; Titchener, 1908) either by speeding peripheral processing of the attended stimulus (Sternberg & Knoll, 1973) or by determining when information in sensory storage is to be sampled by the decision mechanism (Schmidt & Kristofferson, 1963).

Although we do not consider TOJs extensively here, the attention gating model suggests a substantially different interpretation. According to the AGM, the subject performs the light-tone TOJ task by listening for one stimulus, say the tone, as a target. On detecting it, the subject gates the input from the other, visual stimulus into a visual short-term memory. (Alternatively, the subject could detect the visual stimulus and gate the tone into auditory memory.) The contents of memory then consist of several brief episodes of blank screen, a brief flash (that persists because it is not interrupted), and more blank episodes. The decision task involves setting a criterion for how much blank time, if any, should precede the flash in memory in order for the physical stimuli to be judged as simultaneous. If, on a trial, more than the criterion amount of blank space precedes, the tone is judged earlier; otherwise the flash was earlier.

The problem in evaluating a gating attention theory of temporal order judgments with classical data is that the blank events in memory are not easy to identify and to report. This is where RSVP of the numeral stream (as a channel) has the advantage over a channel that contains only one nonblank event. Each event in the gated numeral channel has an identity that can be used to externalize the contents of memory and thereby to give insight into the memory's mechanics. Whether the gating model that here describes temporal order judgments within the numeral stream indeed describes classical temporal order judgments is a question for further research.

Summary and Conclusions

Shifting attention from the letter stream to the numeral stream produces numeral reports with three chief characteristics: clustering, disorder, and folding.

Clustering

The numerals reported from the numeral stream are typically chosen from a cluster of numeral positions centered about 400 ms after target occurrence, with a range of from 200 ms to 600 ms. Subjects make most of their reports from this cluster, whether reporting four numerals or only one (Reeves, 1977). The first reported of the four numerals (or the single reported numeral) provides a measure of an attentional reaction time, that is, a measure of the time taken to shift attention from the letter stream to the numeral stream (Sperling & Reeves, 1980). (Like a motor reaction time, the attention reaction time includes several component latencies, such as the time taken to identify the target and first numeral, as well as the latency for the shift of visual attention.) The clustering of the reported numerals around the first reported numeral is explained by the quick opening, followed by a quick closing, of an internal attention gate that allows numerals to flow into a visual short-term memory (VSTM).

Disorder

The order of the clustered numerals is independent of stimulus order at fast numeral rates (P_{co} near to chance). Scarborough and Sternberg (1967) similarly found that their subjects could not reliably report which numeral followed a target numeral (poor order information) when they viewed only one stream of characters, even though the subjects could reliably detect the target numeral (good item information). Not only are adjacent numerals interchanged, but so too are numerals from positions several tenths of a second apart, as shown by the $D(x)$ curves in Figure 6. Such disorder suggests that stimuli have feature representations in VSTM that are adequate for recognition but have temporal representations that are not veridical.

Folding

The high probability of disorder found in the clustered numerals can be explained by a strength model in which numerals that occupy different stimulus positions have different precedences $V(i)$, and these precedences are folded. That a single scale of precedences can account for the order scores was shown by the nonparametric quad test. That the precedences are folded around a central position is shown by the inverted-U-shaped $V(i)$ curves in Figure 12 and the corresponding U-shaped P_{iBj} curves in Figures 7, 8, and 9. As an example, the order of precedences in the top left panel of Figure 7 was not the veridical order, 1, 2, . . . , 7 but rather the order 6, 5, 7, 8, 4, 3, 2, which is folded around Position 6.

