

Measuring and Modeling the Trajectory of Visual Spatial Attention

Shui-I Shih
University of Southampton

George Sperling
University of California, Irvine

In a novel choice attention-gating paradigm, observers monitor a stream of 3×3 letter arrays until a tonal cue directs them to report 1 row. Analyses of the particular arrays from which reported letters are chosen and of the joint probabilities of reporting pairs of letters are used to derive a theory of attention dynamics. An attention window opens 0.15 s following a cue to attend to a location, remains open (minimally) 0.2 s, and admits information simultaneously from all the newly attended locations. The window dynamics are independent of the distance moved. The theory accounts for about 90% of the variance from the over 400 data points obtained from each of the observers in the 3 experiments reported here. With minor elaborations, it applies to all the principal paradigms used to study the dynamics of visual spatial attention.

We explored a method of measuring the trajectory of spatial attention that is analogous to measuring the trajectory of subatomic particles in a Glaser bubble chamber (Gray & Isaacs, 1975). In the bubble chamber, a three-dimensional space is filled with a superheated liquid. A particle traveling through the liquid causes rapid localized boiling—microscopic bubbles—along its path. The bubbles can be photographed to indicate the particle's track and the tracks of decay and reaction products that it might produce. The complete tracks of all the bubbles define the trajectory of the particle (see Figure 1). In our procedure, a two-dimensional 3×3 array is filled with letters. These letters change 7 to 10 times per second. Thus each letter occupies a little cube in a three-dimensional space-time volume. During a movement of voluntary attention, some of the letters along the attention trajectory are entered into memory. The track of the remembered letters defines the trajectory of attention through the three-dimensional volume.

Once attention trajectories have been measured in a three-dimensional array of letters, it would be desirable to know whether these trajectories are the same trajectories as have been inferred from other procedures, such as partial report, which involves transfer from iconic memory (Neisser, 1967) to durable storage (Coltheart, 1980) of letters in a single two-dimensional array. A

strong test of the possibility of equivalent attention trajectories in different experimental paradigms requires that all paradigms be tested with the same observers and with similar stimulus materials. Therefore, in addition to the main experiment, which measured attention shifts in three-dimensional letter arrays, two supplementary experiments were conducted with two-dimensional letter arrays: partial-report and whole-report procedures, each with and without poststimulus masks, all with the same 3×3 letter arrays.

Whereas estimating an attention trajectory from the letter-filled three-dimensional displays is relatively straightforward, estimating attention trajectories from partial-report experiments is complicated, requiring a full model of information decay, information transfer rates from iconic to short-term memory, and observers' strategies (e.g., Gegenfurtner & Sperling, 1993). Here, we built on earlier models by incorporating the newly measured attention trajectories, variability in cue interpretation time, and various additional factors to arrive at a general, computational model of spatial selective attention in early visual processing.

Experiment 1, which measured attentional trajectories, was based on the principle outlined above, and it yielded a highly complex data set for each observer, involving more than a hundred data points, each point defined, typically, by 500–1,000 observations. Experiments 2 and 3, which provided various measures of iconic memory and of perceptual acuity also, yielded hundreds of data points for each observer. A strong proof of a theory is that it makes efficient and quantitatively accurate predictions of hundreds of data points from a variety of different experiments, and this is what is claimed for the theory proposed here. In addition to accounting for the experiments reported herein, the theory is applicable to a full gamut of experiments that have been used to measure the dynamics of spatial attention. These are reviewed first.

Overview

Principal Paradigms for Investigating the Mechanisms of Covert Attention Shifts

Although eye movements and attention shifts typically are coupled, it has been clearly shown that observers can move attention

Shui-I Shih, Department of Psychology, University of Southampton, Southampton, United Kingdom; George Sperling, Departments of Cognitive Sciences, Neurobiology and Behavior, and Institute of Mathematical Behavioral Sciences, University of California, Irvine.

The original experimental work was supported by Office of Naval Research Perceptual Sciences Program Grant N00014-88-K-0569 and by the U.S. Air Force Office of Scientific Research, Life Sciences, Visual Information Processing Program, which also supported the theoretical work and preparation of the article. We thank Claus Bundesen and Richard Shiffrin, who served as both readers and reviewers, for their helpful comments on the article, and we thank Karl Gegenfurtner for his assistance in programming Experiment 1.

Correspondence concerning this article should be addressed to George Sperling, Department of Cognitive Sciences, University of California, Irvine, California 92697. E-mail: sperling@uci.edu



Figure 1. Particle trajectories in a bubble chamber. A particle that enters the chamber at the lower left creates microscopic bubbles as it encounters nuclei along its path. It splits to produce two continuing trajectories. The composite trajectory produced by all the ensuing particles in the presence of surrounding electric and magnetic fields identifies and characterizes the original particle and its descendants.

from one spatial location (or object) to another without any change in the eye position (e.g., Eriksen & Hoffman, 1972, 1973, 1974; Klein, 1980; Posner, Nissen, & Ogden, 1978; Posner, Snyder, & Davidson, 1980; Sperling & Melchner, 1978; Wolford & Morrison, 1980). An attention shift without eye movements is called a *covert* (or endogenous) attention shift. All the *attention shifts* mentioned in the present text refer to covert attention shifts. The characteristics of attention shifts have become the focus of many studies in visual attention as discussed below.

Attempts to characterize attention shifts have used five basic paradigms, all of which involve spatial cues: (a) simple reaction times (RTs; e.g., Hughes & Zimba, 1985, 1987; Posner et al., 1978; Posner et al., 1980; Remington & Pierce, 1984; Rizzolatti, Riggio, Dascola, & Umiltà, 1987; Shulman, Remington, & McLean, 1979; Shulman, Wilson, & Sheehy, 1985), (b) choice RTs (e.g., Colegate, Hoffman, & Eriksen, 1973; Egly & Homa, 1991; Eriksen & Hoffman, 1972, 1973, 1974; Eriksen & Webb, 1989; Hoffman, 1975; Jonides, 1980, 1983; Klein, 1994; Murphy & Eriksen, 1987; Musseler, 1994; Podgory & Shepard, 1983; Posner et al., 1980, Experiments 3 and 4; Shaw, 1978; Tsal, 1983), (c) discrimination (e.g., Cheal & Lyon, 1989; Cheal, Lyon, & Gottlob, 1994; LaBerge & Brown, 1986; Lyon, 1990), (d) partial report (e.g., Averbach & Coriell, 1961; Sperling, 1960; also see Coltheart, 1980, for reviews), and (e) attention gating (e.g., Reeves & Sperling, 1986; Sperling & Reeves, 1980; Sperling & Weichselgartner, 1995; Weichselgartner & Sperling, 1987). Major theories of attention shifts—spotlight (e.g., Posner et al., 1980), zoom lens (e.g., Eriksen & Yeh, 1985), gradient (e.g., LaBerge & Brown, 1989), premotor theory (e.g., Klein, 1980), and attention gating (Reeves & Sperling, 1986)—have recently been reviewed (e.g., Cheal et al., 1994; Egly & Homa, 1991; Eriksen & Murphy, 1987; Rizzolatti, Riggio, & Sheliga, 1995; Shiffrin, 1988; Sperling & Weichselgartner, 1995; Yantis, 1988). Thus, our brief discussions here first focus on the experimental paradigms rather than the theories. We then introduce a new paradigm—choice attention gating—and argue that this paradigm has important advantages in the investigation of attention shifts.

Reaction-Time (RT) Tasks

The influence of attention is inferred in an RT task when the observer responds more quickly to stimuli at a location A when he or she is attending A than when he or she is attending to another location, B. The speed of the attention switch is inferred from the time it takes, after a cue to attend to A while the observer is attending elsewhere, for responses to targets at A to achieve the same advantage as when attention is initially directed to A. RT paradigms typically vary (a) the retinal eccentricity of possible target locations, (b) the foreperiod between a location cue and target appearance, and (c) the validity of the cue. Usually, the simple RT task involves detecting a luminance increment; the choice RT task involves a discrimination (e.g., between two characters, X and O). Relative to a trial with no cue or a neutral cue, RTs to targets are fast (a benefit) if the targets appear at cued locations; RTs are slow (a cost) if targets appear at invalidly cued locations. The magnitude of the costs and benefits increases as the foreperiod increases from 0 to 100 or 300 ms, depending on the cue type (Cheal & Lyon, 1991; Klein, 1994; Wright & Ward, 1994) and the task (see Murphy & Eriksen, 1987).

At least four technical problems arise from the procedure of RT tasks. First, the manipulation of distance is usually confounded with the stimulus eccentricity (see Cheal & Lyon, 1989; Sagi & Julesz, 1985). Thus, it is not clear whether the increase in RT for the “far” cued location is data limited (due to poorer visual processing capabilities at greater eccentricities) or resource limited (Norman & Bobrow, 1975; due to resources required to program and execute long-distance attention shifts). Second, there commonly is no control over visual persistence. This makes it difficult to determine how much time has been used to shift and focus attention—a slow attention shift may still encounter legible visible persistence at the shifted-to location. This requires a poststimulus mask (Sperling, 1963) to eliminate the persistence (Cheal & Lyon, 1989; Lyon, 1990). Third, the manipulation of cue validity causes observers to prepare differentially for different trials but typically only in proportion to the probability of target occurrence at different locations; that is, probability matching occurs (Eriksen & Murphy, 1987; Eriksen & Yeh, 1985; Shaw, 1978). Because cues do not necessarily produce 100% conformity, true costs and benefits may be underestimated. Fourth, the foreperiod between cue presentation and target onset has a complex effect on performance; long foreperiods (e.g., 200–1,000 ms; Shulman et al., 1985) are particularly difficult to evaluate because of changing hazard functions and strategies. With choice RT, for example, the maximum benefit is achieved with a foreperiod of 150–250 ms. The possibilities of replacing facilitatory effects with inhibitory effects (Posner & Cohen, 1984) and of making eye movements during long foreperiods further complicate the interpretation.

Perhaps the most serious problem in using spatially cued RT tasks has been an ambiguity in the theoretical interpretation of the results. Several investigators (e.g., Duncan, 1980; Lappin & Uttal, 1976; Shaw, 1982; Sperling, 1984; Sperling & Doshier, 1986) have argued strenuously that attention can operate concurrently at different levels of processing. Although some studies have suggested that spatial cuing enhances perceptual sensitivity at the cued location (see below), RT advantages for cued locations may not necessarily involve better perceptual processing of stimulus informa-

tion but may instead merely involve a change in response criterion at the decision level, analogous to the difference between a change in sensitivity and a change in bias in signal-detection theory (Green & Swets, 1966). Although a change in decision criterion could be considered to be an attention process, the main focus of most investigators has been attention processes that enhance perceptual processing.

Spatially Cued Discrimination or Identification

The influence of attention is inferred in a discrimination task when the observer responds more accurately to stimuli at a location A when he or she is attending to A than when he or she is attending to another location, B. The speed of the attention switch is inferred from the time it takes, after a cue to attend to A while the observer is attending elsewhere, for responses to targets at A to achieve the same accuracy advantage as when attention is initially directed to A. The primary dependent measures are accuracy and sensitivity (or discriminability). Perhaps because of technical problems similar to those discussed above, early studies using accuracy of form discrimination found relatively small effects, typically less than 5% (e.g., Egly & Homa, 1984; Grindley & Townsend, 1968; Van der Heijden, Schreuder, & Wolters, 1985). These were followed by numerous studies that indicated perceptual processing indeed may be better at cued locations than at uncued locations (e.g., Chastain, 1991; Downing, 1988; Muller, 1994; Muller & Findlay, 1987; Possamai & Bonnel, 1991; Shaw & Shaw, 1977).

Among the most rigorous investigations of the dynamics of attention shift using spatially cued form discrimination are those of Cheal and Lyon (1991, 1994; Lyon, 1990). They resolved several technical problems encountered in RT tasks. Specifically, in their experiments, (a) a poststimulus mask is used to constrain visual persistence, (b) 100% valid cues are used to discourage random distribution of attention, (c) short foreperiods (17–267 ms) or even negative foreperiods are densely sampled to characterize performance as a function of the time available for an attention shift, and (d) stimuli are first scaled for equal discriminability as a function of eccentricity before the effect of distance on attention shifts is evaluated. Cheal and Lyon (1991) found that discrimination accuracy increases as the foreperiod increases from 0 to 100 ms and asymptotes thereafter. More important, accuracy does not vary as a function of eccentricity for most foreperiods when the stimuli are scaled to be equally discriminable at all eccentricities (over the range tested, from 2 to 10 degrees of visual angle [deg]). However, it is noteworthy that poststimulus cuing (i.e., cue onset is after target onset) also facilitates discrimination sensitivity (Lyon, 1990, Figures 1 and 2). This implies that the facilitation is partially due to better perceptual processing, partially due to better processing of the contents of memory, and partially due to other factors.

Cued Partial Reports

In a partial-report task, the observer is cued to report particular items from a large array of items. The to-be-reported items are available only briefly. The influence of attention is inferred when the observer recalls more requested items in response to an attention cue than when no cue is given. A high speed of attention switching is inferred when the observer is able to report a large

fraction of the requested items when there is only a brief interval before the requested items become unavailable. Auditory cues were originally used by Sperling (1960), and visual cues by Averbach and Coriell (1961), and there are useful reviews by Coltheart (1980) and Long (1980) of an extensive subsequent literature. Because partial reports have been used with briefly presented visual displays, the problem for attention theories is that the decay of memory for the display (iconic decay) is confounded with the attention process involved in making the partial report.

Several authors have attempted to isolate the attention switching components from the iconic decay components within a theoretical framework (e.g., Averbach & Coriell, 1961; Bundesen, 1990; Gegenfurtner & Sperling, 1993; Rumelhart, 1970). These theories all propose an initial distribution of processing resources (“attention”) that changes following a partial-report cue; the theories differ in how resources are defined and in other details (see Gegenfurtner & Sperling, 1993, for a review). The overriding problem is that the number of assumptions in all these theories is considerable, and it would be desirable to obtain more direct measurements of the attention processes involved.

Attention Gating

The attention-gating paradigm is a variant of a partial-report procedure in which the presentation of many items in a single brief presentation is replaced with a stream of items (rapid serial visual presentation; RSVP) in one or a few locations. The cue directs the observer to report only items at a particular time and place. The time course of attention shifts measured by the gating paradigm is consistent with the time course measured by Cheal and Lyon (1991) using spatially cued discrimination. Although the attention-gating paradigm was developed for the task of defining the dynamics of attention shifts (Reeves & Sperling, 1986; Sperling & Reeves, 1980; Sperling & Weichselgartner, 1995; Weichselgartner & Sperling, 1987), it so far has not been applied to measuring choice RTs of attention shifts to different locations. We describe the paradigm and its application below.

A New Paradigm: Choice Attention Gating

RSVP Streams

Attention-gating paradigms use rapid serial presentations (streams) of arrays to fill a space-time cube with items, any of which might be requested for report. In Reeves and Sperling’s (1986) study, for example, the observer maintained fixation between two adjacent streams of characters: a target-containing letter stream and a to-be-reported numeral stream in which numerals occurred at rates of from 4 to more than 13 numerals per second. The observer’s task was to detect the target and then to attempt to report the simultaneous item (and the next three items) from the to-be-reported stream. This involved a change of task (from searching for a letter target to remembering numerals), and it also involved a spatial shift of attention from the search location to the to-be-reported stream.

The data from an attention-gating experiment are the items reported by the observer. When the observer has to shift attention spatially from the target location to the location of the to-be-

reported stream, the most frequently reported item occurs about 300 to 400 ms after the target onset (Reeves & Sperling, 1986). Indeed, the temporal distribution of reported items (in the gating procedure) and the distribution of simple motor RTs (to the same targets) are strikingly similar, suggesting that many of the processes involved in shifts of visual attention are shared with those that underlie motor responses (Sperling & Reeves, 1980). Usually attention shifts are studied in choice paradigms in which the direction of the attempted attention shift depends on the cue to shift attention. The attention-gating paradigm has hitherto been investigated only with "simple" attention shifts (the direction of the shift is independent of the cue) and has not been attempted with choice attention shifts. The attention-gating procedure is here elaborated to a choice attention-gating procedure in which the direction of the attention shift depends on the particular tonal cue received.

Choice Attention Gating With Three Tonal Cues

In the current application, a stream of 3×3 letter arrays was presented to the observers in rapid succession. Each observer was instructed to fixate at the center of the display throughout the entire stimulus presentation. The row to be reported was indicated to the observer by a tonal cue, whose onset coincided with the onset of a particular array in the stream and whose pitch indicated the row. The report in correct left-to-right sequence of items from the cued row and items from arrays presented not too long after the array simultaneous with the cue was differentially rewarded according to a payoff matrix. Reports of items from the wrong row or wrong column or from late-occurring arrays were penalized. Observers could earn the largest payoffs by shifting attention as quickly as possible within the cue-determined spatiotemporal window. Throughout the many thousands of trials, observers attempted to optimize their performance.

Outline

Experiment 1 used the choice attention-gating paradigm to investigate the time course of attention shifts in response to a tonal cue. Two different rates of array presentation were used in the streams to ensure that the estimated parameters of the attention shifts were independent of the particular circumstances of measurement.

Experiment 2 investigated iconic memory following brief exposures of the same kind of arrays as were used in Experiment 1 and with the same observers. This offered an opportunity to determine whether the parameters of the attention shifts that were derived from Experiment 1 could apply to attention shifts in iconic memory experiments. To further define the properties of attention-directed retrieval of information from brief displays, we used postexposure pattern masks (Sperling, 1963) in some conditions of Experiment 2 to constrain the period during which retrieval of information was possible.

Experiment 3 examined whether the imperfect performance observed in partial reports was due to a limitation in memory or to perceptual factors. Experiment 3 used a whole-report procedure and independently varied the to-be-reported row, mask delay, presence-absence of null symbols in the nontarget rows, and

presence-absence of an early tonal cue indicating the row to be reported. The purpose of eliminating items from the nontarget rows in some trials was to estimate the best performance the observer could achieve for each row and, thereby, to provide estimates for perceptual acuity at the nine retinal locations at which characters occurred in our arrays.

On the basis of the experimental results, we propose a *strength model of visual selective attention* to account for the results of the three experiments in a common framework of plausible mental processes as well as to relate them to other attention paradigms. The attention model consists of four components: perceptual processing, an attention control mechanism that determines which of the perceptually processed items are stored in visual short-term memory (VSTM), and a decision mechanism that determines which remembered items are reported.

Experiment 1: RSVP Choice Attention-Gating Experiment

Experiment 1 used a choice attention-gating paradigm to investigate the time course of attention shifts in response to a tonal cue.¹ The paradigm combines a partial report with a rapid serial visual presentation (RSVP) procedure. The observer viewed a stream of 3×3 letter arrays presented successively. The observer was instructed to fixate at the center of the display throughout the stimulus presentation. A tonal cue was presented at the onset of one of the letter arrays. The cue told the observer to immediately sample the letters from the row indicated by its pitch. A payoff matrix was used to differentially reward and penalize responses according to their correctness and temporal locations relative to the cue onset.

The stream of arrays was presented so rapidly, with a duration of 50 ms for each array and a stimulus onset asynchrony (SOA, the time between successive stimulus onsets) of 100 or 150 ms, that the decay in iconic memory was relatively unimportant and saccadic eye movements were discouraged (Sperling & Reeves, 1980). Because of the limited capacity of short-term memory (Miller, 1956; Sperling, 1963), the choice attention-gating paradigm coupled with a proper payoff matrix discourages the observer from nonselective transfer of letters from iconic memory into a durable storage (Averbach & Coriell, 1961; Gegenfurtner & Sperling, 1993), that is, VSTM (Phillips, 1974; Scarborough, 1972).

The observer performed two tasks in succession on each trial. First, he or she performed a choice reaction-time (RT) task to indicate which of the three tonal cues had been presented. Second, he or she typed the three letters from the indicated row in the earliest possible array simultaneous with or subsequent to the onset of the tonal cue. In this task, the ordered recall task, the observer was required to report a letter from each column position in the row. The order of report was from left to right.

Method

Observers

Two New York University graduate students, J.S. and S.S., volunteered their services. S.S. was the experimenter. Both observers had normal or

¹ Preliminary reports of this experiment appeared in Shih and Sperling (1996) and Sperling and Shih (1997).

corrected-to-normal vision. Stimuli were viewed binocularly at a distance of approximately 132 cm in a dimly lit room. A chin rest was used to fix the viewing distance. Each observer performed at least 1,000 practice trials before the experimental sessions.

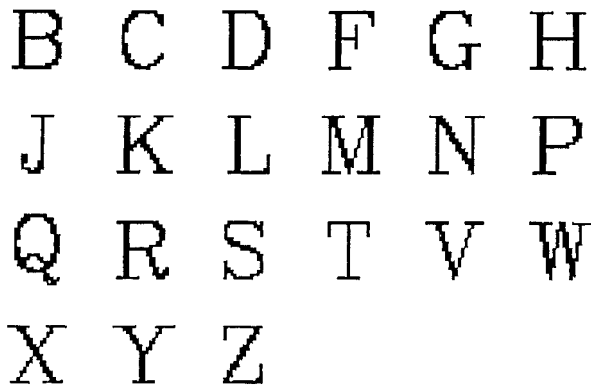
Apparatus

The experiment was programmed to run under the MS-DOS operating system using a Maxum AT (PC clone) as the host computer. An Enhanced Graphics Adapter (EGA) and the Runtime Library (Hall & Gegenfurtner, 1988) were used to control the presentation of visual displays. The visual stimuli were displayed on an EGA monitor with a refresh rate of 60 Hz. A Data Translation audio digital/analog signal generator was used for the presentation of tonal cues through earphones.

Stimuli

Visual stimuli. The stimulus set consisted of 21 uppercase consonant letters (*A*, *E*, *I*, *O*, and *U* were not used; see Figure 2a). The height and width of the letter *M* were 0.56 and 0.48 degrees of visual angle (deg), respectively. The luminance of the letter strokes was about 74 cd/m². *A*

a.



b.



Figure 2. Stimuli used in the experiments. (a) Stimulus letters. The height and width of the letter *M* are 0.56 and 0.48 degrees of visual angle, respectively. A 3×3 array of randomly chosen letters, sampled with replacement, was the arrangement used. (b) Random pattern masks used in Experiments 2 and 3. A mask presentation consisted of a rapid sequence of different pattern masks.

letter array was composed of nine letters randomly sampled with replacement and arranged in a 3×3 array. The size of a letter array was approximately 3.52×3.52 deg². The center-to-center distance between adjacent letters was about 1.52 deg. The fixation point was a central square subtending 0.13×0.13 deg².

Auditory stimuli. High- (1100 Hz), medium- (650 Hz), and low-pitched (250 Hz) tones, respectively, were used to indicate the top, middle, and bottom rows. To eliminate clicks, we presented tones with 20 ms rise and decay times. The tonal cue was presented through headphones for 200 ms.

Design

Two variables were manipulated: the cued row and the SOA (stimulus onset asynchrony). The cued row (top, middle, or bottom), indicated by a tonal cue, was the one from which the observer was instructed to report three letters. The SOA (100 or 150 ms) was the interval between onsets of two consecutive letter arrays. The cued row was chosen randomly with equal probability in each trial, whereas the two SOAs were run in separate blocks. The dependent measures were speed and accuracy in response to a tonal cue and recall probability as a function of a stimulus letter's spatial location and temporal location relative to the cue onset (see below).

Procedure

Figure 3 illustrates the procedure of the choice attention-gating experiment. The observer pressed a key to initiate a trial whenever he or she was ready. A central fixation square was presented for 0.5 s. After 0.5 s of a blank screen, a stream of 15–21 letter arrays was presented. Each array was presented for 50 ms (i.e., three refreshes). Arrays were separated by a 50-ms blank screen (the 100-ms-SOA condition) or a 100-ms blank screen (the 150-ms-SOA condition). A tonal cue was presented at the onset of one of the arrays. The observer was asked to perform, first, a choice RT task and, then, an ordered recall task. Feedback was given after the ordered recall responses.

Temporal location of a tonal cue and critical arrays. The number, *n*, of arrays before the onset of a tonal cue was randomly varied between three and nine from trial to trial. The probability of each *n* was determined by an exponential function, that is, $p(n) = (1/3) \times (2/3)^{n-3}$, $n \in [3, 8]$. Given that a tone had not yet occurred at array *n*, $3 \leq n \leq 8$, the probability that the tone would occur on array *n* + 1 remained constant (at 1/3) with this aging distribution. If the tone had not occurred in any of the arrays [3, 8], it occurred in the ninth array; the probability of this outcome was $p(9) = .0585$. The six consecutive arrays, starting from the one immediately preceding the cue onset, were defined as the critical arrays. After the critical arrays, six more arrays were presented.

Choice RT. Before each trial, the observer placed his or her right index, middle, and ring fingers, respectively, on keys 1, 2, and 3 of the number pad on the right-hand side of the keyboard. This task required the observer to respond, as quickly as possible, to the pitch of a tonal cue by pressing one of the three keys. Keys 1, 2, and 3 corresponded to the low-, medium-, and high-pitched tones, respectively. RTs and correctness were recorded. Feedback was given for "too slow" and "too fast" choice RTs (see below).

Ordered recall task. Ordered recall is the primary task of the present experiment. The observer was required to type the three letters which, to the best of his or her ability, were (a) from the cued row, (b) from the earliest possible array simultaneous with or subsequent to the onset of the tonal cue, and (c) in the correct column position. A payoff matrix (see Table 1) was implemented to encourage observers to (a) not guess randomly (responses outside the critical arrays were penalized), (b) report letters from the earliest possible array after the cue onset, (c) use high

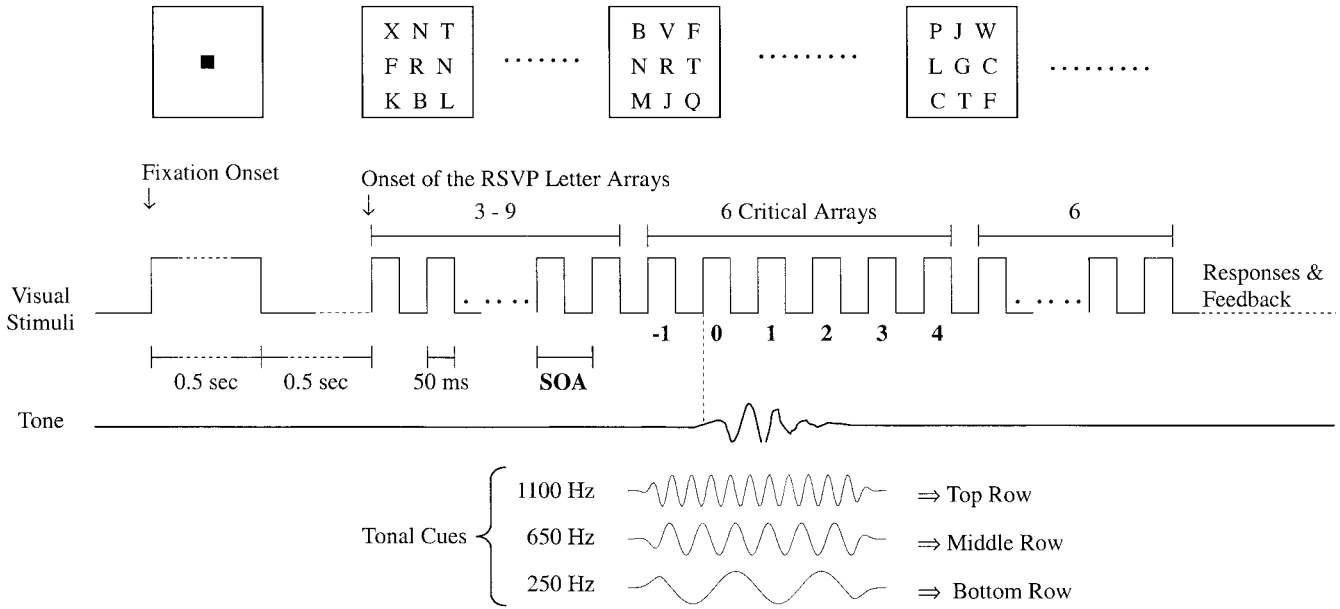


Figure 3. Procedure for Experiment 1: choice attention gating. The observer presses a key to initiate a trial. A central fixation square is presented for 0.5 s followed by 0.5 s of a blank screen. Then, a stream of 15–21 letter arrays is presented. Each array is presented for 50 ms (i.e., 3 refreshes at 60 refreshes per second). Arrays are separated by a 50-ms blank screen, producing a stimulus onset asynchrony (SOA) of 100 ms. (In the 150-ms-SOA condition, the blank screen is presented for 100 ms.) A 200-ms (high-, medium-, or low-pitched) tonal cue is presented through headphones to indicate the (top, middle, or bottom) cued row. The 6 critical arrays are -1 to +4 relative to the onset of the tonal cue (0). There are 3–9 arrays before the critical arrays and 6 arrays afterward. RSVP = rapid serial visual presentation. Feedback immediately after the responses includes a 3 × 6 array of the three letters displayed in the cued row of the 6 critical arrays plus the credits earned for quick and accurate responses.

confidence for reports believed to contain correct item-and-position reports, and (d) use low confidence otherwise. The observer indicated one of two levels of confidence for each letter reported. Uppercase responses indicated high confidence, and lowercase responses, low confidence.

Table 1
Experiment 1: Payoff Matrix—Points Earned for Each Reported Letter as a Function of Response Confidence, Temporal Position, and Spatial Location of the Matching Stimulus Letters

Response confidence	Temporal position relative to the cue onset						
	<0	0	1	2	3	4	>4
Spatial location: Match							
High	-10	10	7	4	2	1	-10
Low	-10	3	2	1	1	1	-10
Spatial location: Match adjacent							
High	-10	0	0	0	0	0	-10
Low	-10	0	0	0	0	0	-10
Spatial location: No match							
High	-10	-10	-10	-10	-10	-10	-10
Low	-10	-2	-2	-2	-2	-2	-10

Note. In cases in which a response could be scored in different ways, the highest possible credit was assigned.

Typed responses were displayed on the screen and the observer was allowed to correct them before pressing the enter key. Performance was summarized in a set of recall probabilities $p(r, c, t)$. Each probability $p(r, c, t)$ was the proportion of matches between the response $R(c)$ for column c and the stimulus letter presented at row r , column c and time t relative to the cue onset. Because the stimulus letters were randomly sampled with replacement, it was likely that a response $R(c)$ would match more than one stimulus letter in the critical arrays. However, these multiple matches only reflect chance-level guessing, which is easily estimated.

Feedback. Feedback included (a) a 3 × 6 array of the three letters displayed in the cued row of the 6 critical arrays; (b) the three letter responses; (c) credits obtained, according to the payoff matrix, in the present trial; and (d) the cumulative credits. A message of “Late!” (or “Early!”) was displayed when the choice RT was longer than 1,000 ms or shorter than 150 ms. No ordered recall was required when this occurred. Although there were no monetary rewards, observers discussed their scores and attempted to improve them. Separate feedback was given for each reported letter. The instructions and feedback clearly indicated that each reported letter should come from the earliest array possible; the reward structure was indifferent to whether reported letters occurred in the same versus different arrays.

Sessions. Observer S.S. completed 40 experimental sessions of 100 trials each, except for Session 1, which consisted of 76 trials (for a total of 3,976 trials), with an SOA of 100 ms, and thereafter 34 sessions of 100 trials each (3,400 trials) with an SOA of 150 ms. Observer J.S. completed 41 experimental sessions of 50 trials each (2,050 trials) with an SOA of 150 ms, and thereafter 34 sessions of 75 trials each (2,550 trials) with an SOA of 100 ms.

Results and Discussion

Overview

All analyses were performed individually for each observer. We first focus on the ordered recall task and then describe the results from the choice RT task. For the ordered recall task, we first show that observers did indeed follow the instructions to shift attention to the row indicated by the tonal pitch. Hence, the analyses that follow are contingent on each cued row. Because the data were collected over a protracted period of time, we tested whether the data are reasonably homogeneous for each SOA condition and for each observer. The data of observer S.S. in the 100-ms-SOA condition were partitioned into two clusters to reflect a practice effect that was indicated by cluster analysis and verified by con-

tingency analysis (Shih & Sperling, 1994). Whether attention shifts simultaneously to all the letters in a row was evaluated by contingency analyses. Also considered were the absence of the expected decline of performance with eccentricity and the absence of interference between the choice RT task and the ordered recall task.

Ordered Recall Task

The proportion of high-confidence responses was larger in the second SOA condition (i.e., 150-ms SOA for S.S., 99%; 100-ms SOA for J.S., 98%) than it was in the first SOA condition (i.e., 100-ms SOA for S.S., 90%; 150-ms SOA for J.S., 96%). Because the proportion of low-confidence responses was small, we did not

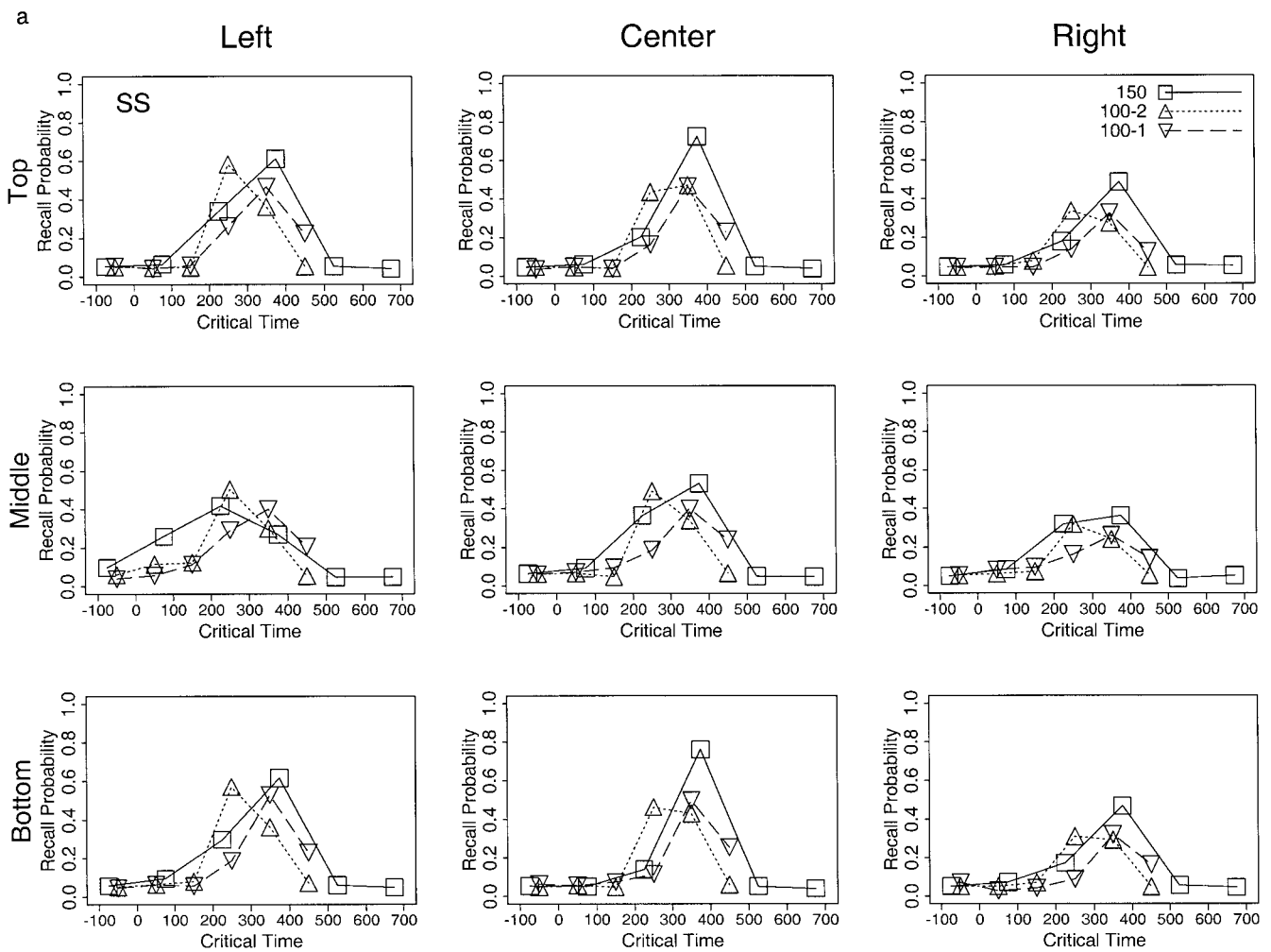


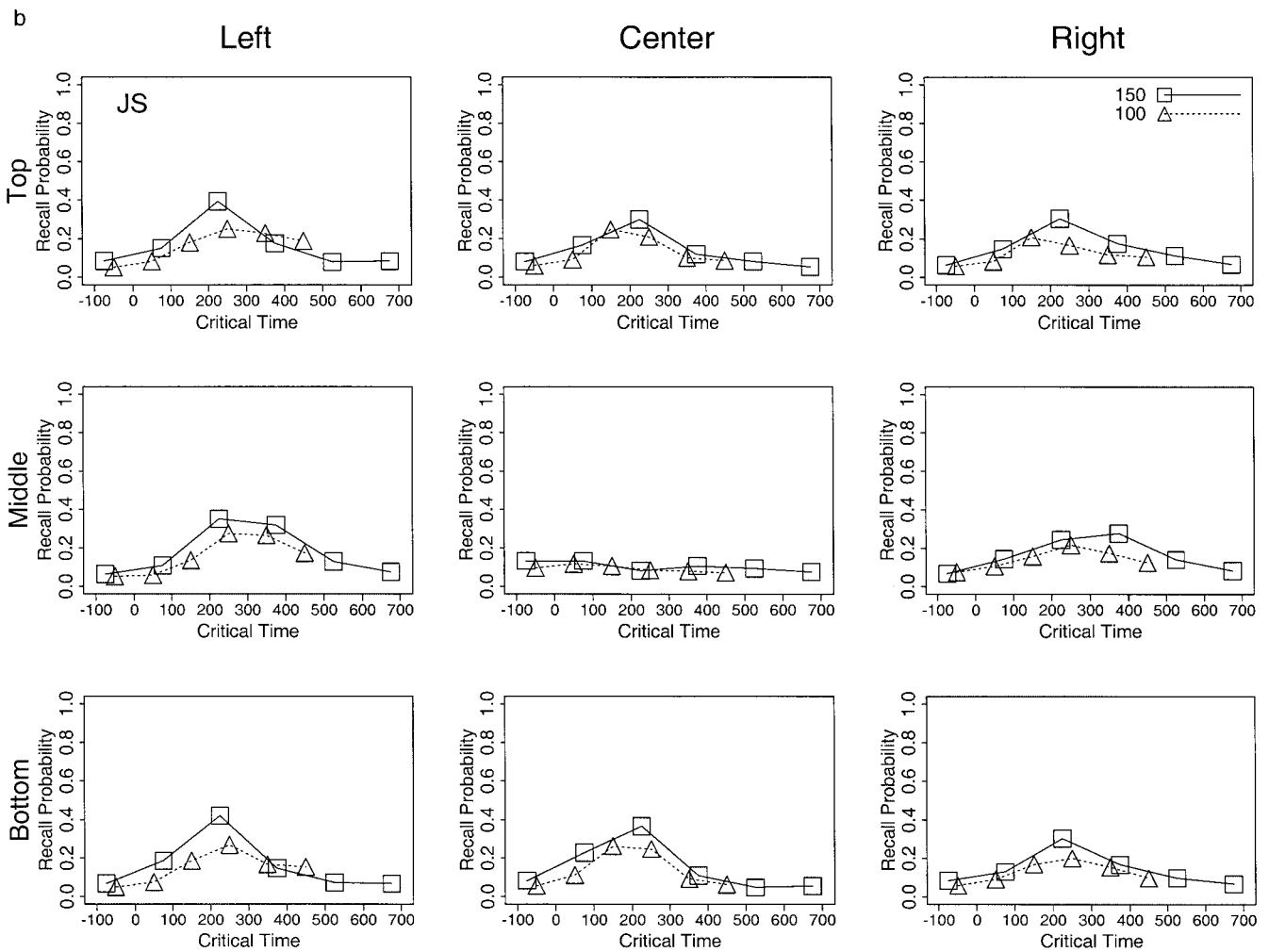
Figure 4. Data from Experiment 1: choice attention gating. Recall probability as a function of the critical time (in milliseconds). Critical time is time from the onset of the tonal cue to the midpoint of the interval during which the cued row of letters is available, that is, the time from tonal cue onset to letter onset plus half of the time between consecutive array onsets (stimulus onset asynchrony; SOA). The location of each panel corresponds to the 3 cued rows \times 3 column positions. Curve parameters are indicated at the top right panel. (a) Observer S.S., (b) observer J.S. 100 and 150 represent the SOA between successive arrays; 100-1 and 100-2 (observer S.S.) represent clusters of data from less-practiced and more-practiced sessions, respectively.

distinguish the low- and high-confidence responses and aggregated them in the following analyses.

Attention shifts are confined to the cued row. Observers were instructed to direct attention to the row indicated by the tonal pitch on each trial. The three letters in their response were to correspond to the three columns of the cued row. How accurately did the letters the observers reported correspond to the intended row and column positions? That is, to what extent did observers make row and/or column confusions? To answer this question, on each trial we matched each response letter (corresponding to a particular cued row R and response column C) to every stimulus letter (row r , column c) in the set of six critical arrays, indexed by t , to yield $p_{R,C}(r, c, t)$, the probability of a match. These data are analyzed in detail in Appendix A. The important result is that row confusions

were statistically and quantitatively nil. Essentially, observers reported letters only from the cued rows. However, there were statistically significant column confusions between adjacent letters. These were most likely to occur in reports of the middle row and for observer J.S. The main column confusions were J.S.'s reporting what he thought was the middle position of the middle row (which he was ignoring) but actually reporting items that occurred in the (adjacent) left and right positions.

Examination of data homogeneity: Clusters. Although observers had practiced for more than 1,000 trials before the experimental sessions, the data collected across sessions might not be comparable because of practice effects that resulted from the complicated nature of the tasks and the long period of data collection. Therefore, before data from the ordered recall were collapsed over



sessions, the homogeneity (over sessions) of the data was evaluated (Shih & Sperling, 1994). A clustering algorithm (FASTCLUS procedure, SAS Institute, 1985) was applied to the sessions within each SOA condition. For a given SOA condition, the necessity of partitioning sessions was first judged by visual inspection of the similarity of the profile of recall probabilities and then verified by contingency analysis (see the *Location × Location Response Contingencies* section below). For J.S., the data collected across sessions were judged to be homogeneous for each SOA condition; hence each consisted of a single cluster. For S.S., sessions in the 100-ms-SOA condition were partitioned into two clusters: Cluster 1 consists of the first 9 sessions, and Cluster 2, the remaining 31 sessions. Because of the chronological succession and the correlation between cluster and performance level, the difference between the two clusters in the 100-ms-SOA condition obviously represents a practice effect. For example, although motor RT was not used to define the clusters, observer S.S.'s mean RT decreased by 81 ms from Cluster 1 to Cluster 2. Therefore, the data within each cluster are treated separately.

Temporal distribution of recall probabilities. To illustrate the dynamics of attention shifts, we present the full set of recall probabilities for each observer graphically in Figure 4. Each panel describes responses at a particular spatial location (cued row *r* and column *c*). The *x*-axis represents critical time, which is defined as the midpoint of the interval during which an array is visually available, that is, the midpoint between the onset of an array and the onset of the next array, equivalently, the time of stimulus onset plus half of the SOA. Critical time reflects that visual persistence (iconic memory) maintains information very well during the relatively short SOAs (i.e., 100 and 150 ms). The *y*-axis is the proportion of matches of the reported letter to the letter that occurred in the same row and column in each of the arrays. The different curves within a panel represent SOA conditions and, for observer S.S., different levels of practice. The chance level is 1/21. Each curve is referred to as a *recall function*, which reflects the time course of attention at a given spatial location for a condition (or a condition and practice cluster) and observer.

The recall functions are unimodal. For all locations, including the fixated one, most recalled letters had been displayed between 100 and 400 ms after the cue onset. For a given cluster, recall functions for positions in the same row were all similar in shape,

except the one for the central location for J.S., where performance was uniformly low. Early in practice, J.S. made a strategic decision to concentrate his attention on the outer positions and to ignore the center position. Afraid he might have to repeat thousands of trials if he were to change strategy, he maintained this unusual strategy throughout (the first experiment only). In the data analyses, this position is treated separately.

Absence of a perceptual acuity effect. It is noteworthy that performance for the top and bottom rows was not inferior to, even better than (in terms of total recall), recall for the middle row, where the observer fixated. Because of physiological constraints, legibility of (perceptual acuity for) a stimulus decreases as its distance from the fovea increases. Because the center-to-center distance between adjacent letters was about 1.52 deg, performance of the more peripherally placed letters was expected to suffer a perceptual acuity loss. However, there was no impairment of performance with increased eccentricity. We return to this paradox later.

Attention reaction time (ART). The time from an attention cue to the time of occurrence of the first reported item in a stream can be considered an ART by analogy to a motor RT (MRT; Sperling & Reeves, 1980). In Experiment 1, there were three attention cues (top, middle, bottom) and three streams within each row (left, center, right). Only one item per stream within a cued row was reported. As Figure 4 shows, items from all locations tend to be reported from approximately the same frames. Therefore, we consider here the mean time of occurrence of the three items reported by the observer. There are two caveats: (a) Only trials in which at least one reported item occurred within the critical arrays are considered, and (b) rather than use the onset time of an item, we consider the midpoint of the interval during which an item was available as the time stamp for the reported item. In the following description of the theory, it becomes evident that the ART, so defined, coincides, approximately, with the mean (midpoint) of the attention window. The overall ARTs, averaged over all trials for the 2 observers, were 288 and 252 ms for observers S.S. and J.S., respectively.

Choice RT Task

MRT distributions: Mean. Table 2 summarizes the number of trials, accuracy, means, and standard deviations of (motor) RT for

Table 2
Means and Standard Deviations of Reaction Times (in Milliseconds) of Correct Motor Responses to the Tones Indicating the Target Row (Top, Middle, Bottom), Number of Trials (N), and Percentage of Correct Motor Responses

Observer, SOA ^a (cluster)	Top row				Middle row				Bottom row				Overall			
	M	SD	N	%	M	SD	N	%	M	SD	N	%	M	SD	N	%
S.S., 100 ms (1)	484	56	304	96	471	56	285	94	463	63	287	98	473	59	876	96
S.S., 100 ms (2)	399	52	1,022	97	383	60	1,037	95	394	52	1,041	96	392	55	3,100	96
S.S., 150 ms	365	45	1,146	95	327	47	1,128	95	349	51	1,126	93	347	50	3,400	94
J.S., 150 ms	454	95	680	96	513	114	659	96	458	79	695	98	474	100	2,034	97
J.S., 100 ms	336	56	844	94	408	102	859	88	356	69	846	93	366	84	2,549	92

Note. SOA = stimulus onset asynchrony.

^a The chronological order of the SOA conditions was different between the two observers: S.S. performed 100-ms-SOA conditions first; J.S. performed 150-ms-SOA conditions first.