Attention Gating Model

We propose attention gating as the explanation of clustering, disorder, and folding. Clustering occurs because attention is allocated to the numeral stream only briefly. Disorder and folding reflect the same mechanism. Items that are present 400 ms after the target receive the most attention and tend to be represented most prominently in VSTM. Items that occur before and after 400 ms receive less attention and therefore have weaker representations in VSTM. At the fastest numeral rates, response ordering is determined entirely by the amount of attention that items receive. At slower rates, there may also be some additional temporal order information. A test of this hypothesis, embodied in the generalized attention gating model, awaits empirical measurements to discriminate item-strength from temporal-order components.

Visual Memory

Clustering, disorder, and folding occur in a memory described as "visual" by the subjects. Further, the subjects have an illusion of correct order in their reports. They believe they have reported items in the correct order and are surprised by the continuing discrepancy between their reports and the feedback of actual presentation order given on each trial. They say that they do not report from a confused memory for order after the trial but, rather, that they accurately report the numerals as they "see" them. Because items are displayed rapidly, each superimposed on the last, they cannot be retained in a retinally based memory

(e.g., Sakitt, 1975); we attribute the results to a postretinal, visual, short-term memory (e.g., Kaufman, 1977; Phillips & Baddeley, 1971; Scarborough, 1972).

References

Beck, J., & Ambler, B. (1973). The effects of concentrated and distributed attention on peripheral acuity. *Perception and Psychophysics, 14*, 225–230.

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of response. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgement and choice*. San Francisco: Holden-Day.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Chandler, J. P. (1965). *STEPIT*. Bloomington, IN: Quantum Chemistry Program Exchange, Indiana University Chemistry Department.

Cherry, C. (1953). Some experiments on the reception of speech with one and with two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Cohn, T., & Lasley, D. (1974). Detectability of a luminance increment: Effect of spatial uncertainty. *Journal of the Optical Society of America, 64*, 1715–1719.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics, 27*, 183–228.

Dosher, B. A., Sperling, G., & Wurst, S. A. (1984). Stereopsis versus proximity-luminance covariance as determinants of perceived three-dimensional structure. *Journal of the Optical Society of America A: Optics and Image Science, 1*, 1254.

Dosher, B. A., Sperling, G., & Wurst, S. A. (in press). Tradeoffs between stereopsis and proximity luminance covariance. *Vision Research*.

Dunlap, K. (1910). The complication experiment and related phenomena. *Psychological Review, 17*, 157–191.

Eriksen, C. W., & Collins, J. F. (1969). Temporal course of selective attention. *Journal of Experimental Psychology, 80*, 254–261.

Eriksen, C. W., & Hoffman, J. E. (1972). Some characteristics of selective perception determined by vocal reaction time. *Perception and Psychophysics, 11*, 169–171.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington DC: Winston.

Grindley, C. G., & Townsend, V. (1968). Voluntary attention in peripheral vision and its effects on acuity and differential thresholds. *Quarterly Journal of Experimental Psychology, 20*, 11–19.

Grossberg, S. (1978a). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17*, 199–219.

Grossberg, S. (1978b). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–371). New York: Academic Press.

Helmholtz, H. (1909). *Handbuch der physiologischen optik* [Handbook of optical physiology] (Vol. 1, 3rd ed.). Leipzig, East Germany: Voss.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception and Psychophysics, 25*, 319–327.

James, W. (1890). *The principles of psychology* (Vol. 1). New York: Holt.

Jonides, J. (1981). Voluntary vs. reflexive control of the mind's eye movement. In J. B. Long & A. B. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Jonides, J. (1983). Further toward a model of the mind's eye movement. *Bulletin of the Psychonomic Society, 21*, 247–250.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kaufman, J. (1977). *Three forms of short-term visual memory*. Unpublished doctoral dissertation, New York University.

Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 259–279). Hillsdale, NJ: Erlbaum.

Landauer, T. K. (1962). Rate of implicit speech. *Perception and Motor Skills, 15*, 646.

Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception and Psychophysics, 10*, 85–89.

Liss, P. (1968). Does backward masking by visual noise stop stimulus processing? *Perception and Psychophysics, 4*, 328–330.