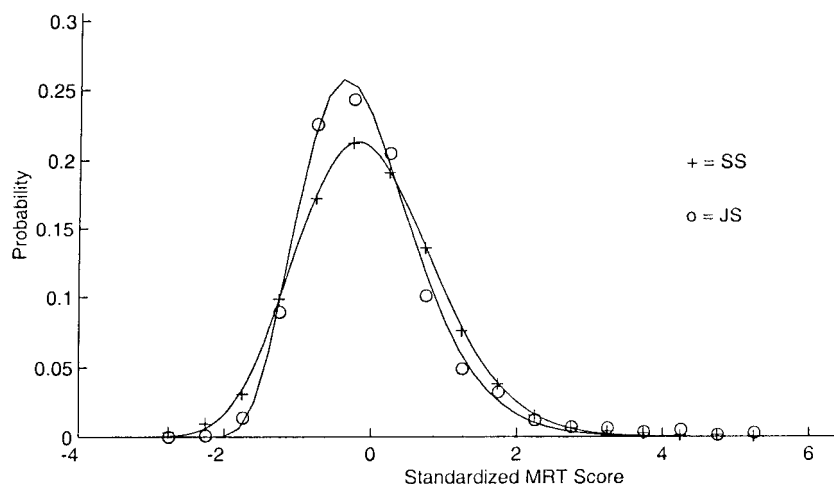


Figure 5. Average distribution of normalized choice reaction-time (RT) responses for each observer, Experiment 1. The distribution of RTs in each condition was normalized to have a mean of 0 and a variance of 1 before combining it in the average. The curves drawn through the points are best fitting Gamma functions of orders 27 (observer S.S.) and 8 (observer J.S.). MRT = motor reaction time.

each tonal cue, SOA, cluster of data (see below), and observer. Accuracy of the choice RT task was generally high (around 95%) except that of J.S. in the 100-ms-SOA condition, in which his response accuracy to the medium-pitched tone was about 88%. There was no indication of a speed-accuracy trade-off. There are two major observations. First, the speed of response to each of the three tonal cues was approximately the same within an observer and condition. The only exception was that observer J.S., who ignored the middle letter of the middle row, responded more slowly to the middle cue. Second, these data indicate that, even though observers had experienced thousands of trials, they continued to make large and highly significant improvements in RTs. Although mean choice RTs were under 500 ms in the first set of data, they were reduced by about 100 ms during the ensuing thousands of trials. It is obvious that a detailed model for these data would have to take into account the continuing improvement, throughout the experiments, of the observers' speeds in interpreting and responding to the attention cues.

MRT distributions: Shape. The RT distribution for the tonal attention cues is important in relation to the model. Inspection of the individual RT distributions suggested that changes in mean RT with practice, and to a lesser extent with row, were the most important differences between distributions; other factors were quite similar over conditions. To characterize the shape of the RT distributions, we converted the distributions within each condition to z scores and computed a grand average for each observer in which all RTs were included. Specifically, in each condition, the mean was subtracted from the RT distribution, the resultant distribution was divided by its estimated standard deviation, and a weighted average of these distributions was computed for each observer in which each individual RT had equal weight. The average RTs for the 2 observers are shown in Figure 5.

The curves drawn through the RT distributions represent Gamma functions of order 8 for observer J.S. and 27 for observer S.S. These Gamma functions account for more than 99.5% of the

variance of the observers' frequency distributions² and are incorporated in the model.

Common components in ART, MRT. The ARTs (defined above) and the MRTs have common initial stages: the perception and interpretation of the attentional cue. Therefore, we expected, and found, a trial-to-trial correlation between ARTs and MRTs. The ART-MRT correlations for the 2 observers, averaged over all conditions, are .25 (S.S.) and .19 (J.S.). These correlations are not very large, but that is misleading. We see later, from the model, that randomness within the attention window itself would greatly reduce the ART-MRT correlation. Indeed, the model shows that average covariance between ART and MRT is 0.51 of the covariance that would be expected if all the randomness in ART and MRT were shared except that produced within the attentional window. It follows that a very large component of the variability in the MRT must be due to variability in cue interpretation time as subsequent stages in MRT are not shared with ART. Conversely, the MRT distributions themselves can be (and are) used to characterize distributions of cue interpretation times.

Does the choice RT task interfere with attention shifts? Whether and how the choice RT response to an auditory stimulus in the first task affects the shift of visual attention to a target embedded in a briefly flashed multielement array in the second task were investigated by Pashler (1991). He found that accuracy in the second task did not vary with the SOA between tone and array nor with the RT in the first task. In other words, an RT response to a tone could occur independently of shifting visual attention (see also Sperling & Reeves, 1980). With practice, our observers improved their RT in the choice RT task by more than

² Whereas the data for both observers are constrained to have a variance of 1, the best fitting Gamma function for observer J.S. has a variance of only 0.82; it "ignores" a small number of outlier points that contribute significantly to the variance of J.S.'s data.

100 ms (see Table 2) whereas their performance in the ordered recall task remained stable. This result suggests a high degree of independence between these two tasks and probably very little, if any, interference between them. Furthermore, similar results were obtained in follow-up experiments using the paradigm of Experiment 1 but omitting the choice RT task (Shih, 1996). Therefore, we conclude that, in the present experiment, attention shifts are not significantly affected by the concurrent choice RT task.

Location × Location Response Contingencies: Do Response Items Tend to Come From the Same Frame?

The main results of Experiment 1 were as follows: For every spatial location, (a) the maximal recall occurred at about the same time, and (b) the widths of the temporal recall functions were about the same. However, all the data considered so far were concerned only with averages of many trials. There are some questions that require the analysis of responses within an individual trial before averaging. In particular, do letters from the same stimulus frame occur together in the stimulus report more often than might be expected by chance (i.e., from the width of the attention gate)? The simplest approach to this analysis is to determine what two-way contingencies exist in the temporal locations of the reported stimulus letters. For example, does report of the upper-left letter from frame *i* make it more likely that the upper-center or upper-right letters will also be reported from frame *i*? The answers to these questions further define the microprocesses of spatial attention.

A two-way contingency matrix. Two-item contingencies between responses were derived from two-way contingency tables, an example of which is illustrated in the top panel of Table 3. The top panel of Table 3 is derived from observer S.S., data for Experiment 1, SOA = 150 ms, and it includes all reports of the top row. The columns show the temporal positions of the reported letter for the left column, and the rows show the temporal position of the reported letters for the right column.

If the reports for the left and right column positions were independent, the frequencies in the top panel of Table 3 would be determined by the product of the marginal probabilities. Five cells in the top panel of Table 3 have significant statistical deviations from these expected frequencies: Three cells on the diagonal have greater frequencies than expected, and two cells adjacent to the diagonal have lower frequencies than expected. These data indicate that observer S.S. reported the left and right letters from the same frame (namely, Frames 1 and 2) significantly more often than would be expected by chance. Despite the high statistical significance, the actual correlation is quite small. The Pearson product-moment correlation for the top panel of Table 3 is .093; the correlation accounts for less than 1% of the variance of the data (i.e., $r^2 < .01$).

A restricted subset of data. The top panel of Table 3 shows all the responses. Responses for temporal positions larger than 4 were grouped together because these responses were not distinguished during data acquisition. Including the grouped position (greater than 4) data has the advantage of considering all the data. It has

Table 3
Observed Frequency as a Function of the Temporal Position of the Reported Letter for the Left Column and the Temporal Position of the Reported Letter for the Right Column, Cued Row Top, 150-ms Stimulus Onset Asynchrony Condition, Observer S.S.

Right column: Temporal position relative to the cue onset	Left column: Temporal position relative to the cue onset							Total
	-1	0	1	2	3	4	>4	
<i>7 × 7 contingency matrix (r = .0926)</i>								
-1	6	2	19	37	1	2	2	69
0	7	9*+	33	30*-	4	3	1	87
1	7	16	115***+	77***-	13	7	11	246
2	31	32	148***-	409***+	30	23	20	693
3	3	6	21	38	7	3	3	81
4	2	4	23	40	4	0	2	75
>4	16	19	102	183	12	16	31***+	379
Total	72	88	461	814	71	54	70	1,630
<i>6 × 6 contingency matrix (r = .0870)</i>								
-1	6	2	19	37	1	2		67
0	7	9	33	30*-	4	3		86
1	7	16	115***+	77***-	13	7		235
2	31	32	148***-	409***+	30	23		673
3	3	6	21	38	7	3		78
4	2	4	23	40	4	0		73
Total	56	69	359	631	59	38		1,212

Note. A plus or minus following a number indicates that the observed frequency is greater (plus) or smaller (minus) than chance-expected frequency at the designated significance level as indicated by a chi-square test of independence. A significance level of .001 keeps the overall alpha at about .05.
* $p < .05$. ** $p < .01$. *** $p < .001$.

two disadvantages: It unfairly groups on the diagonal some letters that might have fallen off the diagonal if they were not grouped together in the single greater-than-4 cell, and it imprecisely locates the positions of all the greater-than-4 letters at Position 5. Therefore, the bottom panel of Table 3 presents the analysis of a restricted subset of data in the top panel of Table 3 in which both responses occurred within the critical arrays (Temporal Positions $-1, \dots, +4$). In the bottom panel of Table 3, the correlation falls (from .093 in the top panel of Table 3) to .087. Subsequently, we restricted our analysis to this subset of the data. We did so primarily because of subsequent awkwardness in dealing with the imprecision of the aggregated positions (greater than 4). As a result of considering only the subset of data in which both responses fell within the critical arrays ($-1, \dots, 4$), the amount of available data was reduced somewhat, the correlation coefficients changed slightly and unsystematically, and all inferences and conclusions remained the same as when considering all the data.

Forty-one contingency matrices. The contingency matrix in the bottom panel of Table 3 illustrates the contingencies between reports of the left and right spatial locations (columns). We must also consider the contingencies between reports of the letters from the center and the right locations and between the left and the center locations. Thus, for each reported row, there were three contingency matrices like Table 3. There were three reported rows (top, middle, bottom), 2 observers, and two SOAs. For observer S.S., there were two clusters for the 100-ms-SOA condition. For J.S., the center location of the middle row was excluded from the contingency analysis because of a universally low performance

level for that location. Altogether there were 41 contingency matrices; 27 for observer S.S. and 14 for observer J.S. Table 4 exhibits the 41 Pearson correlations for these 41 matrices. For observer S.S., the average observed correlation was .082; for observer J.S., it was .063.

Overall correlations within the contingency tables were quite low. For example, .087, which was the correlation for the restricted subset of the data in the bottom panel of Table 3, was similar to the average correlation (.0823) for observer S.S.; the average correlation in the contingency matrices for observer J.S. was .063. The range of correlations also was quite small. Therefore, it is no surprise that the hypothesis that, for each observer, the correlations in all the contingency matrices were the same was rejected for only 2 of the 41 correlations. Nevertheless, there was a small trend for the left and center locations to have a higher correlation than the right and center or the left and right locations. The interpretation of the overall similar, low correlations in terms of the interplay of an invariant attention window and a variable cue interpretation time are considered fully in the Model of Spatial Shifts of Visual Attention section.

Summary

Experiment 1 investigated the time course of attention shifts in response to a tonal cue using a choice attention-gating paradigm. The bell-shaped recall functions and contingency analyses indicated that observers most often reported letters from a cluster of positions centered at about 200–300 ms after the onset of an

Table 4
Summary of the 6 × 6 Contingencies for Experiment 1: Pearson Product–Moment Correlations as a Function of Contingent Spatial Locations, Cued Row, Cluster, and Observer

Observer, SOA (cluster)	Row	Left–center	Center–right	Left–right	<i>M</i>
S.S., 100 ms (1)	Top	.0827	.2074*	.0384	.1095
	Middle	.1350	.2253*	.0962	.1522
	Bottom	.0764	.0577	.0876	.0739
S.S., 100 ms (2)	Top	.0891	.0704	.0204	.0600
	Middle	.0816	.0544	.0448	.0603
	Bottom	.1081	.0527	.0818	.0809
S.S., 150 ms	Top	.1089	.0530	.0870	.0830
	Middle	.0478	.0815	.0385	.0559
	Bottom	.0636	.0849	.0484	.0656
S.S., overall					.0823
J.S., 100 ms	Top	.1018	.0398	.0389	.0601
	Middle	—	—	.0952	.0952
	Bottom	.0421	.0870	.0283	.0525
J.S., 150 ms	Top	.0252	.1120	.0511	.0628
	Middle	—	—	.1025	.1025
	Bottom	.1326	.0106	.0192	.0541
J.S., overall					.0633

Note. Values designated (by an asterisk) as statistically significant are those for which the observed Pearson correlation value deviates from the grand mean correlation of the given observer at a significance level of .05, without adjusting for the number (27 or 14) of z tests performed for each observer. For J.S., the center location of the middle row was excluded because of a universally low performance level for that location.

* $p < .05$.

attention cue, with a range of 100–400 ms. Attention shifted simultaneously to all three cued locations in an entire row; there was no tendency to scan from left to right or in any other systematic direction. The analysis of contingencies between letters reported from the same row indicated a small but consistent greater tendency to report letters from the same array than would be expected from the temporal uncertainty in individual reports. The similarity of attention shifts for the middle row (which involves no spatial movement of attention) and for the outer rows (which involve attention movements up or down) indicates that spatial distance traversed plays no measurable role in this kind of attention shift. These findings support a quantal (episodic) theory of attention shifts (e.g., Sperling & Weichselgartner, 1995).

Experiment 2: Partial Report

Experiment 2 was designed to determine whether the trajectories of the attention shifts measured in Experiment 1 could be applied to predict the outcome of a partial-report paradigm. A tonal cue was used to signal the observer as to which one of three rows to report from a briefly exposed 3×3 letter array. The traditional method of using a partial-report accuracy to directly indicate the duration of iconic memory confounds the decay constant of memory with the time constants of the attention shift to the row to be reported. (A prolonged iconic decay with a slow attention shift yields data similar to a faster decay with a faster attention shift.) By using the independent measure of attention shift time from Experiment 1, we can make an unconfounded estimate of the decay of iconic memory.

The decay constant of iconic memory derived by applying an a priori attention shift function can be compared to the decay constants of iconic memory estimated from different paradigms in which a poststimulus pattern mask follows a brief exposure (e.g., Gegenfurtner & Sperling, 1993; Irwin & Brown, 1987; Loftus, Busey, & Senders, 1993; Loftus, Duncan, & Gehrig, 1992). In masking methods, the total amount of information reported indicates the cumulative strength V of iconic memory up to the time of mask onset. The time derivative of V , dV/dt , gives the instantaneous value of iconic memory and can be used to estimate decay constants in masking (vs. partial-report) methods. Therefore, in addition to simple measurements of iconic memory in a partial-report paradigm, Experiment 2 investigated partial reports with masking stimuli over a wide range of mask delays. A comprehensive model, such as is proposed in the Model of Spatial Shifts of Visual Attention section of this article, ultimately must make consistent, good predictions for all these paradigms.

Method

Observers, Apparatus, and Stimuli

The 2 observers and the set of equipment were the same as in Experiment 1. The stimulus set was the same as in Experiment 1, except that Y was excluded. This exclusion was due to observers' complaints that Y was confused with X and K . Eighteen letterlike patterns, shown in Figure 2b, were used as poststimulus masks. Letter arrays were constructed in the same way as in Experiment 1. Mask arrays were composed of 9 randomly sampled (with replacement) patterns which were arranged in the same way as letter arrays (i.e., 3 rows \times 3 columns).

Procedure

Figure 6 illustrates the procedure and conditions for the present experiment. The observer pressed a key to initiate a trial whenever he or she was ready. A central fixation square was presented for 0.5 s and was followed by a blank screen of a variable interval between 0.5 and 1 s. A single 3×3 letter array was then presented for 50 ms (i.e., three refreshes). A 200-ms tonal cue and a stream of visual masks were respectively presented at a variable interval relative to the stimulus onset (see the *Design* section). The stream of mask arrays consisted of seven repetitions of a set of five different mask arrays that were constructed for each trial. Each mask array was presented for 16.7 ms (i.e., one refresh) yielding a total masking time of 583 ms per trial.

The observer was asked to perform an ordered recall task—that is, to report the three letters that were from the cued row and in correct column position. The report order was from left to right. Typed responses were presented on the screen, and the observer was allowed to correct them before pressing the enter key. Feedback was given after the ordered recall. The feedback included an ordered set of letters displayed in the cued row and the responses. Each observer completed 40 sessions of 100 trials each.

Design

Three variables were manipulated independently within observers: the cued row, the cue delay, and the mask delay. The cued row (top, middle, or bottom), indicated by a tonal cue, was the one from which the observer was instructed to report three letters. The cue delay ($-200, 0, 200, 400,$ or 800 ms) was the onset time of a tonal cue relative to the onset of the letter array. The mask delay (100, 200, or 400 ms or infinite) was the onset time of the stream of mask arrays relative to the onset of the letter array. No mask was presented when the mask delay was infinite. The dependent measure was the recall probability for each condition.

Results

Results are presented in Figure 7, a and b, respectively, for observers S.S. and J.S. The location of each panel in each figure corresponds to the location of 3 cued rows \times 3 column positions. Curve parameters are the cue delays. For each cue delay condition, the curve maps recall probability to mask delay. The guessing rate was $1/20$.

The results of the present experiment generally replicated findings of similar studies (for reviews, see Coltheart, 1980; Gegenfurtner & Sperling, 1993; Long, 1980). Performance (recall probability) increased as the mask delay increased for each cue delay, decreased as cue delay increased for each mask delay, and was best for the middle row. Analysis of the full implications of these results is deferred until after the model has been presented.

However, it is noteworthy that, for every row, performance failed to asymptotically approach fully correct report of all three letters even when the tone was presented 200 ms before the presentation of the letter array and no visual mask followed thereafter. Experiment 3 was designed to determine whether asymptotically imperfect performance was due to a memory limitation or to perceptual factors.

Experiment 3: Whole Report

The purpose of Experiment 3 was to determine how perceptual factors related to the retinal placement of elements in the 3×3 array might limit performance in reports of a row of three letters.

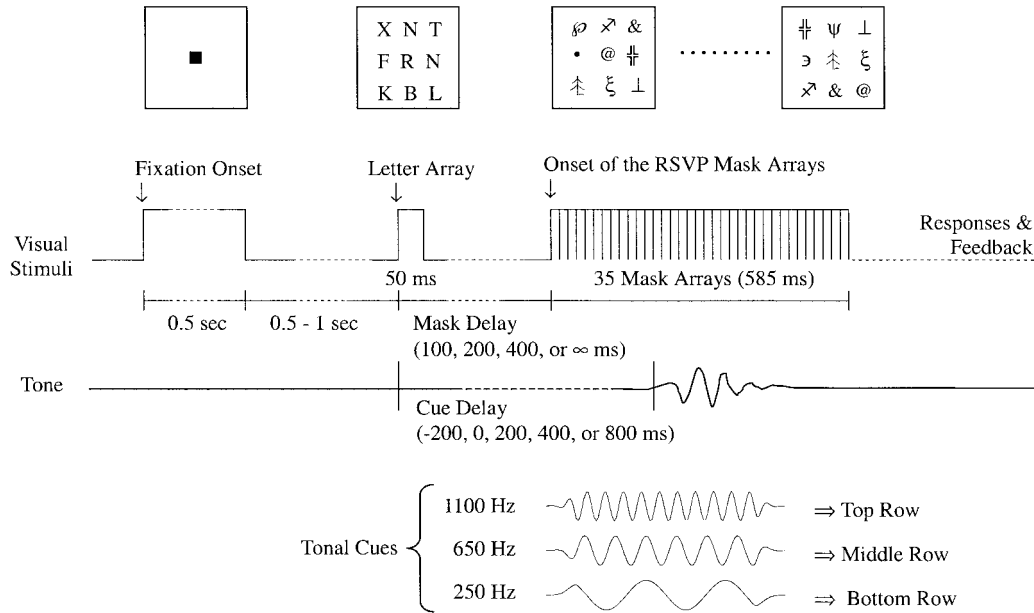


Figure 6. Procedure for Experiment 2: partial reports with four mask delays and five cue delays. An initial key press produces a central fixation square for 0.5 s, followed by a blank screen for a variable interval of between 0.5 and 1 s. A 3 × 3 letter array is presented for 50 ms (i.e., three refreshes). A 200-ms tonal cue is presented at one of five cue delays to indicate which row is to be reported. After one of four mask delays, a stream of 35 mask arrays is presented at 60 Hz for a total masking time of 583 ms. Feedback is given immediately after the responses. The correspondence between the tonal cues and cued rows is the same as in Experiment 1. RSVP = rapid serial visual presentation.

One expects the best performance when the row of three letters is presented in isolation, and the observer knows the location of the row well in advance of its presentation. Thus, the reference stimulus condition was the distractor-absent condition in which non-target letters were eliminated. That is, only three letters in one of three rows were presented; the other two rows were both blank. In order to evaluate a perceptual interference in 3 × 3 arrays, non-target letters were replaced with nonletter symbols in the stimulus array (the distractor-present condition). That is, only three letters in one of three rows were presented; the other two rows were both filled with nonletter symbols. The nonletter elements were assumed to be perceptually equivalent to letters, and they presumably did not exert the interference at higher postperceptual levels that letters might.

To determine how perceptual interference depends on stimulus availability, we independently varied the duration of iconic memory by varying the delay of visual masks. A fixed 250-ms interval between the fixation offset and the stimulus onset was used to eliminate the temporal uncertainty about the stimulus's occurrence. The presence and absence of a tonal cue was independently varied from trial to trial. The onset of a tonal cue, if present, coincided with the fixation offset (i.e., the cue delay equals -250 ms).

The -250-ms cue delay should give observers ample time to direct attention to the cued row before stimulus onset (see Sperling & Weichselgartner, 1995) but insufficient time to move their eyes in response to the tonal stimulus (because that would be a disjunctive reaction time [RT]). Thus, the performance of observers on the

upper and lower rows relative to the middle row provided a measure of perceptual losses at these locations.

Method

Observers, Apparatus, and Stimuli

The 2 observers, apparatus, the set of letter stimuli, and the set of masks were the same as in Experiment 2. The distractor was a letterlike symbol that was made of an X superimposed on an O. A stimulus array consisted of three letters in one of a 3 × 3 array and either distractors or nothing in the other rows. The configuration and size of a stimulus array were the same as in Experiments 1 and 2.

Procedure

The procedure and conditions for Experiment 3 are illustrated in Figure 8. The observer pressed a key to initialize a trial whenever he or she was ready. A central fixation square was presented for 2 s. If a tonal cue was present in a trial, it was always presented at the offset of the fixation square. At the offset of the fixation square, a blank screen was displayed for a fixed interval of 250 ms. The stimulus array was then presented for 50 ms. A stream of mask arrays was presented at a variable interval relative to the stimulus onset. The stream of mask arrays for a trial was constructed similarly to that in Experiment 2. The total masking time was again 583 ms per trial. The observer was required to perform an ordered recall task. That is, he or she reported the three stimulus letters in the correct column positions. The report order was from left to right. The typed responses were presented on the screen, and the observer was allowed to correct them before pressing the enter key. Feedback, including the stimulus letters and

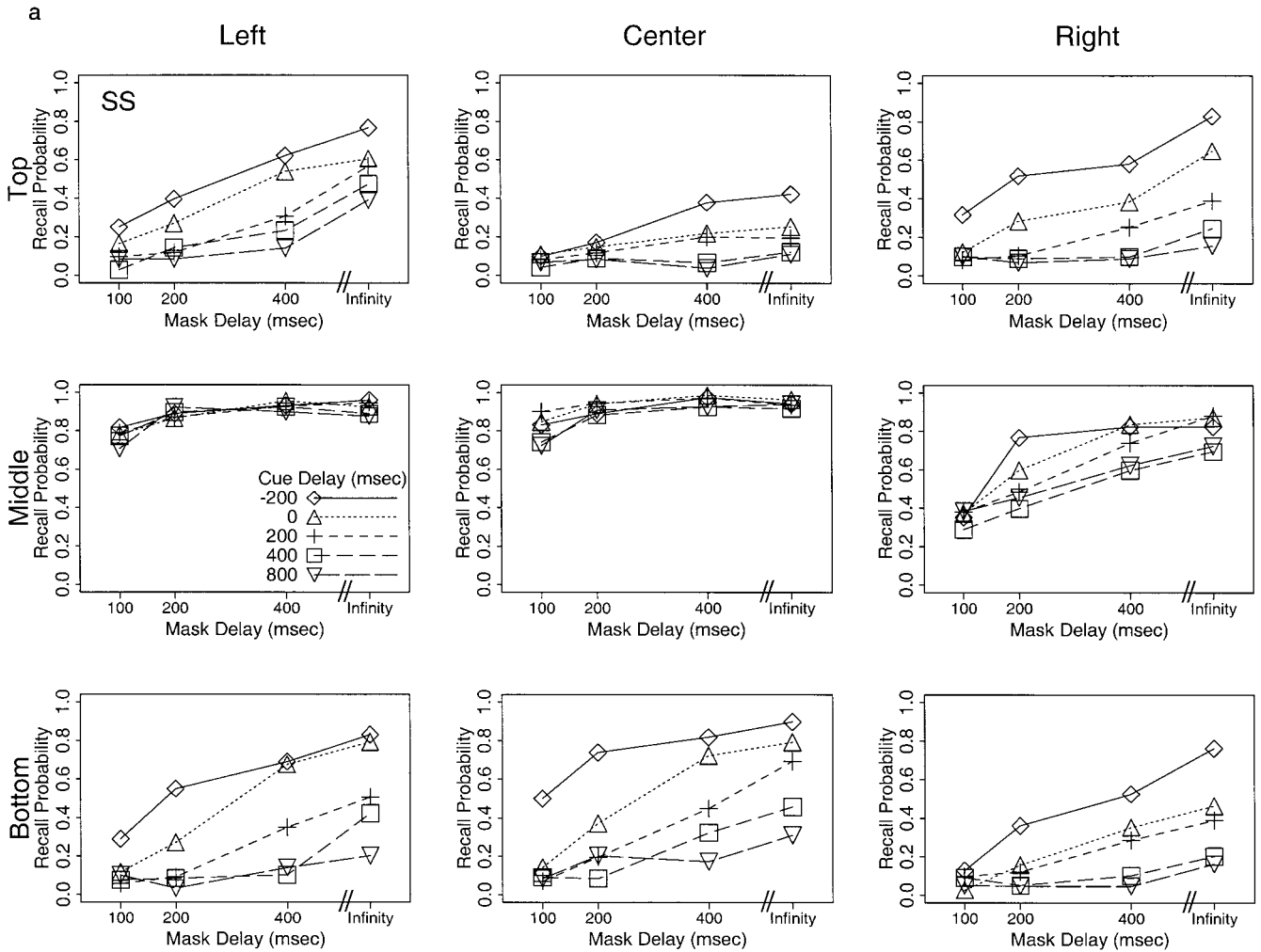


Figure 7. Data from Experiment 2: partial-report recall accuracy as a function of mask delay. The various curves represent different cue delays (see middle left panel). The location of each panel corresponds to the 3 row \times 3 column positions. (a) Observer S.S., (b) observer J.S.

the responses, was given after the ordered recall. Each observer completed 25 sessions of 100 trials each.

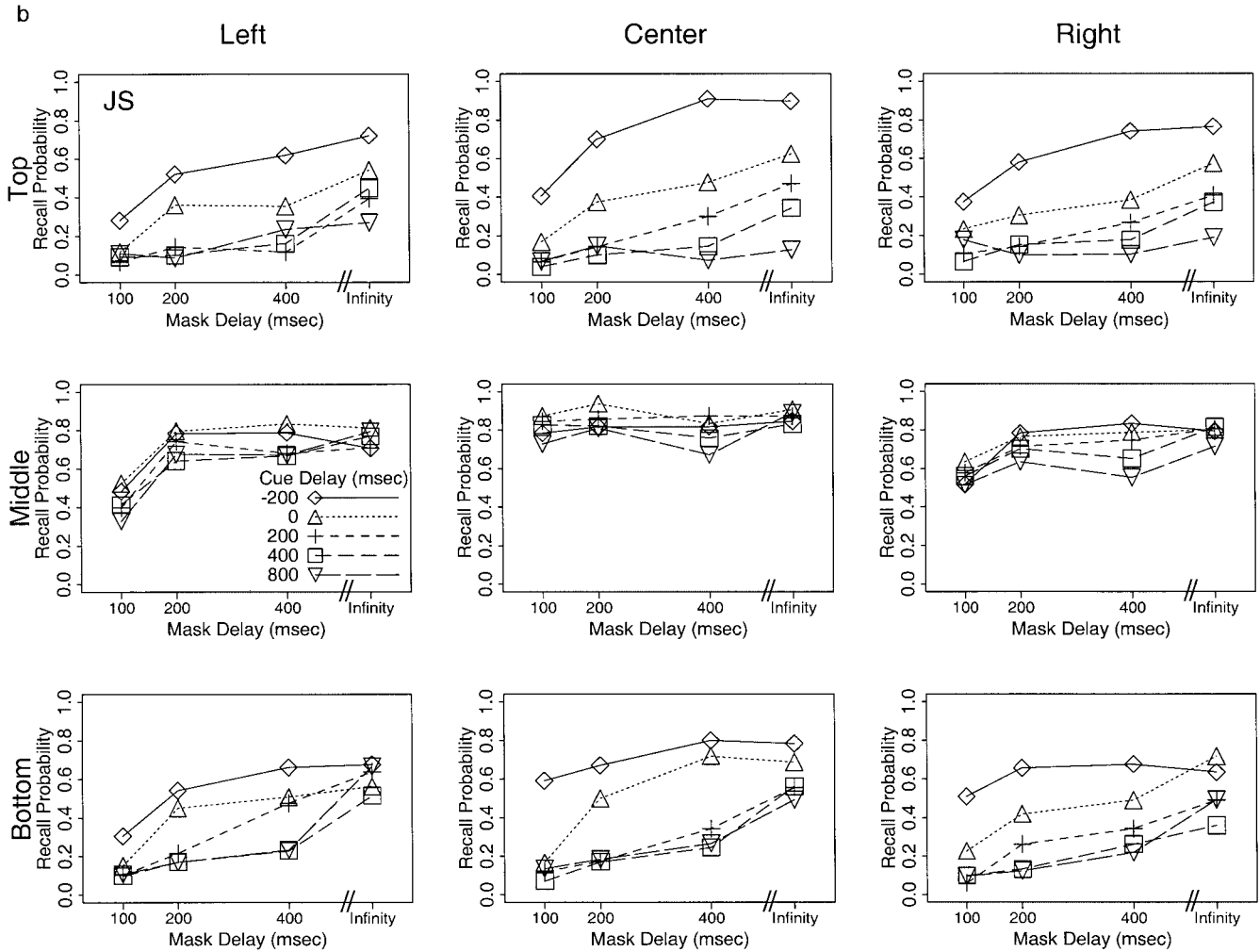
Design

Four variables were manipulated within observers, namely, the cued row, mask delay, presence-absence of a tonal cue, and presence-absence of distractors. The cued row (top, middle, or bottom) was the one in which three stimulus letters were presented. The mask delay (50, 100, 200, or 400 ms) was the onset time of a stream of mask arrays minus the onset of the stimulus array. A 200-ms tonal cue either was presented 250 ms before the stimulus or was absent in a trial. The correspondence between the pitch of the tonal cue and the cued row was the same as in Experiments 1 and 2. Distractors were either present or absent in the noncued rows in a stimulus array. Examples of stimulus arrays, with and without distractors, and of mask arrays are presented at the top of Figure 8. The dependent measure was the recall probability for each condition.

Results

Results for observer S.S. and J.S., respectively, are shown in Figure 9, a and b. In each part of the figure, the location of a panel reflects data gathered from the corresponding location in the 3 \times 3 stimulus array. Curve parameters are absence-presence of tonal cues and of distractors. Each curve shows recall probability as a function of mask delay. The guessing rate was 1/20. As expected, performance improved as the mask delay was increased, and performance was best in the middle row, followed by the top and bottom rows.

Effects of presence-absence of distractors and tonal cues were evaluated using a matched *t* test, whose sample size was 36 (3 Rows \times 3 Columns \times 4 Mask Delays). Performance for conditions with distractors was significantly ($p < .00001$) worse than it was for conditions without distractors: $t(35) = 6.49, p < .00001$, for S.S., and $t(35) = 8.32, p < .00001$, for J.S. For observer S.S.,



although performance for conditions with tonal cues was significantly better than it was for conditions without cues, $t(35) = 6.10$, $p < .00001$, the difference was not significant if the distractors were absent, $t(35) = 0.77$, *ns*. For observer J.S., there was little difference in performance between conditions with and without tonal cues, $t(35) = 0.13$, *ns*.

Almost perfect performance was achieved at the middle row when enough time was available for processing (mask delays of 200 and 400 ms). However, performance for the other two rows was still below the perfect level even in the easiest condition, that is, cue at -250 ms, without distractors, and with a mask delay of 400 ms. From these results, we conclude that the performance deficiency observed in full reports of the top and bottom rows, as well as in partial reports, was due not to a memory limitation but to factors related to perceptual acuity.

Performance in conditions with distractor characters was significantly worse than in conditions without distractors. We attribute this effect to perceptual interference. However, performance for the middle row was less impaired by distractors than performance for the top and bottom rows.

Model of Spatial Shifts of Visual Attention

We propose a strength model of visual selective attention to relate the observers' performances in the current study to underlying processes of perception and attention.³ The three experiments provided 486 (for S.S.) and 420 (for J.S.) primary data points that consist of the probabilities of correct reports. In addition, there were another 972 (for S.S.) and 504 (for J.S.) data points contained in the contingency tables for Experiment 1. The breakdown of these data points is presented in the *Data To Be Accounted For* section. Our aim was to quantitatively describe all these data with a single model and single, common core of parameters.

The exposition is divided into three sections: (a) a summary of basic assumptions, (b) an overview of model parameters, and (c) a stage-by-stage explanation of the model. Estimation of the model's parameters is considered in Appendixes B, C, and D.

³ Preliminary reports of this theory appeared in Shih and Sperling (1995) and Sperling and Shih (1998).

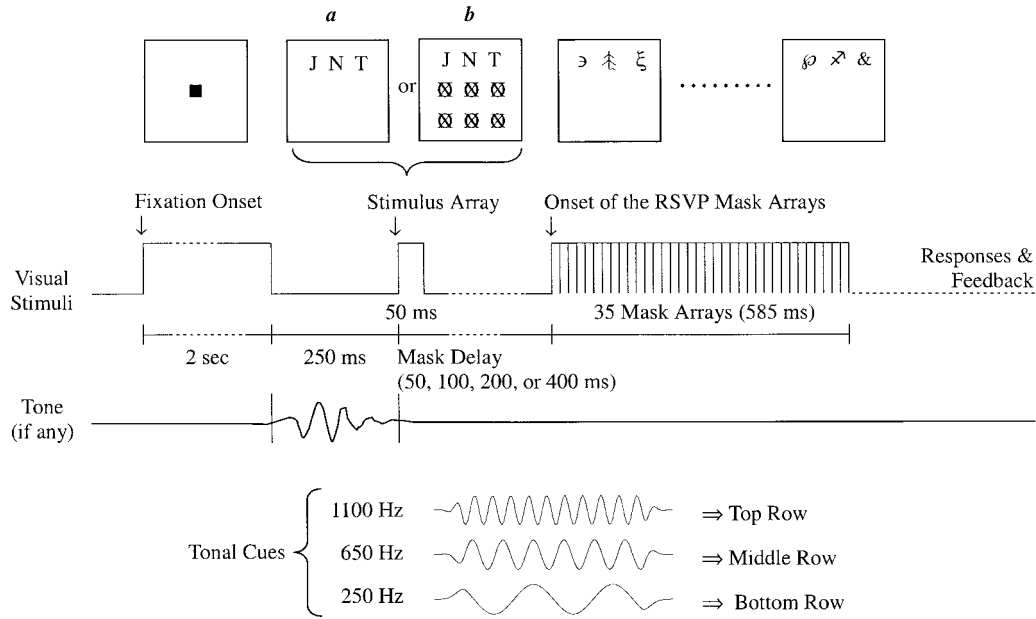


Figure 8. Procedure for Experiment 3: 16 whole-report conditions. An initial key press produces a central fixation square for 2 s followed by a 250-ms blank screen. One of two types of stimulus array is presented for 50 ms (three refreshes). A stream of mask arrays occurs after one of four mask delays. Feedback is given after the responses. With probability .5, a tonal cue is presented 250 ms before stimulus onset. The stimulus consists of one row of three letters and either two rows of distractors or two rows of blanks. The distractors are superimposed letters X and O. Masks occur only at the location of the cued row, and tonal cues always indicate the cued row. RSVP = rapid serial visual presentation.

Basic Assumptions

1. The rate of perceptual processing varies as a function of retinal location (perceptual acuity factor).
2. All stimulus letters are represented perceptually in iconic memory, which decays exponentially.
3. Only some letters in iconic memory are transferred to visual short-term memory (VSTM). The spatial and temporal characteristics of the transfer from iconic memory to VSTM are determined by the attention-gating function, which opens a gate that allows information to flow from iconic memory to VSTM simultaneously from all locations.
4. The instantaneous flow of information from iconic memory to VSTM is given by the product of stimulus availability and the current value of the attention-gating function.
5. Integrating instantaneous information flow over the interval during which a letter is perceptually available gives the letter's strength value.
6. Strength is perturbed by additive internal noise that represents the inaccuracy of a letter's representation in VSTM.
7. Some nonstimulus letters (distractors, e.g., residual letters from previous trials) are present in VSTM.
8. A response for each retinal location is generated by selecting the letter (stimulus or distractor) with the highest strength at that location.

The attention-gating function, which is defined in Assumption 3, arises from three successive attention states, and it has to be elaborated to deal with temporal variability and information trans-

fer prior to any attentional cue. A more complete specification of Assumption 3 follows.

- 3a. The attention-gating function is determined by a sequence of three consecutive attention states (i) pre-cued state (await and interpret attention cue), (ii) cued state (transfer from iconic memory to VSTM), (iii) recovery state (stop transferring). In Experiments 2 and 3, where only a single letter array is presented, the third state is sufficiently delayed that only the first two states are relevant.
- 3b. The time taken to interpret the attention cue and to initiate the gating function has random variation from trial to trial.
- 3c. In the three experiments, the tonal cue is assumed to initiate the attention-gated transfer to VSTM of the row indicated by the cue. When the onset of the visual display precedes the cue, transfer begins with default parameters.

All these relations, illustrated in Figure 10, are explained below.

Model Parameters

Overview

This section briefly outlines the evolution and meaning of the model's parameters. A detailed discussion and the functional form (or equation) for each component is presented in the subsequent *The Model, Stage by Stage* section.

Primary and Secondary Parameters

There are two kinds of model parameters: Primary parameters determine the functional properties of attention and decision mech-

anisms. Secondary parameters are needed in order to apply the model to the various different visual displays and procedures, to reflect the observers' task-dependent strategies, and to describe changes in primary parameters with practice. Parameters are represented by Greek letters.

Primary parameters. Six primary parameters of the model were estimated, and one was chosen arbitrarily as a reference against which to measure the others. The six estimated parameters are the overall sensitivity γ ; the cue interpretation time τ ; the width ν and slope (proportional to λ^{-2}) of the true attention window; the time constant α of iconic memory; and σ_s , the standard deviation of strength-noise in VSTM. The strength of the strongest distractor in memory μ_d is arbitrarily chosen as 100 ms. The shape of the probability distribution for cue interpretation times is taken as the Gamma function that described the observer's motor reaction times (MRTs); the variance is taken as the data variance of MRTs.

Secondary parameters. These deal with retinotopic inhomogeneities and experiment-specific details. For example, the information-processing rate varies over the nine spatial locations in which letters are displayed in the three experiments, and these inhomogeneities vary from observer to observer. Such inhomogeneities are unimportant to a theory of attention. On the other hand, it would be unfortunate if inaccurate estimation of visual inhomogeneities were to produce poor fits of the model to data. Therefore, we arbitrarily use eight parameters, $\gamma(r, c)$ for row r and column c , to describe the processing rates at the eight noncentral letter locations relative to the middle location.

The stimuli contained nine letters in a 3×3 array in all conditions of Experiments 1 and 2. In Experiment 3, all stimuli had one row of letters and the six remaining locations contained either blanks or nonletter characters. In the model, the nonletter characters are treated equivalently to letter characters. Blanks adjacent to a letter make it easier to process than do adjacent characters, and this is reflected in a rate parameter ω_{E3} that multiplies the $\gamma(3, c)$. The subscript E3 indicates it is specific to Experiment 3.

In Experiments 2 and 3, only one letter array was presented to the observer and it was preceded by a blank field; in Experiment 1 the target array was embedded within a sequence of more than 20 letter arrays and was always preceded by a letter array. The different information-processing rates in Experiment 1 relative to Experiments 2 and 3 are represented by two parameters, $\omega_{E1}(t/b, \cdot)$, which represents the altered rate for the top and bottom rows, and $\omega_{E1}(m, \cdot)$, which represents the altered rate for the middle row. Also, modeling Experiments 2 and 3 requires only two attention episodes; Experiment 1 requires three: (a) await and process attention cue, (b) open access to VSTM of cued row, and (c) terminate access to VSTM. The time duration between onset of State 2 (cued) and State 3 (recovery) in Experiment 1 is expressed by the parameter ν . In Experiment 2, the observer is sometimes forced to select letters for report before an attention cue has occurred. The parameters $\beta(r, \cdot)$ $r = 1, 2, 3$, represent the observer's a priori distribution of attention to row r . Finally, the parameter $\Delta\tau$ is an adjustment to cue interpretation time τ in Experiment 1 that represents two different levels of practice.

Evolution of Parameters

For illustration, suppose that a basic set of parameters was first determined to account for performance in Experiment 1, the choice attention-gating experiment. Experiment 1 provided excellent estimates of the attention-gating parameters but very poor estimates of iconic memory, because no stimulus was on for more than 150 ms before it was overwritten by a subsequent stimulus. Therefore, Experiments 2 and 3 required a new parameter that was not important for Experiment 1: the decay constant of iconic memory (α).

In Experiment 1, it was not useful to transfer any stimulus characters into VSTM before the cue because there were far too many characters, and VSTM would be completely overloaded. In Experiments 2 and 3, there were never more than nine characters presented, so observers did transfer letters into VSTM prior to the cue. This required parameterizing the precued state of attention, which we did in terms of a priori biases to report particular rows [$\beta(r, \cdot)$].

The visibility of a frame within a rapid sequence (in Experiment 1) differs from the visibility of isolated flashes (in Experiment 2); this requires two parameters because the difference is not the same in the middle [$\omega_{E1}(m, \cdot)$] versus the upper and lower rows [$\omega_{E1}(t/b, \cdot)$]. Finally, in Experiment 3, some frames had only one row. This required a parameter ω_{E3} to account for the different visibility of letters in a row surrounded by empty space relative to a row within a full array.

In actual fact, however, parameters were estimated jointly for all three experiments. The experimental variables and the parameters are summarized in Table 5, which shows how they apply to the various experiments and conditions.

The Model, Stage by Stage

Perceptual Processing

Stimulus contrast and relative perceptual acuity. For a wide range of illumination, visual processing rate, to a first approximation, is independent of the absolute luminance and dependent on point contrast, the normalized deviation of stimulus points from the mean luminance of the display. Our stimuli were bright letters on a dim surround, producing a very high contrast that remained constant for all conditions. Rather than getting sidetracked with a theory of contrast saturation, we simply define the effective contrast of our stimuli as 1. The effective letter contrast $L(r, c, k, t)$ at a location (r, c, t) in frame k is 1.0 if there is a letter being presented in row r , column c , at time t ; otherwise, contrast is zero.

The perceptual strength of a stimulus letter is determined by the product of effective contrast $L(r, c, k, t)$ and the local visual processing capacity. The maximum processing capacity (sensitivity) is γ , and it applies to the central location. The local processing capacity is defined relative to the central location, as $\gamma \times \gamma(r, c)$, where r, c represents the locations in the 3×3 arrays. For the central location, $\gamma(2, 2) = 1$ by definition; for the other eight locations, $\gamma(r, c) < 1$, representing a peripheral loss of processing capacity. There are eight local processing-rate parameters $\gamma(r, c)$. Undoubtedly, relative perceptual acuity could be characterized by fewer than eight parameters, but it is not our goal here to develop a theory of peripheral falloff of perceptual processing capacity,

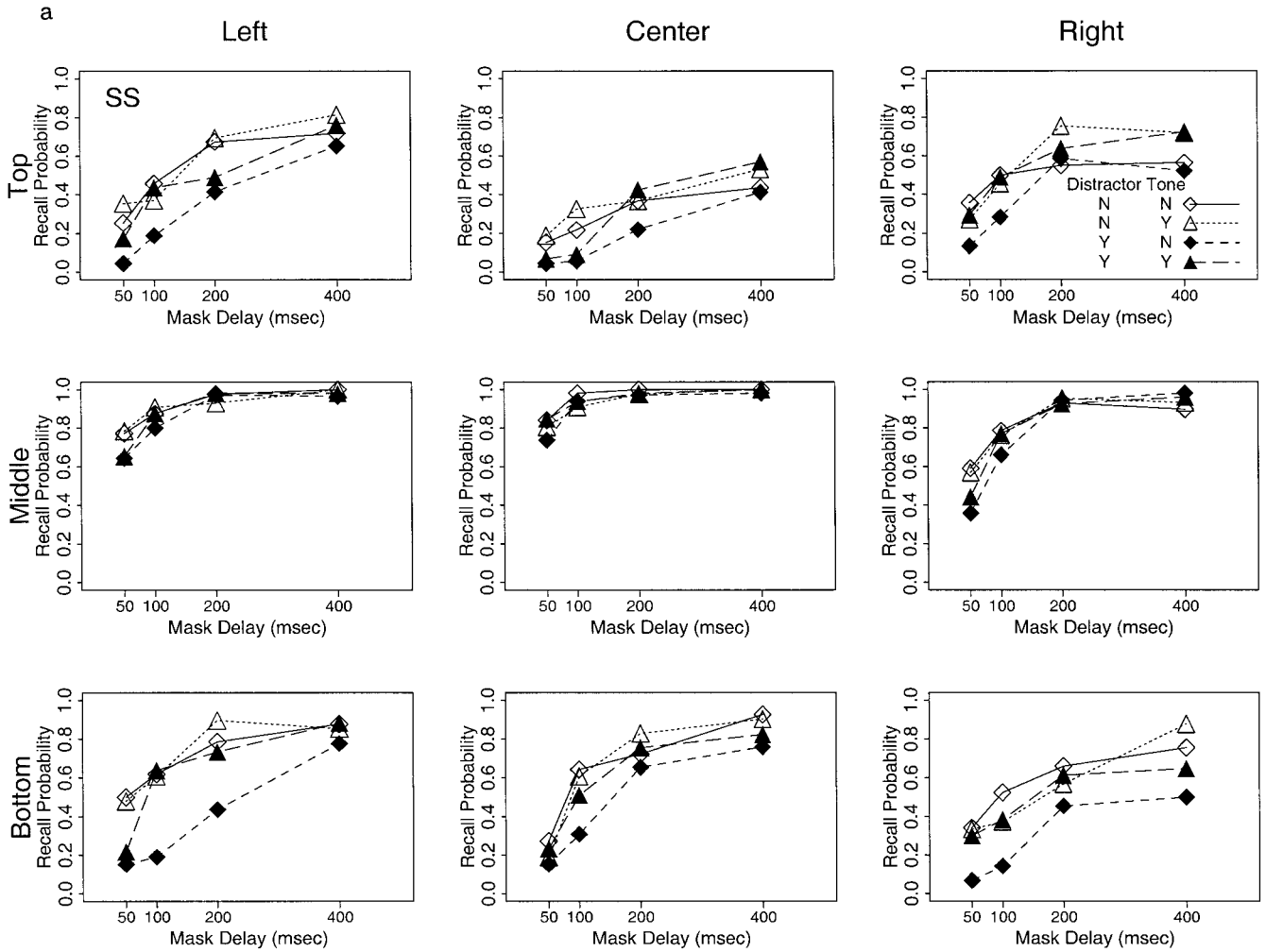


Figure 9. Experiment 3: whole-report recall probability as a function of mask delay. The different curves represent presence (Y) and absence (N) of masks and of tonal cues (see upper right panel; N = no; Y = yes). The location of each panel corresponds to the 3 rows \times 3 column positions. (a) Observer S.S., (b) observer J.S.

merely to represent it accurately. In the model, local processing capacity is assumed to remain constant over all experiments and conditions; that is, it represents an intrinsic retinotopic capacity limit that determines the local rate of information processing in each of the various types of stimulus presentation.