Liss, P., & Reeves, A. (1983). Interruption of dot processing by a backward mask. *Perception, 12*, 513–529.

Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin, 88*, 785–820.

Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Nickerson, R. (1965). Response times for "same"–"different" judgments. *Perceptual and Motor Skills, 20*, 15–18.

Nickerson, R. (1967). Expectancy, waiting time and the psychological refractory period. In A. F. Sanders (Ed.), *Attention and performance III (Acta Psychologica*, Vol. 27), pp. 23–24. Hillsdale, NJ: Erlbaum.

Nickerson, R., & Burnham, D. (1969). Response times with nonaging foreperiods. *Journal of Experimental Psychology, 79*, 451–457.

Norman, D. A. (1967). Temporal confusions and limited capacity processors. *Acta Psychologica, 27*, 293–297.

Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science, 22*, 73–74.

Pierce, J. R., & Karlin, J. E. (1957). Reading rates and the information rate of a human channel. *Bell Systems Technical Journal, 36*, 497–616.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. I. Pick, Jr., & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10–15.

Reeves, A. (1977). *The identification and recall of rapidly displayed letters and digits*. Unpublished doctoral dissertation, City University of New York.

Reeves, A., & Sperling, G. (1983). Attentional theory of order information in short-term visual memory. *Mathematical studies in perception and cognition* (No. 83-7, p. 52). New York: New York University, Department of Psychology.

Reeves, A., & Sperling, G. (1984). Visual temporal order perception. *Investigative ophthalmology and visual science* (Association for Research in Vision and Ophthalmology Supplement), *25*, 69.

Remington, R. W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology and Performance, 6*, 726–744.

Sakitt, B. (1975). The locus of short term visual storage. *Science, 190*, 1318–1319.

Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology, 3*, 408–429.

Scarborough, D. L., & Sternberg, S. (1967, April). *Processing items and their order in sequential visual displays*. Paper presented at the meeting of the Eastern Psychological Association, Boston.

Schmidt, M. W., & Kristofferson, A. B. (1963). Discrimination of successiveness: A test of model attention. *Science, 139,* 112–113.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.

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

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

Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 14–22.

Snodgrass, J. G. (1969). Foreperiod effects in simple reaction time: Anticipation or expectancy? *Journal of Experimental Psychology Monographs, 79* (3, Pt. 2).

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498).

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31.

Sperling, G. (1970). Extremely rapid visual scanning. *Bulletin of the British Psychological Society, 23,* 58.

Sperling, G. (1984). A unified theory of attention and signal detection. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 103–181). New York: Academic Press.

Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science, 174,* 307–311.

Sperling, G., & Dosher, B. A. (in press). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* New York: Wiley.

Sperling, G., Pavel, M., Cohen, Y., Landy, M. S., & Schwartz, B. J. (1983). Image processing in perception and cognition. In O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 359–378). New-York: Springer-Verlag.

Sperling, G., & Reeves, A. (1976). Reaction time of an *Un*observable response. *Bulletin of the Psychonomic Society, 10,* 247.

Sperling, G., & Reeves, A. (1977, August 23). *Attentional theory for order information in short-term visual memory.* Paper presented at the Tenth Annual Mathematical Psychology Meeting, University of California, Los Angeles.

Sperling, G., & Reeves, A. (1978). Measuring the reaction time of a shift of visual attention. *Investigative ophthalmology and visual science* (Association for Research in Vision and Ophthalmology Supplement), *17,* 289.

Sperling, G., & Reeves, A. (1980). Measuring the reaction time of an unobservable response: A shift of visual attention. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 347–360). New York: Academic Press.

Sperling, G., & Reeves, A. (1983). Gating model of visual attention. *Bulletin of the Psychonomic Society, 17,* 354.

Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America, 58,* 1133–1145.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629–685). New York: Academic Press.