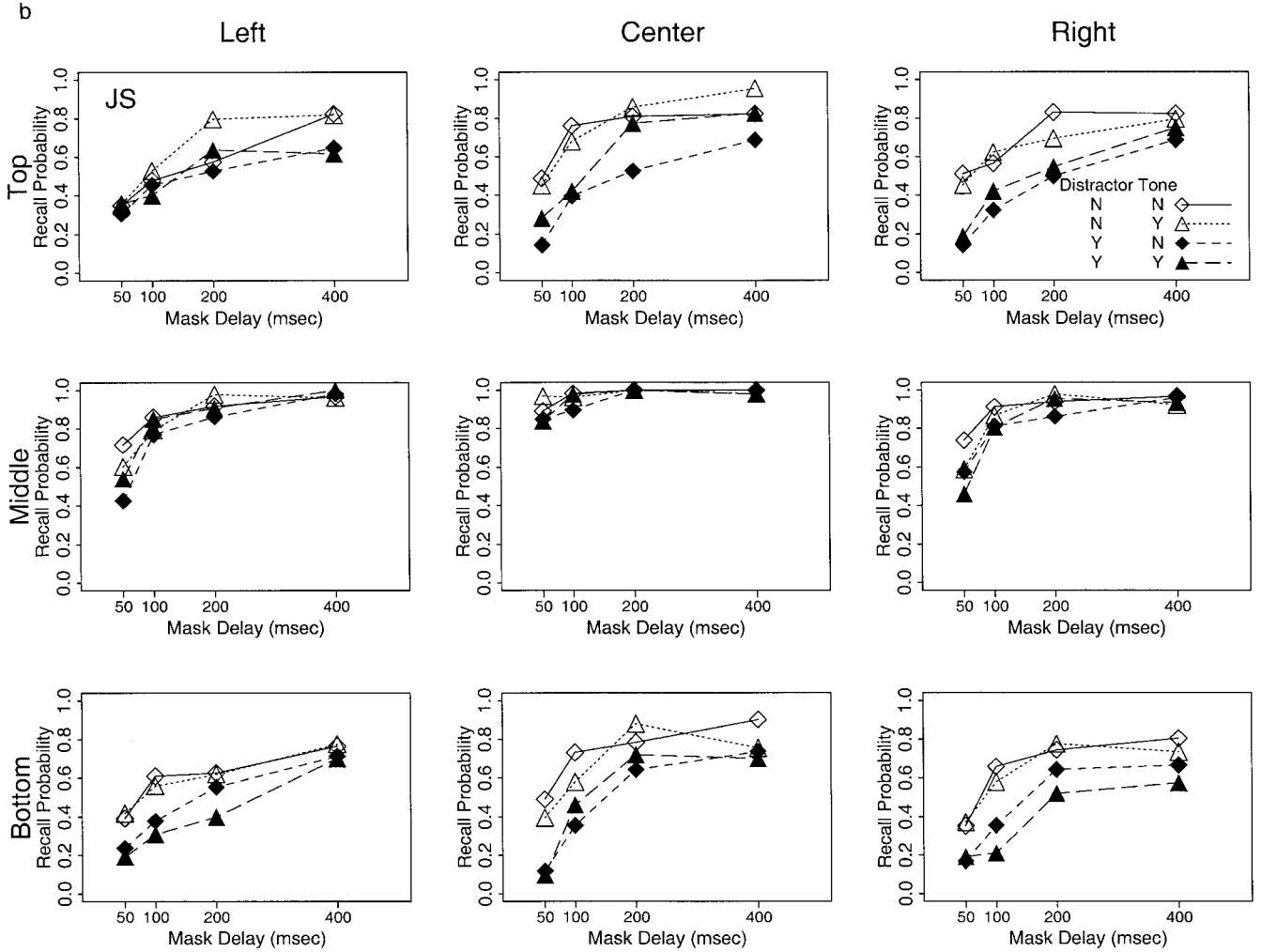
Iconic decay and stimulus availability. For specificity, let k be an integer index on the frames in a display sequence, with $k = 0$ representing the target frame, positive integers representing frames after the target, and negative integers representing frames before the target. Three steps are involved in computing the stimulus availability $B(k, t)$ for a stimulus letter in frame k as a function of time t .

1. Stimulus contrast $L(r, c, k, t)$ takes only the values $\{0, 1\}$. $L = 1$ during any refresh interval in which a letter is painted on the monitor screen; $L = 0$ otherwise. L is multiplied by local processing rate $\gamma(r, c)$.

2. Perceptual interference is parameterized. The brief exposure of a 3×3 array following a blank preexposure field in Experi-

ments 2 and 3 is regarded as the standard reference condition. To account for the differences in information-processing rates between the standard condition and the sequential displays of Experiment 1, the model uses two secondary parameters ω (one for more central areas, the other for more peripheral areas) that multiply $L\gamma(r, c)$. The parameter $\omega_{E1}(t/b, \cdot)$ multiplies the top and bottom rows; $\omega_{E1}(m, \cdot)$ multiplies the middle row; E1 indicates Experiment 1. In Experiment 3, the distractorless displays (in which six of nine letters were replaced with empty space) are multiplied by ω_{E3} to represent faster processing in distractorless displays relative to the standard condition, which has distractors.

3. After a letter in frame k is turned off, its availability does not go immediately to zero but decays exponentially according to the decay constant α of iconic memory (Loftus et al., 1992). Once a subsequent letter or masking stimulus is superimposed on a letter in frame k , its visibility is assumed to immediately go to zero. Thereby, stimulus availability $B(k, t)$ of a letter in frame k involves,



first, the product of five factors: effective stimulus contrast L , perceptual sensitivity γ , local processing rate $\gamma(r, c)$, iconic decay D , and display interference $\omega_{En}(r)$; and second, the time of presentation of a subsequent item at the same location, $L(r, c, k + 1, t)$:

$$B(k, r, c, t) = D[\alpha, L(r, c, k, t)\gamma\gamma(r, c)\omega_{En}(r), L(r, c, k + 1, t)]. \quad (1)$$

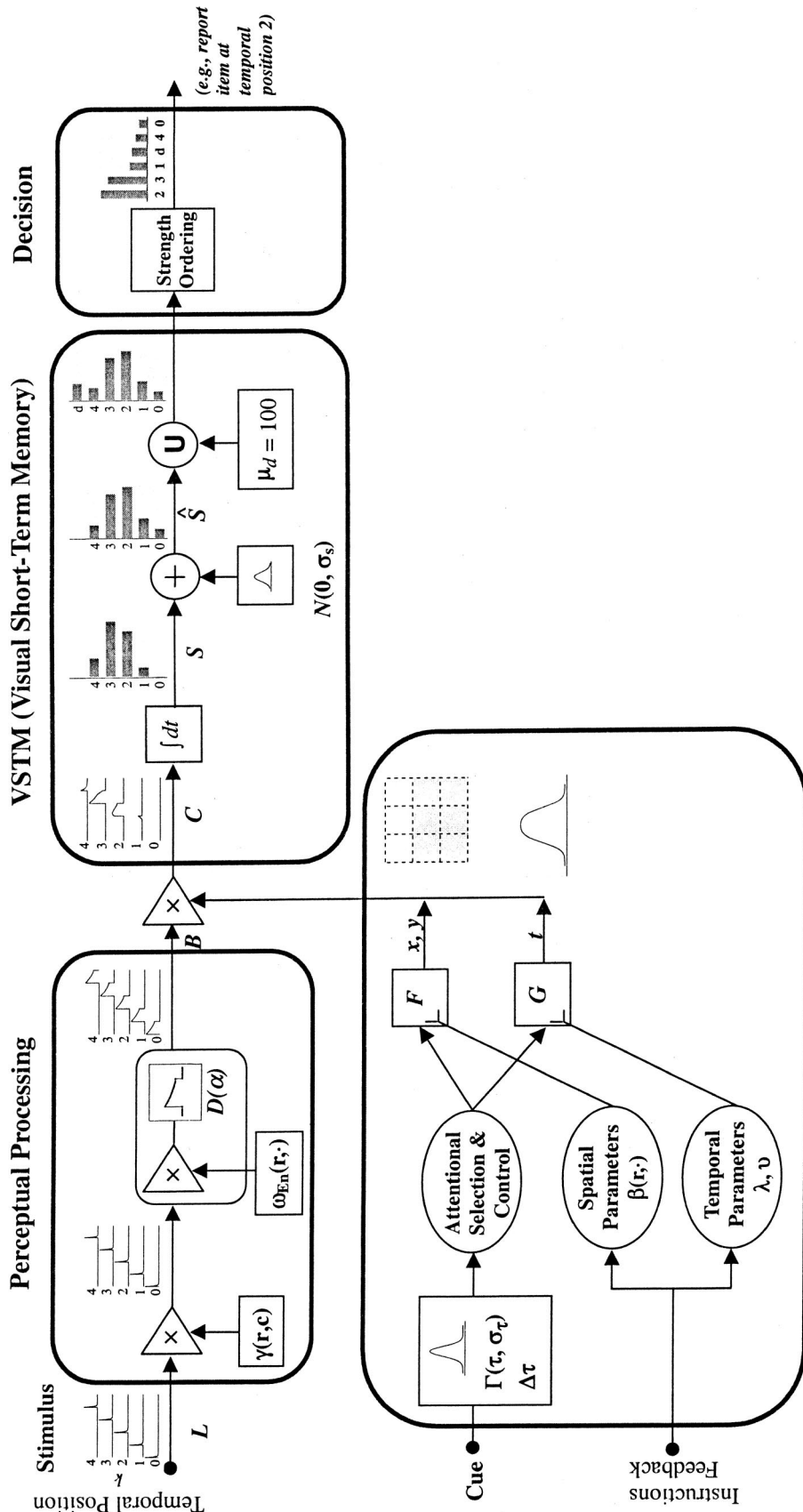
The subscript En represents the Experiment number (1 or 3), and Greek letters represent estimated parameters.

Attention Components

Episodic attention theory. Experiment 1 measured the time course of attention shifts in response to tonal cues. A priori, in modeling the time course of such attention shifts, any space-time trajectory might be possible. Here, we rely on the episodic theory of Sperling and Weichselgartner (1995) and the results of Experiment 1 to restrict and characterize the space-time trajectory of attention. According to the episodic theory, visual attention con-

sists of a sequence of discrete attention episodes. Each episode is characterized by a real-valued function $F(x, y)$ that describes the spatial distribution of attention for that episode. In the original theory, the apparently smooth shift of attention from one location x_1 to another location x_2 resulted from the gradual transition between just two successive discrete states in which attention was focused first on x_1 and then x_2 . In the present theory, the smooth transition between two attention states is produced by two processes acting jointly: the intrinsically smooth rise and fall of attention (parameter λ) and the probabilistic mixing (over trials) of different attention switching times due to the variability σ_t^2 of cue interpretation time.

Three attention states generate an attention window. Applied to Experiment 1, the episodic attention theory requires three relevant states of attention (Episodes, E) defined relative to the onset of the tonal attention-directing cue: E_0 , a precued state (await and interpret attention cue); E_1 , a cued state (transfer from iconic memory to VSTM); and E_2 , a recovery state (stop transferring). Corresponding to each attention state E_j is a function $F_j(x, y)$ that describes the spatial allocation of attention during the interval that the state j is in effect. In the recovery state $j = \text{post}$, it is assumed



Attention Control Mechanism

Figure 10. Strength model of visual selective attention. Arrows indicate the direction of information flow. Complex operations are indicated by boxes, multiplication by triangles, and simple operations by circles. Model parameters that must be estimated are indicated by Greek letters. Functions and transformations are indicated by capital letters. L (luminance) represents the visual input as a function of space and time. γ represents the processing rate at the fixation point; $\gamma(r; c) < 1$ describes the peripheral loss of perceptual processing rate in row r , column c of the stimulus relative to the central location. The visual input is subject to perceptual interference. In Experiment 1, interference from a nonblank prior stimulus is $\omega_{E1}(m, \bullet)$ (middle row) and $\omega_{E1}(l/b, \bullet)$ (top and bottom rows). Subscript E1 indicates Experiment 1. In Experiment 3, absence of spatial interference from distractor elements is ω_{E3} . Subscript E3 indicates Experiment 3. D indicates the persistence function of iconic memory (decay time constant α), and B indicates the resulting stimulus availability. After a cue interpretation time that has a Gamma distribution (of order j) with mean τ and variance σ_τ^2 , the attention control mechanism outputs a true temporal attention window of width ν and slope λ^{-2} (the curve adjacent to G). In Experiment 1, τ decreased with practice; this is represented by the parameter $\Delta\tau$. Feedback gradually influences attentional parameters; this is indicated by the “synapses” terminating in boxes F and G . In Experiments 2 and 3, the precue row biases are $\beta(r; \bullet)$. Central availability C is the product of the stimulus availability B and the true attention window $F(x, y)G(t)$. The integral of C over time ($\int dt$) gives the initial strength S of an item in visual short-term memory (VSTM) to which Gaussian noise σ_s is added. The noise-perturbed stimulus items share VSTM with intrinsically generated distractor items, the largest of which has strength $\mu_d = 100$ ms. The decision mechanism is a strength ordering process. At each location, it outputs items in order of decreasing strengths.

Table 5
Experimental Variables and Parameters

Major variables		Parameters
Experiment 1: Choice attention gating		
Cued row	v	Width of true attention window
Stimulus onset asynchrony	λ	$(\text{Slope})^{-1}$ of true attention window
	τ	Mean of cue interpretation time
	$\Delta\tau$	Variation of τ due to strategy
	σ_τ	Standard deviation of the cue interpretation time
	γ	Perceptual sensitivity
	$\gamma(r, c)$	Perceptual acuity re $\gamma(2, 2)$
	$\omega_{E1}(t/b, \bullet)$	Scalar to $\gamma(r, c)$ for the top and bottom rows due to spatial-temporal interaction
	$\omega_{E1}(m, \bullet)$	Scalar to $\gamma(r, c)$ for the middle row due to spatial-temporal interaction
	σ_s	Standard deviation of internal noise
	$\mu_s = 0$	Mean of internal noise
	$\mu_d = 100$	Mean of distractors
Experiment 2: Partial reports		
Cued row	λ	$(\text{Slope})^{-1}$ of true attention window
Mask delay	$\beta(r, \bullet)$	Precued attention state
Cue delay	τ	Mean of the cue interpretation time
	σ_τ	Standard deviation of the cue interpretation time
	α	Iconic decay time constant
	γ	Perceptual sensitivity
	$\gamma(r, c)$	Perceptual acuity re $\gamma(2, 2)$
	σ_s	Standard deviation of internal noise
	$\mu_s = 0$	Mean of internal noise
	$\mu_d = 100$	Mean of distractors
Experiment 3: Whole reports		
Target row	λ	$(\text{Slope})^{-1}$ of true attention window
Mask delay	$\beta(r, \bullet)$	Precued attention state
Presence of distractors	τ	Mean of the cue interpretation time
	σ_τ	Standard deviation of the cue interpretation time
	α	Iconic decay time constant
	γ	Perceptual sensitivity
	$\gamma(r, c)$	Perceptual acuity re $\gamma(2, 2)$
	ω_{E3}	Scalar to $\gamma(r, c)$ due to the absence of spatial distractors
	σ_s	Standard deviation of internal noise
	$\mu_s = 0$	Mean of internal noise
	$\mu_d = 100$	Mean of distractors

that no letters are entered into VSTM so attention allocation is represented by $F_{\text{post}}(x, y) = 0$. Let t_j be the instant at which an attention state E_j comes into effect. The rate at which it comes into effect is described by a (cumulative probability distribution) function $G(t - t_j)$ that goes monotonically from 0 to 1 as t increases. The next state comes into effect at time t_{j+1} according to $G(t - t_{j+1})$. Thereby, the effective duration of E_j is $G(t - t_j) - G(t - t_{j+1})$, and this difference describes an attention-gating function. These relations are illustrated in Figure 11, a-c.

The experiments involve only the discrete case of three rows times three columns, and attention is directed only to a whole row at one time, with exactly the same attention-gating function $G(t -$

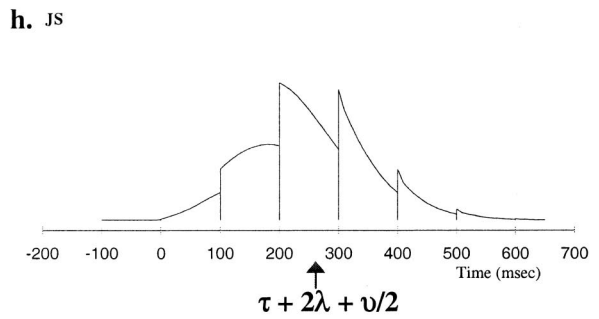
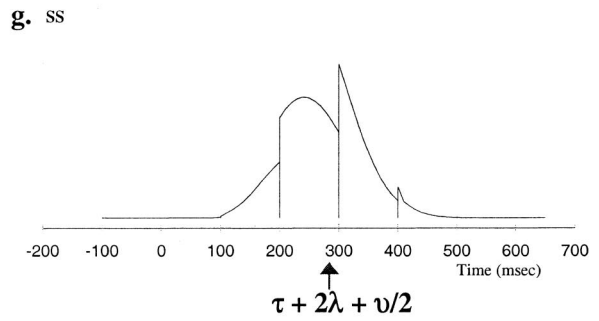
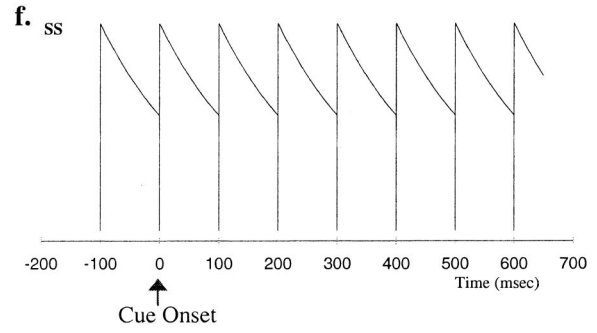
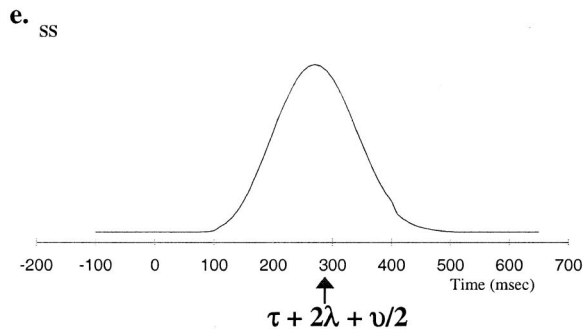
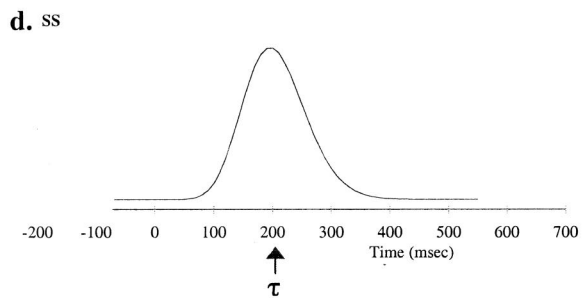
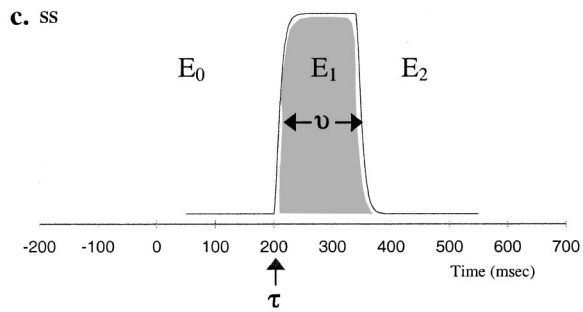
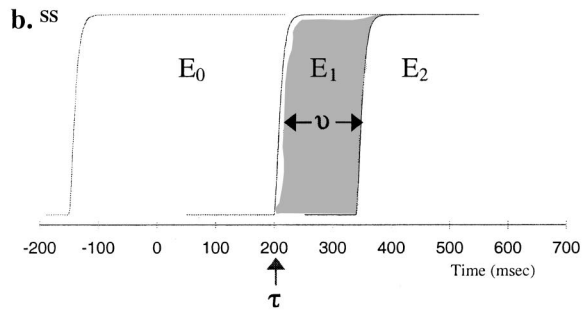
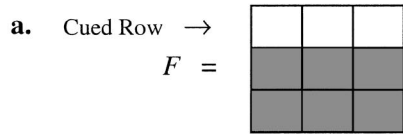
$t_j) - G(t - t_{j+1})$ being applied to the different columns. Therefore, it is convenient to replace x and y in $F_j(x, y)$ with simply a row index r .

The precued spatial attention function $F_{\text{pre}}(r)$ represents the bias to report letters from a given row prior to receiving and interpreting an attention cue (relevant to Experiments 2 and 3):

$$F_{\text{pre}}(r) = \beta(r, \cdot), \beta(r, \cdot) \subset \{\beta(1, \cdot),$$

$$\beta(2, \cdot), \beta(3, \cdot)\}. \quad (2)$$

In Equation 2, r takes the values t, m, b , representing the top, middle, and bottom rows, respectively. The dot indicates the same



β applies to all three columns in the indicated row. The β are nonnegative real numbers, representing the initial row-bias parameters (estimated from the data).

The cues of Experiment 1 direct the observer’s attention to one of three cued rows $r_c = \{\text{top, middle, bottom}\}$. Once the observer has interpreted the attention-directing cue, he or she is in the “cued” attention state A_{cued} . The spatial distribution of attention F_{cued} in A_{cued} depends, obviously, on the cued row. We assume F_{cued} takes the value 1 for the cued row and 0 for all other rows:

$$F_{\text{cued}}(r) = 1 \quad \text{if } r = r_c \text{ and } 0 \text{ otherwise.} \quad (3)$$

The attention-gating function. When the same kinds of trials are presented repeatedly in attention-gating experiments, such as Experiment 1, the observer sometimes reports items from a frame that begins 200 ms after the attention cue and sometimes from frames that begin 300 or more ms later. This variability in response is here modeled by two processes: first, variability in cue interpretation time and, second, the duration of the attention window that encompasses several items.

Even if an attention switch were instantaneous, if it occurred at different moments on different trials, because of variability in cue interpretation time, the data would mimic a continuous transition from one attention state to the next.

The gradual opening and closing and consequently relatively long duration of an attention window is evident in experiments (e.g., Reeves & Sperling, 1986) in which, on a single trial, the observer reports four items from a rapid stream of items. Suppose, for a given stream, the most frequently reported item (also the item most frequently reported first in the response) comes from Position 4 after the attention cue. Subsequently (and less frequently) reported items will come from Positions 5 and 3, and then 6 and 2, and so on, describing a gradual opening and closing of a temporal attention window centered on Item 4.

The two processes, (a) the temporal shape of the true attention window and (b) variability in switching time, can be distinguished by means of the contingency data of Experiment 1. This aspect of the theory is fully developed in Appendix B. For the moment, only the following considerations are important: (a) Our numerous explorations have shown that the shape of an attention-gating function is not highly constrained by the present data; therefore, it has been convenient to assume (b) that, except for a shift in mean, the distribution of cue interpretation times is given by the distribution of MRTs and (c) that the true attention window (as in

Sperling & Weichselgartner, 1995) is the difference between two second-order cumulative Gamma functions with time constant λ and separation v (variance $2\lambda^2 + v^2/12$).⁴ To a first approximation, the attention-gating function is formed by the convolution of these components, and its variance Σ^2 is the sum of the component variances $\Sigma^2 = \sigma_\tau^2 + 2\lambda^2 + v^2/12$ (see Figure 11).⁵ Again, to a first approximation, only the attention-gating function is needed to predict the primary data; the decomposition of Σ^2 into its component variances is needed only for the contingency data of Experiment 1 (joint probabilities of pairs of responses).

Given the assumptions above, the cued attention state E_1 is defined by the product of the spatial filter $F_{\text{pre}}(r)$ and the temporal attention window—attention-gating function, $G(t - t_1) - G(t - t_2)$ —as follows: Let $\varphi(x)$ be a second-order Gamma function, and let $G(t - t_i)$ be a cumulative Gamma distribution with starting time t_i :

$$\varphi(t) = \frac{1}{\lambda^2} t e^{-t/\lambda} \quad \varphi(t) = 0, \quad t < 0 \quad (4a)$$

$$G(t - t_i) = \int_{-\infty}^{t-t_i} \varphi(x) dx \quad (4b)$$

$$E_1 = A_{\text{cued}}(r, t) = F_{\text{cued}}(r) [G(t - t_{\text{cued}}) - G(t - t_{\text{post}})], \quad (5)$$

where $t_{\text{cued}} = \tau$ and $t_{\text{post}} = \tau + v$.

⁴ The variance of the attention window is composed of the sum of two variances, the variance of the second-order Gamma function $2\lambda^2$ plus the variance $v^2/12$ contributed by the time interval v between the beginning and end of the attention window (i.e., by a rectangular distribution of width v).

⁵ Convoluting the true attention window with the distribution of cue interpretation times is strictly correct only when the output (the probability of reporting a letter) is strictly proportional to the height of the function (as in Bundesen’s, 1990, theory) and when considering the report of only a single item. In the present model, the transformation from height of the attention-gating function to probability of report depends on the magnitude of memory noise; that is, there is an extra degree of freedom in the transformation, and so that convolution is only an approximation. In fitting the model, this approximation was used to hasten the computation for finding optimum parameters. The final evaluation of model parameters used an exact (Monte Carlo) computation.

Figure 11 (opposite). Deriving visual short-term memory (VSTM) inputs from the attention-gating function and iconic inputs. (a) Spatial allocation of attention; attention to the top row is indicated. (b) Three consecutive episodes of attention. The time indication of the x -axis is relative to the onset of the tonal cue: E_0 , precued, awaiting the attention cue; E_1 , items from cued row enter VSTM; E_2 , postcued, items stop entering VSTM. v is the width of the attention window; the slope is determined by a second-order Gamma function with time constant λ . τ is the mean cue interpretation time (observer S.S.). (c) The true attention window transposed down from (b) and placed directly on the axis, assuming zero variance for cue interpretation time τ (observer S.S.). (d) The assumed probability density function of cue interpretation time (based on motor reaction-time data; observer S.S.). (e) The attention-gating function derived by convoluting (c) and (d); τ is subtracted from (c) and (d) before convoluting and then restored (observer S.S.). (f) A temporal sequence of iconically decaying stimuli at a stimulus onset asynchrony of 100 ms (observer S.S.). (g) The product of the attention-gating function (e) and the input (f) is the input to VSTM (see Figure 10; observer S.S.). (h) The input to VSTM (observer J.S.).

1. The expected starting time for the cued state (t_{cued}) relative to the tonal cue is the cue interpretation time τ . It may depend on strategy and experimental conditions and has a trial-to-trial variance σ_τ^2 .
2. Although the standard deviation of choice RT diminishes by about 10–20 ms with practice (see Table 2), in the present model, variance σ_τ^2 of cue interpretation time is treated as a constant for an observer that does not vary with the particular states involved in the transition, nor with practice, nor between experiments or conditions.
3. The interval v between the cued and postcued attention state is relevant only for Experiment 1; v is assumed to be independent of condition (SOA and cued row).

Equation 5 holds for any attention state; only the particular function $F_{\text{cued}}(r)$ and the particular duration v make it particular to the cued-attention state. Note also that because attention $A(r, t)$ is described by a product of a purely spatial function $F_k(r)$ and a purely temporal function $G(t - t_{\text{cued}}) - G(t - t_{\text{post}})$, it is *space-time separable*. The implications of *separable* are very powerful. By taking logarithms of the space and time functions, the product in Equation 5 is converted into a sum. This illustrates that a monotonic transformation (the logarithm of the variables) produces a linear theory in which the effects of attention to a set of locations in space and attention to an interval in time simply add. The combination of the spatial function and temporal attention functions in Equation 5 suffices for Experiment 1, and we call $A_{\text{cued}}(r, t)$ the *attention-gating function* (see Figure 11c). (The attention-gating function is, below, decomposed into the *true attention window* and the *variability of cue interpretation time*.)

The attention-gating function $A_{\text{cued}}(r, t)$ is determined by the auditory cue to report a particular row and is assumed to be identical for every point in that row. It multiplies the iconic visual input $B(k, r, c, t)$ to determine which items enter VSTM. The output of the attention gate is C (see Figure 11, g and h), the product of stimulus availability $B(k, r, c, t)$ from Equation 1 and the attention-gating function $A_{\text{cued}}(r, t)$, and C is then the input to memory. The three functions, iconically available items $B(k, r, c, t)$, the attention-gating function $A_{\text{cued}}(r, t)$, and their product, are illustrated in Figure 11.

VSTM

VSTM accumulates (mathematically integrates) the attentionally selected information, $C(k, r, c, t) = A_{\text{cued}}(r, t)B(k, r, c, t)$ over time. Whether an item (k, r, c) that has been stored in memory ultimately appears in the response is determined by its memory strength $S(k, r, c)$,

$$S(k, r, c) = \int_{t_k}^{t_{k+1}} C(k, r, c, t) dt. \quad (6)$$

Subsequently, a decision component selects the particular letters to be reported from among the letters in VSTM.

Additive random memory noise. There are two destructive processes in VSTM. First, it is assumed that the strength of a letter i is not perfectly recorded. This strength uncertainty is represented by additive noise, that is, by adding to each letter strength $S(k, r, c)$ an independent Gaussian random variable with a mean of zero

and a variance of σ_s^2 . The unit of VSTM strength is contrast (which is dimensionless) \times milliseconds. Without internal strength–noise, the model would be completely deterministic with respect to which letters are reported from a particular display. Internal additive noise scales the internally recorded strengths. That is, noise determines the amount $\Delta s = s_i - s_j$ by which the strength of letter i must exceed the strength of letter j in order for the probability p_i of report of letter i to exceed the probability p_j of report of letter j by an amount $\Delta p = p_i - p_j$. The standard deviation of internal noise σ_s is an estimated parameter that differs between observers but not between conditions or experiments.

Internally generated distractor letters. The second destructive process in VSTM is quite different from error in the value of encoded strength. It is assumed that distractor letters are present in VSTM and that these compete with traces of relevant stimulus letters for output priority. Distractors are letters from outside the critical arrays or residual traces of letters from previous trials. As only one letter is reported at any spatial location, it is obvious that only the strongest distractor at a location has to be considered. Originally, we assumed that the strength of the strongest distractor was approximated by a random variable: the mean μ_d and variance σ_d^2 . However, preliminary model explorations indicated that σ_d was not a useful parameter for predicting the data, so it was set to zero, and μ_d functions like a simple threshold. Items with strengths below μ_d are not reported. We arbitrarily set μ_d to 100 ms, which constrains the estimates of the other parameters.

Decision Mechanism

The model produces responses by selecting from among letters in VSTM. At each location, the strongest letter is reported. If more reports are required (as in Reeves & Sperling, 1986), the next strongest letter is reported, and so on. The model has no mechanism for location confusions; each spatial location is treated as an independent channel with its parameters and decision process.

Statistical variability in performance. The only sources of variability in the model are noise in VSTM (σ_s^2) and the cue interpretation time (σ_τ^2). There is no variability in precued attention allocation nor in any of the other parameters. Iconic decay is assumed to be the same for all locations in spite of evidence to the contrary (Farrell, Pavel, & Sperling, 1990). Obviously, these are simplifications, but the model already is quite complex. This model predicts all the primary data (probability of reporting a letter at location r, c in array k) in all experiments and conditions with reasonable accuracy (see below). To predict the contingency data (joint occurrences of two letters in Experiment 1), we must add one significant detail to the theory.

A Signal-to-Noise Theory: $\sigma_\tau/\sigma_{\lambda+v}$

The signal-to-noise theory relates the ratio of the standard deviation of the cue interpretation time σ_τ over the standard deviation of the true attention window $\sigma_{\lambda+v} = \sqrt{2\lambda^2 + v^2}/12$ to the correlation between response letters' temporal locations in the stimulus sequence. In the choice attention-gating experiment, this ratio provides a means of answering the question of whether the letters in a report have a greater probability of coming from the same frame than would be expected by chance.

Consider first the effect of variations in cue interpretation time τ . Suppose the width of the true attention window is extremely small (e.g., $\sigma_v = 1$ ms) and the variance of the cue interpretation time is large (e.g., $\sigma_\tau = 100$ ms). Because the extremely narrow attention window can sample only one frame, all the reported letters would necessarily come from the same frame. The particular frame that was sampled on any trial would depend on the value of the cue interpretation time on that particular trial. The correlation between the frames on which adjacent reported letters occurred would be ≈ 1.0 .

Suppose instead that the variance of the cue interpretation time were small (e.g., $\sigma_\tau = 0.01$ ms) and the width of the true attention window were large (e.g., $v = 1,000$ ms). Then any variation in the frame number from which the left, center, and right reported letters were derived would be entirely random, each stimulus letter having an equal chance of being reported from any frame that occurred within the true attention window. The correlation between the frames on which adjacent reported letters occurred would be ≈ 0 .

These simple examples illustrate that by appropriate choice of σ_τ and $\sigma_{\lambda+v}$, the correlation between frames from which pairs of letters are reported can vary from nearly 1.0 (when $\sigma_{\lambda+v} = 0$ and $\sigma_\tau = M$) to nearly 0 (when $\sigma_{\lambda+v} = M$ and $\sigma_\tau = 0$), where M is a large number. A more precise specification of $\sigma_{\lambda+v}$ and σ_τ can account for the correlations observed in the 6×6 contingency tables (see Tables 3 and 4). The heuristics above can be made precise in a formal signal-to-noise theory (see Appendix B). Because of various complications, even the formal algebraic theory is only a rough approximation; estimating parameters requires Monte Carlo simulations (see Appendixes C and D).

Data to Be Accounted For

The data for Experiment 1 consisted of five homogeneous data subsets: 2 observers \times 2 SOAs and, for 1 observer only, two clusters for the 100-ms-SOA condition that represent significantly different levels of practice. Each homogeneous subset of data (a cluster) can be summarized in two ways: (a) primary data, the probability of report at six temporal sampling points in nine spatial locations, yielding 54 data points in each cluster (48 for J.S. with data points involving the center column of the middle row excluded); hence, there were 162 primary data points (i.e., 3 clusters \times 54 data points per cluster) for S.S. and 96 (i.e., 2 \times 48) for J.S.; and (b) secondary data, nine 6×6 contingency tables (three for each row) of joint responses' temporal locations between two spatial locations (i.e., left-center, left-right, center-right) within each row, yielding 972 secondary data points (504 for J.S.; see [a]).

The correlations in the contingency tables for Experiment 1 were nearly all less than .1, and so the correlations, although nonzero, account for less than 1% of the variance of the data. The primary determinants of such contingency table data are the marginal probabilities, and these are already estimated in the primary predictions. To make really good predictions of the contingency table data, it would be necessary to further refine the predictions of the marginals, and this seems an unnecessary exercise. Predicting the correlations directly is far more instructive. Therefore we are concerned with only the correlations. Instead of treating the 36

points in each contingency table as separate data, we summarize each table by just a single number, the correlation, reducing the secondary data set from 972 and 504 to 27 and 14 for the observers S.S. and J.S., respectively. The correlations are treated separately from the other data in the analyses, although the same model parameters are used in all predictions.

The data of Experiment 2 consisted of the probability of a correct response at each of the nine tested locations (3 cued rows \times 3 columns per row). There were 20 conditions: 5 Cue Delays \times 4 Mask Delays yielding 180 data points for each of 2 observers. The average number of observations per point was 67.

The data of Experiment 3 consisted of the probability of a correct response at each of the nine tested locations. There were 16 conditions in Experiment 3: the Presence–Absence of Distractors \times the Presence–Absence of a Tonal Warning Cue \times 4 Mask Delays. This yielded 144 data points for each observer, each data point with an average of 52 trials.

In sum, the total number of primary data points to be accounted for was 486 for S.S., and 420 for J.S. The 972 and 504 secondary data points for observers S.S. and J.S., respectively, in the contingency tables are summarized by 27 and 14 correlation coefficients.

Model Predictions

Primary data. Finding the optimum parameters for the model involved many complexities that are described in detail in Appendix D. The criterion of optimality was maximizing R^2 , the amount of variance in the data that is accounted for by the model, corrected for the number of free parameters (Judd & McClelland, 1989). The model accounts for 91% and 88% of the variance in the 486 and 420 points of primary data for the 2 observers, S.S., J.S., respectively. Table 6 presents, for each observer, the parameter estimates along with the preset and derived parameters. The model-predicted and observed recall probabilities for all these data are presented in Figures 12–14 for the 2 observers.

Secondary data. The 972 and 504 secondary data points in the contingency tables (for joint occurrences of two reports in Experiment 1) are summarized by 27 and 14 correlation coefficients, respectively, for observers S.S. and J.S. Monte Carlo simulations were used to establish confidence intervals around the predicted correlations. The 99% interval was used because of the large number of comparisons (27 for S.S., and 14 for J.S.). For J.S., one of the observed correlations fell outside the model-predicted 99% interval. For S.S., one observed correlation fell outside the model-predicted 99% interval and one outside the 95% interval. Both of the deviant correlations were associated with the 100-1 cluster, which occurred very early in terms of the overall data collection (see Appendix C, Table C1). Obviously, the model's predictions do not differ significantly from the actual secondary data. However, the correlations in the Location \times Location contingency tables, although small, were extremely useful in establishing the signal-to-noise property of the attention-gating function: the ratio of standard deviation of the cue interpretation time and of the true attention window. And this property of the model (and of the data) establishes the parallel (vs. serial) acquisition of information within the window of attention.

The parameters. The model has 6 estimated parameters that deal with the basic processes of detecting and responding to the

Table 6
Preset, Estimated, and Derived Parameters, and Goodness of Fit

Parameter		Unit ^a	Observer S.S.			Observer J.S.		
Preset parameters								
Mean of internal noise	μ_s	ms	0			0		
Mean of distractors	μ_d	ms	100			100		
Estimated parameters								
Cue interpretation time	τ	ms	200 ^b			140 ^b		
Time constant of attention transition	λ	ms	6			5		
Width of true attention window	ν	ms	160			171		
Iconic decay time constant	α	ms	171			150		
Standard deviation of internal noise	σ_s	ms	66			82		
Perceptual sensitivity	γ		3.36			3.15		
Perceptual acuity relative to γ_{22}	[γ_{rc}]		0.27	0.16	0.23	0.34	0.42	0.36
			0.80	1.00	0.40	0.52	1.00	0.53
			0.33	0.36	0.23	0.33	0.38	0.33
Precued attention state	[β_r]		0.47			0.27		
			0.66			0.70		
			0.39			0.48		
Stimulus \times Row visibility adjustments to γ_{rc}								
$r = \text{top, bottom, Exp 1}$	$\omega_{vb}(1)$		1.92			1.03		
$r = \text{middle, Exp 1}$	$\omega_m(1)$		0.50			0.76		
Exp 3	$\omega(3)$		1.32			1.69		
Adjustment to τ	$\Delta\tau$	ms	40 ^c			—		
Goodness of fit	R^2		0.91			0.88 ^d		
Derived parameters								
Standard deviation of τ (based on observed MRT)	σ_τ	ms	53			91		
Standard deviation of AGF ($\Sigma^2 = \sigma_\tau^2 + \sigma_\lambda^2 + \sigma_\nu^2$)	Σ	ms	71			104		
Signal-to-noise ratio	$\sigma_\tau / \sigma_{(\lambda+\nu)}$		1.13			1.82		

Note. Dash in cell means row parameter is not applicable. Exp = experiment; MRT = motor reaction time; AGF = attention-gating function.

^a Blank indicates a dimensionless parameter. ^b For J.S., performance for Experiments 1–3, and for S.S., performance for Experiments 2 and 3 and for Cluster 2 in the 100-ms-stimulus-onset-asynchrony (SOA) condition. ^c For S.S., adjustment for the 150-ms-SOA condition and Cluster 1 in the 100-ms-SOA condition. ^d For J.S., performance for the central location in Experiment 1 is excluded from the predictions because of his decision to ignore that location.

attention cue, with the shift of attention, and with the storage and retrieval of items from short-term memory (see Table 6). There are 15 additional parameters: 8 parameters to describe perceptual acuity in the nine stimulus positions (the ninth is an overall sensitivity parameter and counted among the basic six); 3 parameters to describe aspects of the stimuli in the three experiments; 3 parameters to describe default strategies in Experiments 2 and 3; and for 1 observer, a parameter $\Delta\tau$ to describe a considerable speedup with practice. The basic 6-parameter model is quite simple. However, to deal in detail with the myriad experimental procedures and observer strategies in the three experiments requires corresponding detail in the model's interfaces.

Discussion

Interpreting the Parameters

The Form of the Attention-Gating Function

Apparently different functions for reports of one versus four items. Experiment 1 requested only one item from the temporal stream at any location. (In Experiments 2 and 3 only one item was presented at each location, so there was no incentive to quickly

terminate attention gating.) According to the model, a single item will most often be reported from the center or near center of the attention-gating function where strength is greatest; subsequent items tend to be reported from positions both before and after the central positions. A procedure that requests only a single item at a location gives little information about the tails of the attention-gating function.

Reporting one versus reporting four items. More revealing in learning about the full shape of the attention-gating function is a procedure such as that of Reeves and Sperling (1986), which requests four items in the temporal sequence. They found that the first of the four reported items had the same temporal distribution as the single reported item when only one was requested. And the next three items flesh out the shape of the tails of the attention-gating function. Therefore, the appropriate comparison to the attention-gating function determined here is with the distribution of the first-reported item.

Although Reeves and Sperling (1986) obtained data from 3 observers and three different visual attention cues, they displayed data for only 1 observer and one cue. Their example of data at SOAs of 110 and 145 ms (comparable to our SOAs of 100 and 150 ms) is quite similar to the present data, especially considering that

they used a stream of numerals with a visual attention cue whereas Experiment 1 used a stream of letters and an auditory cue.

There is no obvious scale that can be used to compare widths of attention-gating functions of different shapes. We use the standard deviation (the square root of the variance of the normalized function) even though it gives too much weight to the tails. Reeves and Sperling's (1986) 3 observers have standard deviations (for their four-item attention-gating functions) of 188, 228, and 260 ms compared with the present observers' standard deviations (see Table 6) of 72 and 104 ms for one-item gating functions.

Conclusions. (a) The form of the attention-gating function for procedures that request one report in a temporal stream can be derived from procedures that request four reports but not vice versa. (b) When an attention-gating function is estimated from single-item reports, it is narrower than the gating function estimated from four-item reports. (c) The shape of the single-item gating function depends on the item rate (1/SOA) in the to-be-reported stream, especially at high rates, whereas the gating function determined from four reported items (per stream) is invariant with item rate. (d) The limited capacity of visual short-term memory (VSTM) does not permit simultaneous multiple reports from a location and reports from multiple locations. If these were possible, we would expect to observe the same attention-gating function (at each location) as did Reeves and Sperling (1986).

Cue Interpretation Time

The observers in the present experiments performed two tasks. In the attention task, a tonal cue directed them to report one of three rows; in the reaction time (RT) task, the cue directed them to make one of three finger responses. According to the model, the cue interpretation time is a common element in both tasks—followed in one task by opening an attentional gate and in the other by preparing and executing a motor response. The signal-to-noise contingency analysis (the extent to which letters tended to be reported from the same array) enabled estimates of the variability σ_τ of the cue interpretation time. In preliminary model explorations, for both observers, the estimated values of σ_τ fell within a few milliseconds of the standard deviation of their motor RTs (MRTs). The coincidence of these values was astounding considering how different the data were from which they were derived (Temporal Location \times Location contingencies in attentional reports, MRTs).

We wish to test to what extent cue interpretation time is common to both attention-gating function and the MRT response. Note that every trial contains an MRT and a mean time of occurrence of the reported letters (attention reaction time [ART]). We therefore computed the trial-by-trial correlation between the ART and the MRT. This is the *data correlation*.

If the only source of variation in MRT were cue interpretation time, then one could derive a corresponding ART–MRT correlation from a Monte Carlo simulation in which the cue interpretation time (representing MRT) on a trial is correlated with the mean time of occurrence of reported letters (the ART) on that same trial. In fact, because randomness within the attention window contributes additional variance to the ART, the appropriate measure is ART–MRT covariance, not correlation. We ask, how does the data ART–MRT covariance compare to model ART– τ covariance? If

only cue interpretation time contributes to MRT variance, then we expect data ART–MRT covariance to equal model ART– τ covariance.

The Monte Carlo simulations were used to generate data sets for all the experimental conditions from which model ART– τ covariance was computed. Averaged over all trials and conditions for each observer, the ratio of data-to-model ART–MRT covariance was 0.45 (observer S.S.) and 0.56 (observer J.S.).⁶ This is a large common component and was the basis for using each observer's observed MRT distribution as the cue-interpretation-time distribution in the model.

Iconic Decay, Perceptual Acuity

The time constants of the decay of iconic memory are highly dependent on stimulus conditions; the present values (171 and 150 ms) are typical of what would be expected under good, light-adapted conditions. The perceptual acuity parameters $\gamma(r, c)$ seem to reasonably reflect the falloff of acuity with distance from fixation. One exception is the left column of observer S.S.'s center row, which has a very high acuity value. This reminds us that an attention strategy that remains consistent between conditions and experiments is inherently confounded with acuity estimates.

Evaluation

The parameters of the model have quite reasonable values, that is, approximately what one might a priori expect to find. The agreement of variability in cue interpretation time estimated from correlations between reports of letters in an array and direct measurements of the variability of MRTs indicates great self-consistency.

Overall, the model gives an excellent account of an enormous mass of detailed data. The large amount of data places severe constraints on any model, and the fact that the current model succeeded made it worthwhile to consider carefully how it did so. We now consider how the model relates more generally to other experiments and to a wide range of attentional processes.

Computational Models of Partial Report and Iconic Memory

Partial Report

In partial-report experiments (e.g., Experiment 2), observers are shown a brief display of more items than they can report (because of the limited capacity of short-term memory). At a certain delay, usually after the display has been turned off, they are given a cue to report some randomly chosen part of the display (partial report).

⁶ To find a common processing stage shared by a covert attention-gating response and an overt finger response, Sperling and Reeves (1980) used a different approach. They varied the difficulty of detection of attention cues. The increase in cue interpretation time as cue difficulty increased was 1.4 times greater for attention versus motor responses, suggesting that, even if there were a partially common process, the attention response required more of it.

a

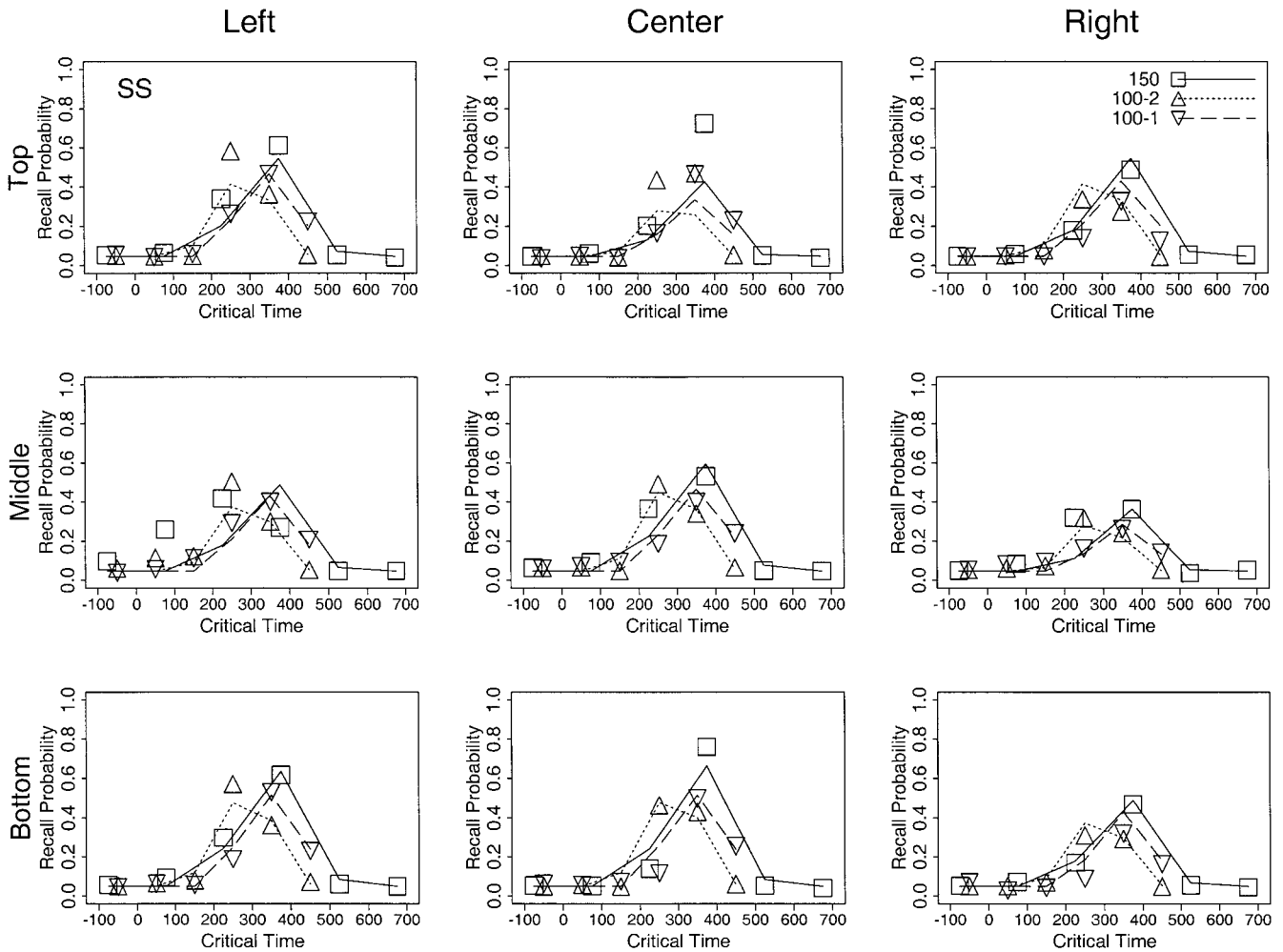


Figure 12. Model predictions for Experiment 1: choice attention gating. Recall probability as a function of the critical time (in milliseconds). Critical time is the time from the onset of the tonal cue to the midpoint of the interval during which the cued row of letters is available. The lines represent model predictions; the symbols represent data. The location of each panel corresponds to the 3 cued rows \times 3 column positions. Curve parameters are indicated at the top right panel. (a) Observer S.S., (b) observer J.S. 100 and 150 represent the SOA between successive arrays; 100-1 and 100-2 (observer S.S.) represent clusters of data from less-practiced and more-practiced sessions, respectively.