Sternberg, S., & Scarborough, D. L. (1969). Parallel testing of stimuli in visual search. In *Proceedings of international symposium on visual information processing and control of motor activity* (pp. 179–188). Sofia, Bulgaria: Bulgarian Academy of Sciences.

Stone, S. A. (1926). Prior entry in the auditory-tactual complication. *American Journal of Psychology, 37,* 284–287.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34,* 273–386.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan.

Tsal, Y. (1983). Movement of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194–214.

Weichselgartner, E. (1984). *Two processes in visual attention.* Unpublished doctoral dissertation, New York University.

Weichselgartner, E., Sperling, G., & Reeves, A. (1985a). Effects of concurrent tasks, distance, and visual obstacles on shifts of visual attention. *Proceedings of the Eastern Psychological Association, 56,* 72.

Weichselgartner, E., Sperling, G., & Reeves, A. (1985b). The effects of distance, concurrent tasks, and visual obstacles on shifts of visual attention. *Mathematical studies in perception and cognition* (No. 85-3, p. 36). New York: New York University, Department of Psychology.

Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory and Cognition, 8,* 521–527.

Wundt, W. (1912). *Introduction to psychology* (R. Pinter, Trans.). London: Allen.

Appendix

Technical Issues in Simulation, Estimation, Scoring, Foreperiods, Persistence, and the Solution of Equation 11

Monte Carlo Simulation Tests of the Quadruple Condition

It is not known what proportions of violations require rejection of Equation 3, inasmuch as the underlying distributions are not known analytically. The validity of Equation 3 was tested conservatively with a Monte Carlo procedure (see *A Monte Carlo simulation*) in which the $V(i)$ were estimated from the data and H was the cumulative normal distribution function. In each simulation, as many trials were run as had been run experimentally (typically 400) to estimate the proportion of violations that would occur in a limited sample when Equation 3 was known to hold exactly. To build up an empirical distribution for this theoretical proportion, each simulation was repeated 100 times with a different random number kernel on each occasion. From the 100 repetitions, a mean proportion of theoretical violations with an associated standard error could be found. The mean proportions of violations of the quad condition were typically 3% at the fastest numeral rate and 2% at all slower rates. In the data, the mean proportions of quadruple violations at the fastest rate were typically 4% and were never more than 1.4 standard errors greater than the simulated proportions. Therefore Equation 3 cannot be rejected at the fastest numeral rate. However, the proportions of quadruple violations in the data were 5% at the medium and 7% at the slow rates. These proportions are small but are nevertheless more than two standard errors greater than the theoretical proportions.

Optimality of the Estimated Precedences, V(i)

The parameters $V(i)$ provided a good fit to the data but were not necessarily optimal, because they were calculated from the data under the independence assumption of Thurstone's Case 5 (Equation 6). To show that the $V(i)$ values used were in fact close to optimal, the predicted order scores $\hat{P}_{iBj}$ were entered into Equation 6 to generate a second-order set of precedences $V''(i)$. If the model were perfect, $V(i)$ and $V''(i)$ would be identical. They actually correlated nearly perfectly ($r = .98$ or better), although the $V''(i)$ were very slightly less in absolute magnitude than the $V(i)$.

Possible Scoring and Report Artifacts

The strength model has small failures but provides a good overall set. However, the model does provide a good overall fit. Is this basic finding artifactual? We have already shown that guessing and forgetting do not explain the item and order information accounted for by the model. Here we consider two other possible artifacts: scoring procedure and response length.