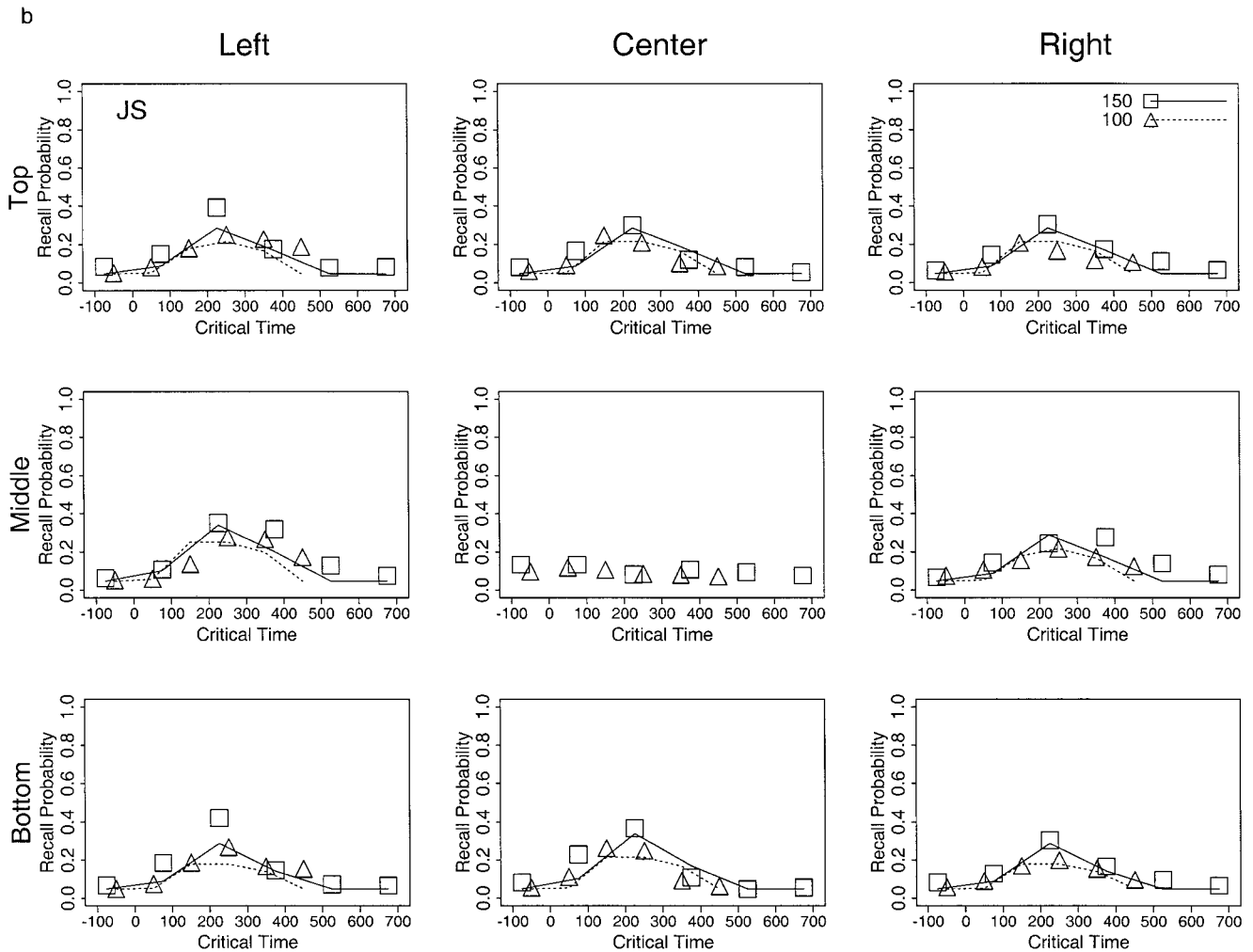
For cue delays of a few tenths of a second or less, partial reports are more accurate than whole reports; this indicates a rapidly decaying sensory memory trace (iconic memory).

In the first mathematical model of partial report, Averbach and Coriell (1961) assumed that prior to the attention cue, information was accumulated nonselectively at every location. After the attention cue, information was accumulated selectively from the cued location(s). If either the nonselective or the selective accumulation were sufficient, the item at the cued location would be reported correctly; otherwise, there would be random guessing. Averbach and Coriell's implicit assumption of two independent stores is extremely implausible, and it was shown to make incorrect predictions in a more comprehensive data set than they originally considered (Gegenfurtner & Sperling, 1993).

Rumelhart (1970)

Rumelhart's computational model of iconic memory assumes that information accumulates at a fixed rate from the visual display. Initially, information is accumulated uniformly from all locations; from the instant the attention cue occurs, it is accumulated only from the cued locations. Accuracy depends monotonically on the amount of accumulated information. Because Rumelhart did not have accuracy data from individual spatial locations, he considered only overall accuracy. His predictions of Sperling's (1960) partial-report data were quite good.

We now know that prior to the attention cue, attention is not distributed uniformly, as Rumelhart (1970) assumed, but is highly focused according to individual biases (Gegenfurtner & Sperling,



1993, and Experiment 2, above). Typically, observers focus on the middle row in arrays of three rows such as Rumelhart considered. This illustrates an important principle: In relatively small, relatively homogeneous data sets (dozens of points), a good fit to data does not by itself validate assumptions; it merely demonstrates that they are consistent. For more detailed analyses of Averbach and Coriell (1961) and Rumelhart (1970), see Gegenfurtner and Sperling (1993).

Bundesen (1990)

Bundesen proposed a general theory of information accumulation, quite similar in spirit to Rumelhart's (1970) but more elaborated. In Bundesen's theory, every item in a visual display is compared to every template in memory to achieve what he called *categorization*, that is, identification. For a continuously visible display, the probability of identification increases as an exponentially limited growth function. Each stimulus location is processed independently, although there is a constraint on the total amount of accumulation. Attention to location or to feature operates multiplicatively, and there is a limited capacity short-term memory.

There is no explicit noise in the model; errors occur when a response is triggered before data accumulation is complete.

To predict results of iconic memory (partial-report) experiments, Bundesen (1990) followed Rumelhart (1970) in assuming that initially information accumulates equally from all locations and that, after the attention cue, accumulation is restricted to the cued locations. There is a rate parameter to describe the exponential decay of visually available information (iconic memory) after termination of the visual display. The theory gives an excellent account of Sperling's (1960) partial-report data, including many fine details.

Comparison of Bundesen's (1990) Model and the Present Model

Bundesen's theory has two components: (a) a formulation of a mechanism of information accumulation and (b) special assumptions that relate this mechanism to specific experiments. The general mechanism is a strength theory. That is, every item is represented by a real-valued number between 0 and 1 (its strength) that represents the probability that it will be reported. In Bundes-

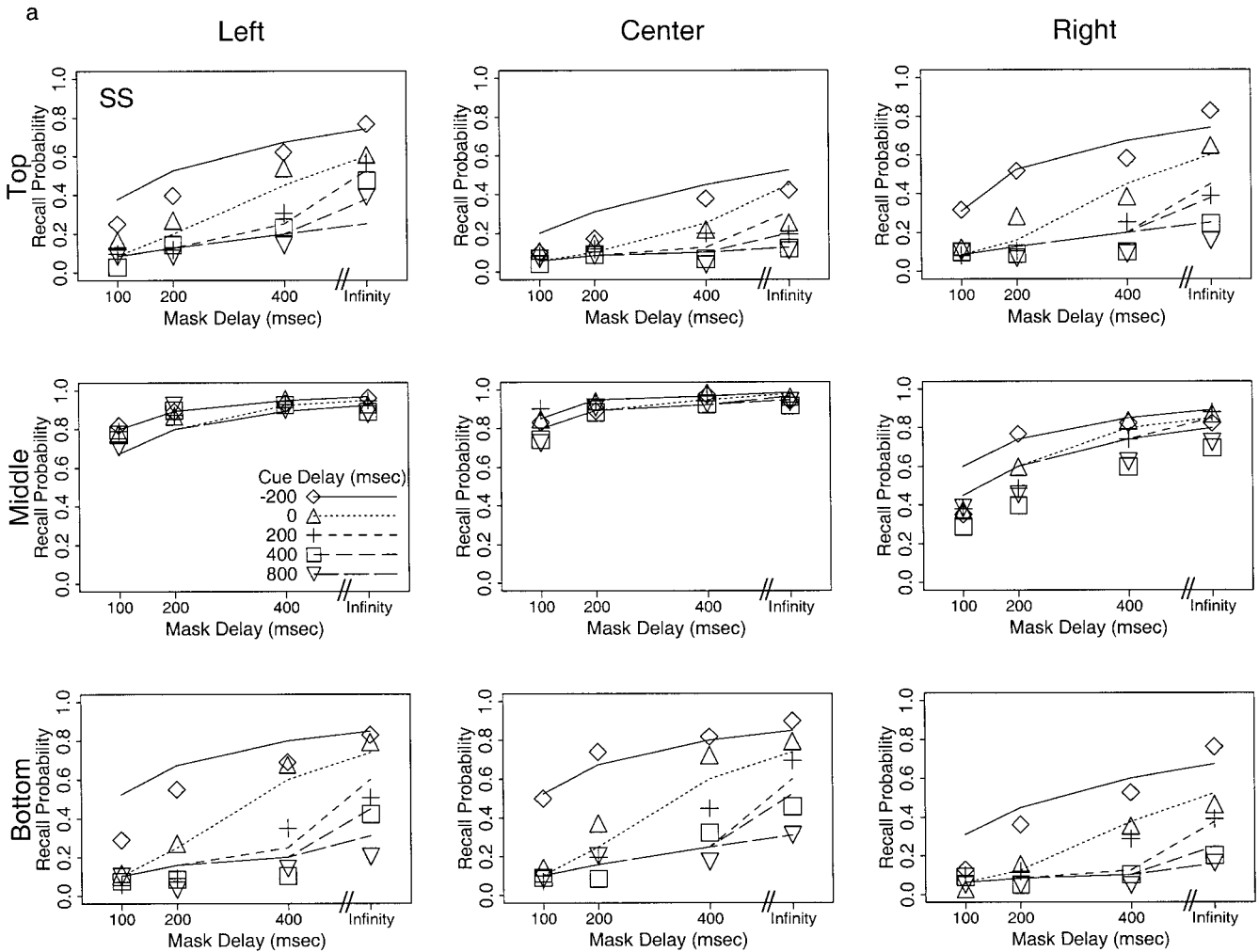


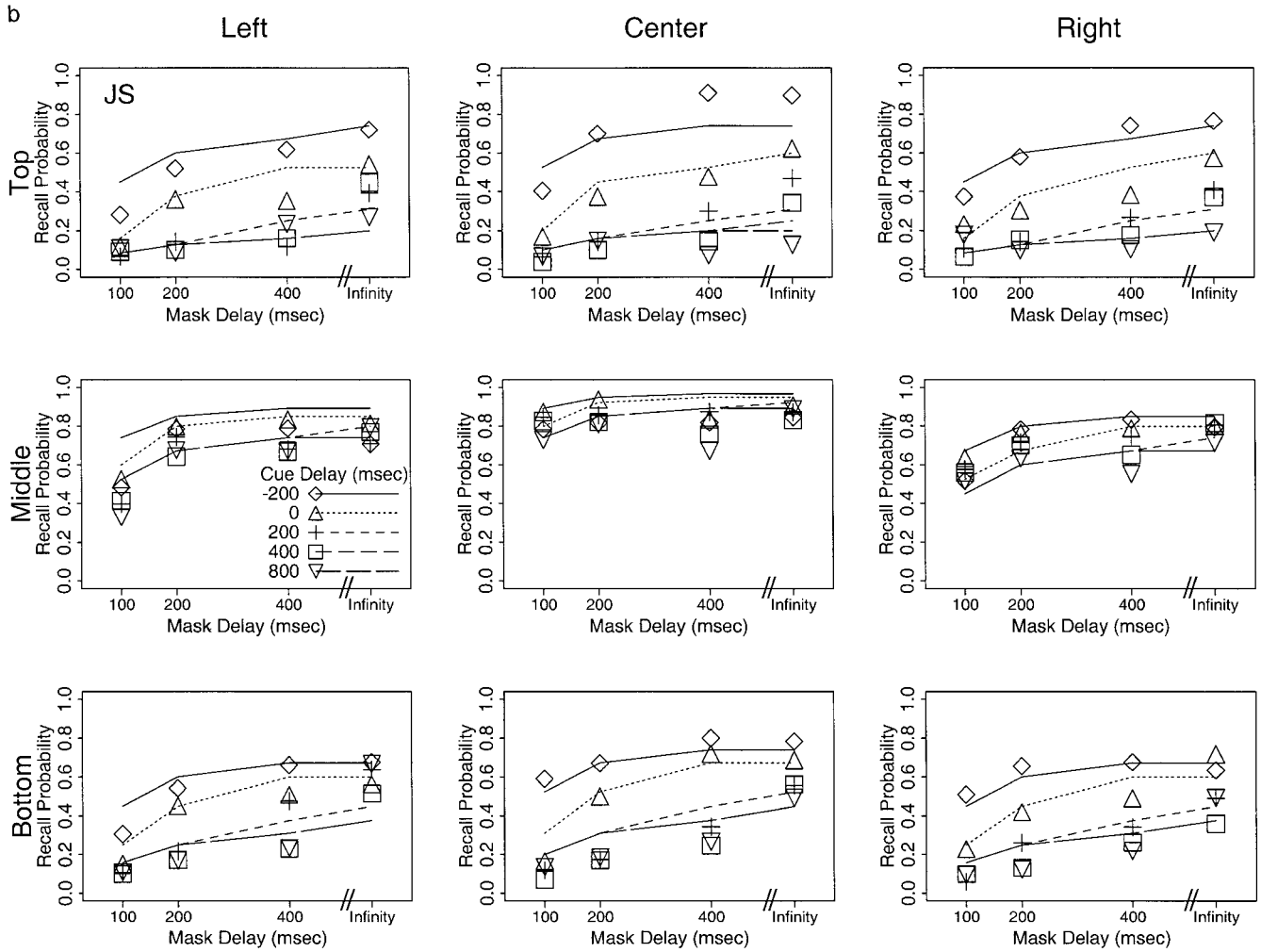
Figure 13. Model predictions for Experiment 2: partial report. Recall probability as a function of mask delay. The lines represent model predictions; the symbols represent data. The location of each panel corresponds to the 3 cued rows \times 3 column positions. The various curves represent different cue delays (see middle left panel). (a) Observer S.S., (b) observer J.S.

en's theory, the strength value directly represents the probability of a report (except when there is competition for space in the limited capacity short-term memory, in which case things get more complicated). For a constant stimulus, strength increases as an exponentially limited growth function.

In the present model, each stimulus item is represented by a strength. The strength of items is perturbed by the addition of random Gaussian noise. Whether or not an item will be reported is determined by its strength relative to other stimulus items and noise items and by the capacity of VSTM. (*Noise items* are traces of previous stimuli that remain in memory; they account for observers' reports of items that were not present in the stimulus.) For a continuously presented item, strength grows linearly with presentation time. The probability of report grows like a cumulative Normal function. Had we assumed exponentially distributed noise, the probability of report would have grown as an exponen-

tially limited growth function, that is, similar to Bundesen's (1990) model. Also, similar to Bundesen's (1990) model, strength accumulates independently at each spatial and temporal location, according to the rate parameters at that location. And, both models represent attention by differentially multiplying the strength values of items in the attended location and unattended locations.

Bundesen (1990) assumed an overall limit on the rate of information accumulation based on the total number of comparisons of stimulus items to memory templates. In the present model, the total *stimulus amplitude* relative to memory *noise amplitude* serves a similar function. Whether there is an overall limit for the entire display (as Bundesen assumed) or whether processing limitations are more local is a matter that requires further study. Bundesen's formulation offers a principled way to characterize the improved performance in Experiment 3 as the similarity of distractor items to stimulus items is decreased. All in all, the basic framework of both



models is quite similar and sufficiently general to accommodate a wide range of experiments. However, the basic frameworks do not account for experiments; specific assumptions are needed to account for experiments, and these, not the framework, discriminate the models.

Some of the assumptions needed to account for iconic memory experiments are quite obvious and were made in conjunction with the defining experiments (Sperling, 1960) prior to any formal computational theory. Namely, there is a limited capacity short-term memory; there is a higher capacity iconic memory (originally called *visual information storage*) with contents that decay quickly after stimulus termination; once an attention cue is received, information from the cued row is selectively transferred from iconic to short-term memory at a specified rate. However, for a computational theory, many more details are needed.

Among the additional assumptions of previous computational theories, some were quite incorrect, others were remarkably prescient; all published theories made good predictions of (limited) data sets. All theories, prior to Gegenfurtner and Sperling's (1993) observations to the contrary, assumed incorrectly that, before an attention cue, information was transferred uniformly from all stim-

ulus locations. On the other hand, all theorists have assumed correctly that information was transferred simultaneously from different locations in a cued row even though, until Experiment 1 (above), there was no evidence for this. Because it yields good predictions, all theorists, including Bundesen (1990) and Gegenfurtner and Sperling (1993) made the assumption that, following a cue, attention transferred immediately and fully to the cued row. This is an unrealistic assumption that would make attention in iconic memory experiments quite different from attention in other contexts.

Iconic memory experiments, by themselves, do not sufficiently constrain a theory of attention; the same theory has to apply to a wider range of experiments. The specific assumptions of the current theory, such as how attention traverses distance (discrete jumps vs. continuous trajectory, the effect of distance moved), the detailed time course of an attention-gating function, how variations in cue interpretation time affect the correlations between adjacent items, the influence of the mixture of observer strategies, the specific nature of errors, and several others, have no counterpart in Bundesen's (1990) theory and are what distinguishes the present effort.

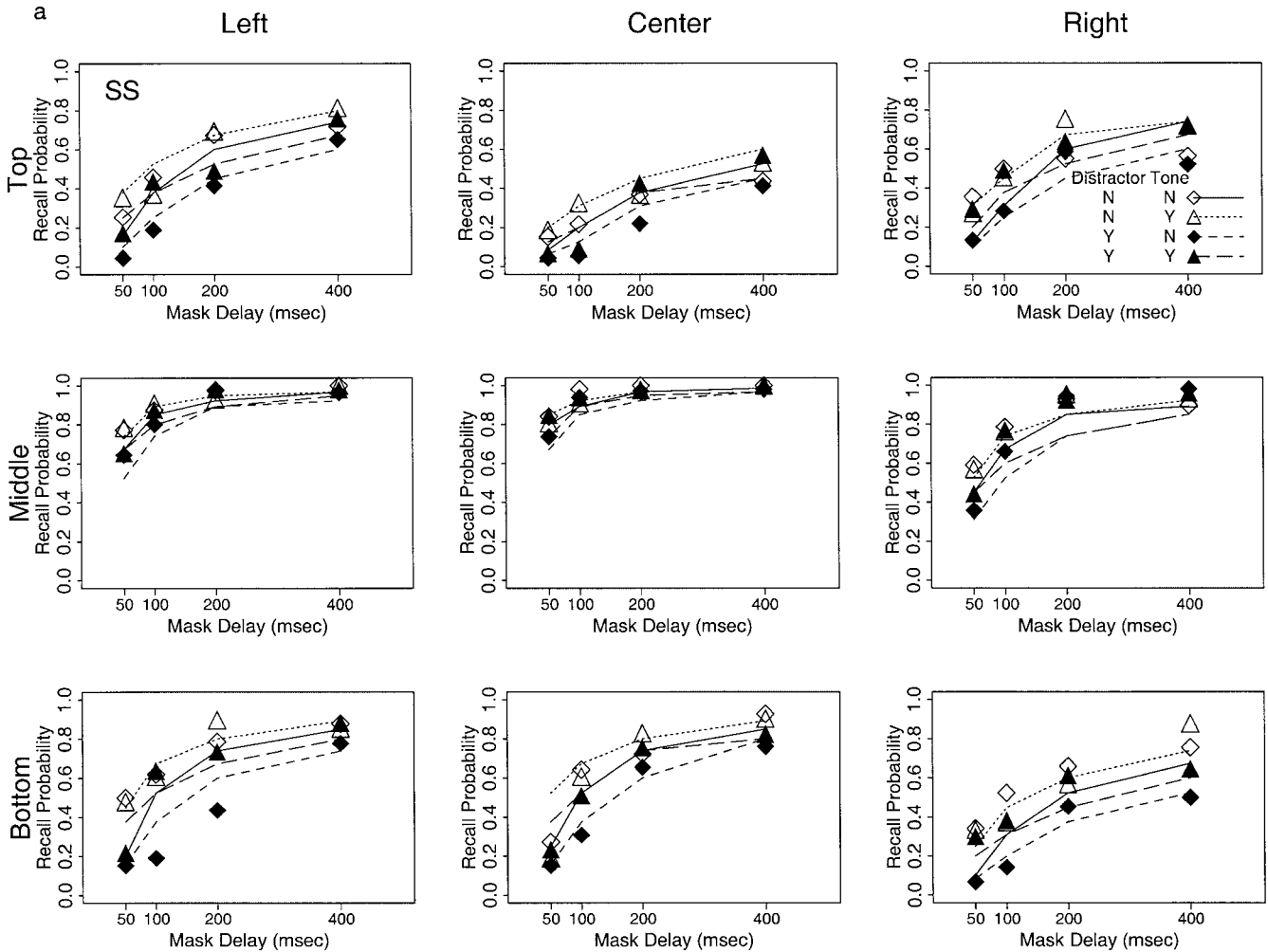


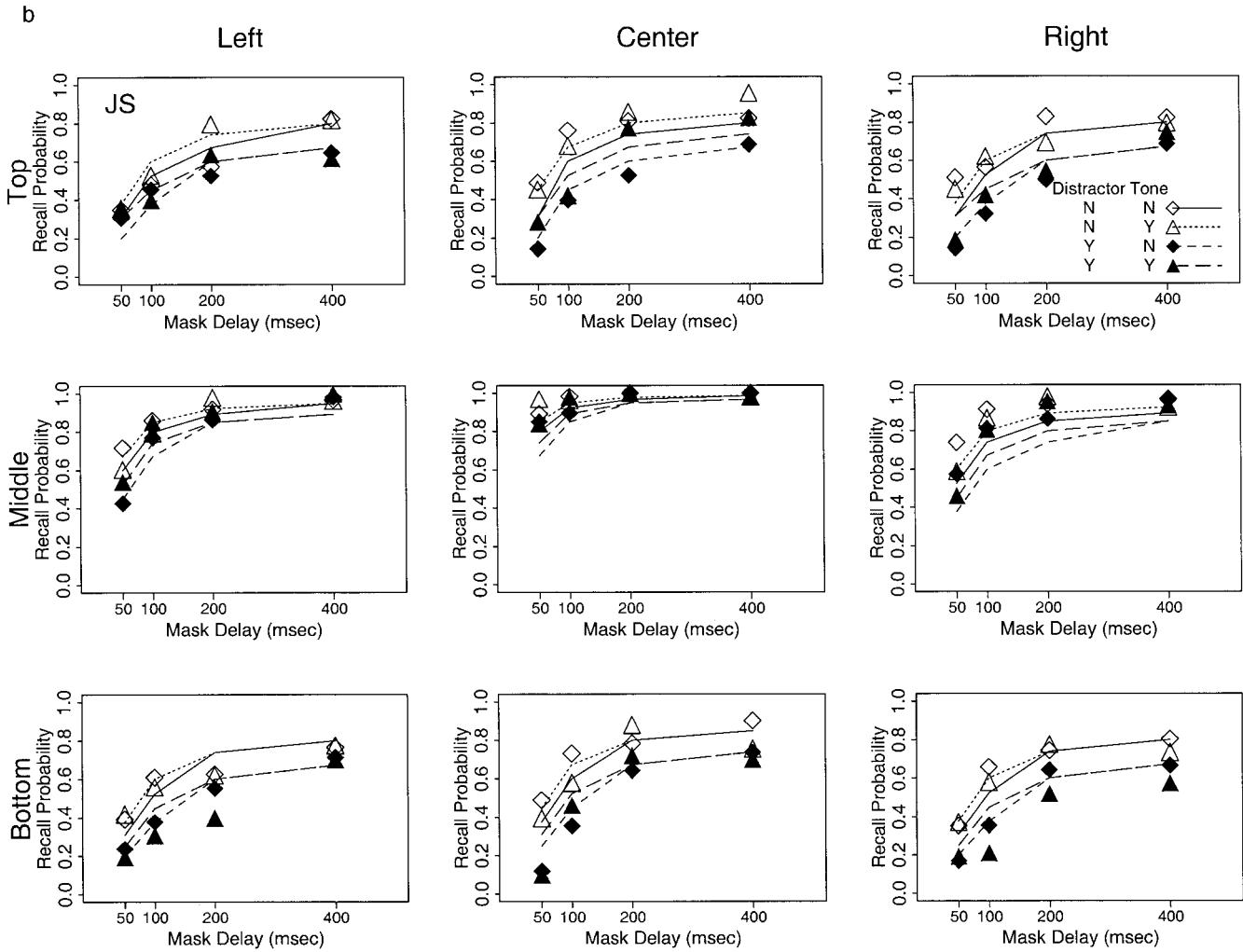
Figure 14. Model predictions for Experiment 3: whole report. Recall probability as a function of mask delay. The lines represent model predictions; the symbols represent data. The location of each panel corresponds to the 3 rows \times 3 column positions. The different curves represent presence (Y) and absence (N) of masks and of tonal cues (see upper right panel). (a) Observer S.S., (b) observer J.S.