Scoring procedure. The possibility exists that our analysis procedure forced item and order scores to be a function of a single variable (what we call *precedence*). Order scores $n(iBj)$ include type (b) events (see Results) in which the subject reports a numeral from Position i but not from Position j. Because we included type (b) events, items rarely reported (low P_i) must have lower precedences than those frequently reported. The analysis thus forces a degree of covariation between item scores and precedences, although how much covariation is not known exactly. However,

if type (b) events are excluded from P_{iBj}, item scores (P_i and P_j) are mathematically independent from P_{iBj} scores. (That is, any empirical relation between them is possible, so long as neither item score is exactly zero.) In fact, when type (b) events were excluded, order scores were found to be similar to the order scores reported here (Reeves, 1977) and we used the type (a) score to illustrate empirical results P_{co} (probability of correct order) and $D(x)$ (the index of disorder as a function of distance). Hence, our scoring procedure, which is appropriate if the strength model is correct, did not artifactually generate the main results (clustering, disorder, and folding).

Response length. A second methodological objection stems from the limited response length. Had the response been longer, it might have included reports that would have invalidated the model. For example, there might be items that either (a) were reported early but infrequently (i.e., low P_i but high precedence) or (b) were reported late (low precedence) but frequently. Both possibilities would reject the strength model, in which $V(i)$ and P_i are positively related. Type (a) items would be reported whenever they occurred, because of their high precedence, and so their absence from the data cannot be an artifact of short response length and actually supports the strength model. Type (b) items would be much less often reported, because their low precedences would exclude them from a short response, and so their absence from the data is not conclusive. However, subjects reported that they had to pay attention to the numeral stream only briefly to avoid having their memory of the numerals overwhelmed by later-coming stimuli, and so it does not appear likely that the short response length excluded enough potential response candidates for the test of the model to be seriously invalidated.

Foreperiod Effect

The target occurred with equal probability at Positions 7 to 20 in the letter stream. Foreperiod, the position of the target within the stream, has a significant influence on motor reaction time (MRT) and attention reaction time (ART), presumably because late-occurring targets are more predictable ("aging," see Nickerson, 1965, 1967; Nickerson & Burnham, 1969; Snodgrass, 1969; Sperling & Dosher, in press). In an earlier analysis of the present experiment, Sperling and Reeves (1980) divided the foreperiods into quartiles and noted that ARTs and MRTs were about 30 ms shorter in the last quartile (Target Positions 18, 19, and 20) than the first (Positions 9, 10, and 11). Foreperiod did not interact with target identity; that is, the two effects on MRT and on ART were additive. Here, we extended the foreperiod analysis to the examination of foreperiod effect on order (P_{iBj}), which, as far as we could determine, was negligible. Specifically, violations of laminarity of P_{iBj} occurred equally in all foreperiod quartiles and were not the result of mixing different foreperiods.

Persistence

That a brief stimulus persists only until stopped by a subsequent item has been argued for masking by several authors (Liss, 1968; Liss & Reeves, 1983; Sperling, 1963). (This is not to imply that subsequent stimuli also stop higher level processes, such as memory comparison, as pointed out, for example, by Hoffman, 1979). Unpublished experiments by one of us (Sperling) show that subsequent superimposed letter characters and letter fragments (visual noise) have similar effects on visual search in a search

paradigm (Sperling, Budiansky, Spivak, & Johnson, 1971). It is therefore reasonable to assume a rectangular pulse for $b_i(t)$. Variations in this assumption, in which items decayed before the next item onset or items persisted through in reduced form to the item after the next one, were tested but did not improve the model fits.

Solution of Equation 11

To solve for V_i', we write Equation 11 as

$$V_i' = S_i'/(1 + \beta \sum_{j=0}^{j=i-1} V_j')$$

and note that $S_0' = V_0'$. By iteration,

$$S_1' = S_1'/(1 + \beta V_0'),$$

and so on, for each successive V_i'. A nonnegative solution, $V_i' \geq 0$, exists only if $\beta > -0.5$. The order component, $H_i' = V_i' - S_i'$, is initially zero for $i = 1$, increases to a maximum, and then decreases at large values of i for which V_i' returns to zero. From the set of V_i', solutions for $a'(t)$ can be obtained, the latter being constrained only at n points for which the S_i' are available; functions $a'(t)$ that have the same integral at the n points would be indistinguishable.