Deriving and Proving the Assumptions for a Theory of Visual Attention

The core of the present theory is based on Sperling and Weichselgartner's (1995) theory. Attention is defined by a sequence of successive episodes; each episode is characterized by a function that describes the spatial distribution of attention during that episode. In this respect, attention is somewhat like the course of saccadic eye movements. The difference is that, whereas long eye movements take longer than short eye movements, attention shifts seem to be independent of distance.

To deal with the three types of experiments considered here and with the new data analyses required combing two earlier models and some significant elaborations. Reeves and Sperling (1986) provided the basic attention-gating model. Sperling and Weichselgartner (1995) showed how the attention window was derived from a succession of attention episodes and applied the theory to

three other important paradigms: spatially cued go/no-go RTs (often used in cost-benefit analyses), spatially cued choice RTs, and spatially cued discrimination experiments. To further develop the theory to deal with choice attention gating, partial report with poststimulus masks, and the 16 whole-report conditions required the following enhancements: explicit consideration of performance at each individual spatial location including parameters to describe visual processing rate at each location, a formulation of iconic memory decay, a second kind of memory interference (item-strength threshold), consideration of the distribution and variance of cue interpretation time, explicit measurements of and incorporation into the model of attentional bias prior to an attentional cue, spatial information-processing parameters that depend critically on parameters of the stimulus presentation sequence, and a more detailed parametric specification of the true attentional window. The analysis relating MRTs to properties of the attention window



and to correlations between the array numbers from which response items are drawn is new and quite complex. We now consider all the assumptions of the attention model in relation to the experiments that specifically tested them.

Discrete Jumps of Attention Rather Than Continuous Sweeps

This assumption was tested in the reanalysis of RT data (Shulman et al., 1979) that had originally been taken as the strongest evidence that a peripheral cue causes attention to move over intermediate positions en route to the peripherally cued location. A continuous sweep of attention would predict a space-time correlation for attentional costs and benefits. That is, when an observer is fixating centrally, and a cue to attend a peripheral target is presented, RT would speed up first at nearby locations en route to the cue and only later at the cued location itself. The episodic theory predicts space-time separability of attentional costs and benefits. A consequence of separability is that a graph of RT facilitation as a function of space x and time t should have all ridges and valleys parallel to the x, t axes; a continuous theory

predicts diagonal ridges and valleys. When graphed, the ridges of the Shulman et al. (1979) data are obviously absolutely parallel to the axes and the episodic theory accounts for 99% of the variation in their data (Sperling & Weichselgartner, 1995).

Attention Does Not Pass Over Intermediate Locations

This has been explicitly tested by presenting stimuli at intermediate locations and noting whether these stimuli influence the response, either by being reported or in other ways. Two studies have shown no effect of intermediate stimuli (A. Reeves, 18 July, 2000, personal communication; Sperling & Weichselgartner, 1995).

The Time Taken for Movements of Spatial Attention Is Independent of the Distance Traversed

Initial tests of this assumption gave ambiguous results because the discriminability of central and of peripheral targets was not equated. Subsequently, studies that have equated detectability of peripheral and central cues have found no effect of distance

traversed on pure movements of attention with the eyes held fixed (e.g., Cheal & Lyon, 1989; Sperling & Weichselgartner, 1995).

The Temporal Shape of the Attention Gate

The temporal shape of the attentional window is best revealed by the attention-gating paradigm that involves rapid serial presentation of a single stream of to-be-reported items. An invariant shape of attention gating is derived from experiments that require the report of four items from the to-be-reported stream (Reeves & Sperling, 1986). The shape of the attention gate when only one item is reported depends on the rate of items in the to-be-reported stream. The shape of an attention-gating function observed with one-item reports can be derived from four-item reports but not vice versa. On the other hand, attending to only a single stream of items cannot resolve issues about the distribution of attention in multiple streams (as in Experiment 1).

Parallel Acquisition Within a Spatial Window

The RSVP (choice attention-gating) paradigm of Experiment 1 was developed specifically to study whether items are acquired in parallel or serially within an attention window. If items had been acquired serially in a regular serial scan, it would have been obvious from the pattern of results. If there had been different serial scans, varying from trial to trial (the common criticism of early parallel theories such as that of Sperling, 1967), there would have been large negative correlations between the frame numbers in which adjacent items were reported rather than the positive correlations actually observed (see Table 4). The choice attention-gating paradigm and the correlation analysis offer a strong proof of parallel acquisition that has not been available from other paradigms. It should be noted that the cued response method as exploited by Doshier and McElree (1992; McElree & Doshier, 1989) has been able to establish parallel memory search in the Sternberg memory retrieval paradigm (which previously had been assumed to be serial), and it could potentially be applied to the dynamics of spatial attention.

Cue Interpretation Time

The time needed to interpret and act on an attention cue is a component of every attention experiment and theory. It involves both perceptual and decision components that are not dissected in any of the experiments reported here. Analysis of the distribution of arrays from which letters are reported in Experiment 1 yields a direct measure of the mean cue interpretation time, and analysis of the tendency of letters in the RSVP choice attention-gating procedure to be reported from the same versus different arrays yields a measure of the variability of cue interpretation times.

Mixture or Pure Strategy?

In the earliest iconic memory experiments (Sperling, 1960), some observers changed their strategy with cue delay; this was especially easy because, in a block of trials, only one delay was presented. In short-delay blocks, observers prepared equally for any possible cue; in long-delay blocks, they prepared to report only one row and, in effect, to ignore the cue. However, that was in an

era when each stimulus letter was drafted by hand, stimulus cards were changed manually, and responses were tabulated on long sheets of paper. In the computer age, the number of trials is enormously larger, and observers quickly reach much higher levels of practice.

To determine whether practiced observers can be induced to change strategy, Gegenfurtner and Sperling (1993) used a partial-report procedure in which they presented observers with either a long or very short cue delay on each trial. In separate, pure blocks of trials, the probability of the short delay was one of 0, .1, .5, .9, or 1.0. If there were different attentional strategies, then the optimal mixture would have been different in the different blocks. However, for each observer, performance was statistically and practically identical in all conditions, indicating that each observer used just one strategy in all conditions.

Time Constant of Iconic Memory Decay

Three approaches to the measurement of iconic decay are considered.

1. The most obvious way to measure the time constant of iconic decay is to use a partial-report cue and to observe the decline of report accuracy with cue delay. That confounds the dynamic characteristics of the attention switch with the decay properties of the iconically decaying image.

2. A completely different measurement was introduced by Loftus and colleagues (Loftus, Duncan, & Gehrig, 1992; Loftus & Hogden, 1988). They matched performance with a stimulus that was flashed briefly and allowed to decay for a period before being replaced by a masking stimulus to performance with a stimulus that remained fully on (for a shorter time period) until it was replaced by a masking stimulus. The equivalent duration of the masked to the iconically prolonged stimulus was used to derive an iconic decay function. Iconic decay had a simple exponential form, similar to other measurements.

3. A third measurement of iconic decay (Gegenfurtner & Sperling, 1993; Irwin & Brown, 1987), similar in spirit to Loftus et al.'s (1992) procedures, involves a partial-report procedure in which various cue delays are combined with various poststimulus masking delays.

The Loftus et al. (1992) procedures are unconfounded by shifts of attention. Therefore, the present model assumes the exponential form derived by Loftus. Incorporating iconic decay into the attention model improves the estimates of attention-gating functions in Experiment 1. Incorporating the attentional window from Experiment 1 into the model improves the estimates of iconic decay in Experiment 2.

Multiplication

To theorists, it has seemed so self-evident that attention exerted its effect by multiplying the internal representation of relevant features or locations of the stimulus, that multiplication assumed an axiomatic status (e.g., Bundesen, 1990; Reeves & Sperling, 1986) and was not directly tested. Multiplication also occurs in theories of gain-control mechanisms in sensory systems and in theories of motion perception (e.g., the Reichardt detector;

Reichardt, 1961), domains in which it can be tested directly (e.g., Lu & Sperling, 2001; van Santen & Sperling, 1984).

When attention is introduced into motion experiments, attention is shown to act multiplicatively (like gain control) in determining the effectiveness of attended features. Blaser, Sperling, and Lu (1999) used an ambiguous motion task in which the saturation of the colors in a grating pattern determined the direction of perceived movement. They found that attending to a color (red or green) was equivalent to multiplying that color's saturation by about 1.3, a multiplicative gain-control effect. Attention multiplied the salience of a stimulus and thereby its effect on the motion system; however, there was no discernible change of color appearance (cf. Prinzmetal, Amiri, Allen, & Edwards, 1998).

To account for the effects of attention on the responses of single neurons in macaque cortical areas V2 and V4, Reynolds, Chelazzi, and Desimone (1999) formulated a comprehensive mathematical model in which attention acted like multiplicative gain control. The conclusion is that, in the (so far) limited number of instances in which it has been possible to quantitatively determine the mode of action of attention, it has been multiplicative.

The Short-Term Memory Is Visual

First, we demonstrate that the limited-capacity short-term memory for letters extracted from rapidly presented letter arrays is neither auditory nor vocal and that items do not enter by subvocal rehearsal. The arguments against subvocal rehearsal are that the transfer rate from iconic to short-term memory can be much too high, for example, 100 letters per second⁷; that letters enter short-term memory more or less simultaneously and independently, not serially (see the *Parallel acquisition within a spatial window* section above); and that there are virtually no acoustic-auditory confusions even with acoustically and vocally confusable visual letters (Sperling & Speelman, 1970). Second, we consider evidence for visual representation in memory.

Although the rate of encoding for short-term memory can be described in terms of items per second, theorists have long argued that items are coded in parallel (Bundesen, 1990; Fisher, 1982; Sperling, 1967) and, by implication, the items are visually represented. In a study involving rapid visual streams, Kaufman (Kaufman, 1978; Sperling & Kaufman, 1978) showed that several items from a rapid stream were retained in memory. She further demonstrated that substituting unfamiliar random shapes for the more frequently used alphanumeric characters did not change the number or distribution of items in short-term memory in any significant way. Memory for unfamiliar, nonverbalizable shapes occurring at 10 per second must be visual, not verbal. That memory for shapes and alphanumeric characters seems to be equivalent suggests that short-term memory for alphanumeric characters also is visual. Of course, if one looks beyond memory for items in visual streams, there have been abundant quantitative investigations of visual (vs. verbal, auditory, or iconic) short-term memory dating to Scarborough (1972) and Phillips (1974).

Items Are Reported in Order of Strength

Response items are assumed to be reported in the order of their strength, with no knowledge of their temporal position except as it

is encoded in their strength. When several items are reported from a stream of items, this leads to a "folding" in the order of report. Items from the middle (the peak) of the attentional gate are reported first, followed in alternating order by weaker early and late items. Reeves and Sperling (1986) reported several statistical tests and analyses of the order of pairs of items to demonstrate that, indeed, strength is the only dimension along which observers can order items, although they have the illusion that they are ordering their report in terms of the true temporal sequence.

The Density Function of Noise in VSTM

The noise density function is intimately related to the psychometric function—the increase in the probability of correct responses as a function of stimulus intensity. For the special case in which the internal strength representation is simply proportional to external stimulus strength, and the internal strength is simply compared with a fixed threshold value to determine whether the response will be correct (e.g., Bundesen, 1990), the cumulative distribution of internal noise would determine the psychometric function. Unfortunately, things are seldom so simple. In preliminary explorations of the present model, a Gamma function was assumed for the noise with the order of the Gamma function as a model parameter to be optimized. However, predictions were indiscriminable for various orders when the Gamma functions were normalized to have the same mean and variance. The Normal density function (which is asymptotically equivalent to high-order Gamma functions) was chosen for convenience.

Summary

The assumptions originally were derived in large part from intuition and for physiological plausibility. With the exception of the noise source, which was assumed to have a Normal density function, it was possible to derive selective experimental tests to substantiate and refine all the major assumptions.

Range of Application of the Attention Model

The Episodic Attention Model

The model assumes that attention can be described as a series of discrete episodes E_i ; each is several tenths of a second in duration. For visual attention, the i th episode in the sequence E_i is characterized by the product of two functions: $F_i(x, y)$, which describes the spatial distribution of attention during the episode, and $G(t - t_i) - G(t - t_{i+1})$, which describes the time period between the onset of state i and of state $i + 1$, the period during which $F_i(x, y)$ remains in effect (see Figure 11).

Attention movements are very much like saccadic eye movements in terms of their dynamic properties in that, as soon as the movement is complete, information is acquired simultaneously

⁷ See Gegenfurtner and Sperling (1993) for a review of some of the numerous measurements that have been made of the rate at which items transfer to short-term memory from briefly presented displays. Most transfer-rate measurements depend on a poststimulus masking paradigm (Sperling, 1963).

from all areas under new spatial attention function $F_{i+1}(x, y)$. This similarity is not unexpected; attention movements and saccadic eye movements undoubtedly evolved simultaneously to serve each other. Just as there are other kinds of eye movements than saccades, there may be other kinds of attention movements than the saccadelike movements being considered here. For example, Khurana and Kowler (1987) found that to sustain smooth pursuit eye movements, it is necessary to devote at least some attention to the moving object that is being visually followed.

The introduction reviewed the principal paradigms for investigating the mechanisms of covert attention shifts. We now review how the model applies to these paradigms.

Spatially Cued Go/No-Go RT

This cost-benefit RT paradigm was exploited by Posner and his students (e.g., Posner, Nissen, & Ogden, 1978). An observer must respond as quickly as possible to a target that may appear either to the left or to the right of fixation. The response is the same ("go") independent of the target. Some time before the target (the foreperiod), an attention cue is presented that, most of the time, correctly indicates where the target will occur (thereby yielding an RT benefit) but occasionally indicates the wrong direction (yielding an RT cost).

In terms of the episodic model, the procedure of Posner et al. (1978) involves three consecutive attentional episodes. The initial episode is characterized by an $F_0(x, y)$ that weights all possible target locations equally. After the attention cue has been presented and interpreted, a new attention episode takes effect with an $F_1(x, y)$ that gives greatest weight to the cued location. In some experimental designs (e.g., Shulman et al., 1979), if a cue has not occurred by a certain time t_2 then it becomes highly likely that the current trial is a catch trial and the correct response will be "no-go." This is represented by a third attentional episode, characterized by F_2 , which is similar in shape to F_1 but of smaller amplitude, thereby reducing tendency to respond. Sperling and Weichselgartner (1995) applied the episodic model quantitatively to an elaboration of Posner et al.'s (1978) basic paradigm (Shulman et al., 1979), which contained more extensive data. The model accounted for 99% of the variance of the data.

Spatially Cued Choice RTs

In this paradigm, a cue indicates the location of a target but not the target's identity. Different targets require different responses; the observer is required to respond as quickly and as accurately as possible. For example, Tsai (1983) presented targets either to the left or right of fixation, and the target (in one of the experiments) could be either the letter *D* or *O*. The observer responded by vocally naming the letter. Performance improved as the foreperiod between the attention cue and target presentation increased. According to the episodic theory, this experiment involves just two episodes. (As a target is presented on every trial, there is no need to prepare for a catch trial—the third episode.) In the initial episode, there is approximately equal attention to both possible target locations. The postcue episode is characterized by selective attention to the cued location. The rate of processing is proportional to the attentional gate, $F_i(x, y) \times [G(t - t_i) - G(t - t_{i+1})]$,

times stimulus availability (i.e., whether the stimulus is turned on or not). Increased processing yields quicker RTs. The episodic theory can make quantitatively accurate predictions in this experiment because the theory has great parametric flexibility; more complex data sets than those currently available are needed to provide a challenge that can usefully be tested statistically.

Spatially Cued Discrimination

This paradigm is similar to spatially cued RTs except that the RTs are not measured; what is measured is merely the probability of a correct response. Whereas RT experiments usually are conducted with easily detected stimuli that yield high levels of accuracy, discrimination experiments typically vary stimulus detectability. In the ideal case, this stimulus variation yields a psychometric function that describes the probability of a correct response as a function of the display parameter that is varied to control accuracy.

The application of the episodic theory to spatially cued choice discrimination is similar to spatially cued discrimination with one additional complexity. The shape of the predicted psychometric function is determined by the cumulative distribution of internal noise. If it is known how the display parameter (e.g., contrast or exposure duration) is represented perceptually, then the psychometric function can be used to characterize the internal noise density function.

Lyon (1987, 1990) briefly presented small test patterns followed by noise masks at four possible locations (North, East, South, West) equally distant from fixation. The test patterns themselves could point in one of four directions. Prior to the pattern, an attention cue indicated (with 100% reliability) which location was to be reported. The data are the probabilities of correct reports as a function of the cue-to-target foreperiod and of the exposure duration of the test pattern.

The episodic theory represents Lyon's (1987, 1990) experiments by just two states: a precue episode in which attention is distributed uniformly (at a very low level) to all target locations and a postcue episode in which attention is focused on the cued location. The theory is formally equivalent to the theory used in this article to predict performance in Experiment 2 (partial report with post-stimulus masking). In partial report, the observer accumulates information about letters at the cued location for storage in VSTM. In cued discrimination, the observer accumulates pattern information about the test patch for the orientation response. Figure 11f applies to cued discrimination but with only one stimulus under the attention window instead of a stream. Sperling and Weichselgartner (1995) used an early version of the episodic model to provide a good fit to Lyon's (1987, 1990) data. A cumulative second-order Gamma function was used to fit the psychometric functions, equivalent in this instance to choosing the density function of internal noise as a second-order Gamma function.

Partial-Report (Iconic Memory) Experiments

The application of the episodic model to partial report involves an initial episode of attention prior to the attention cue followed by the postcue episode in which attention is selectively directed to the cued state. This is equivalent to the model for spatially cued choice

RTs (above) except that the attention gate controls the input to VSTM rather than to a pattern discrimination process. Unlike previous computational models (Bundesen, 1990; Rumelhart, 1970), the initial spatial attention function $F_0(x, y)$ is not uniform but represents an initial bias to report the middle row, nor is the shift of attention instantaneous. The attention-gating function follows a time course determined jointly by the true attention window and the variance of the cue interpretation time (as derived from Experiment 1).

Attention Gating, Single To-Be-Reported Stream

When an attention cue appears, the observer must shift attention from the cue location to the to-be-reported stream of items and report the first one (or first four) that he or she can. The distribution in time of reported items defines the attention-gating function. This paradigm, originated by Sperling and Reeves (1980), has been considered above. The cue can be at a distance from the to-be-reported stream, in which case the attention-gating function defines the time required to shift the required distance, or it may be within the stream itself. These two procedures lead to similar estimates of the gating function illustrated in Weichselgartner & Sperling (1987), indicating that attention shift time is independent of the distance traversed.

Choice Attention Gating

This paradigm has been extensively analyzed here because it is uniquely suited for determining the dynamics of attention shifts. The parameters of attention shifts determined in Experiment 1 were used to generate predictions for Experiments 2 and 3.

An important new finding in Experiment 1 is that the cue to attend to a row causes a simultaneous shift of attention to all locations of the to-be-attended row. A priori, it seemed plausible that the habit of reading from left to right would manifest itself in the selection of the left-most letters from earlier arrays than the arrays from which right-most letters are selected. That this did not happen was quite astounding and counterintuitive. In the absence of eye movements, attention moves simultaneously to all parts of a new location. This is a literal verification of a spotlight model in which one light turns off, the next one turns on, and all parts of the field are illuminated concurrently, no matter where or how the spotlight is pointed (e.g., Sperling & Weichselgartner, 1995).

Saliency Theories of Attentional Processes; Levels of Processing

There is considerable recent interest in a saliency map (Ahmad & Omohundro, 1991; Blaser et al., 1999; Burt, 1988; Koch & Ullman, 1985; Lu & Sperling, 1995; Mozer, 1991; Tsotsos et al., 1995) as a critical microprocess in the mechanism of attention. The saliency map is a theoretical brain location where the moment-to-moment importance (saliency) of locations of the visual field is recorded. The output from the saliency map is assumed to control a myriad of subsequent processes, much as the attention gate does in the episodic attention model. Indeed, the episodic attention model can be interpreted as a specification for the operation of such saliency processes.

Saliency theory is neutral with respect to the point at which the attention gate operates, except that it must occur before consecutive events are stored concurrently. On the other hand, VSTM obviously maintains consecutive events concurrently (i.e., it is not “iconic memory”), and therefore—according to the current theory—there is considerable processing between the attention gate and VSTM.

In Blaser et al.’s (1999) attention control theory, based on apparent motion, attention to color operates by means of multiplicative amplification. Low-level processing of color is controlled by a high-level top-down control process, very much as spatial processing is controlled in the present model. Blaser et al. observed that attending to a color (e.g., the red stripes in red–green grating) multiplied its saliency (for the purpose of motion-direction computation) by a factor of, typically, 1.3. Determining whether an item gains access to short-term memory ultimately requires the neural signal corresponding to the to-be-remembered item to be multiplied by 1, whereas the neural signals that represent discarded items are multiplied by 0. Whether an attention gate requires several neural stages or whether the full 0-to-1 control can occur within a single stage is unresolved.

Summary and Conclusions

Summary

Three experiments were conducted and encompassed within a computational theory.

Experiment 1 introduced a choice-attention-gating paradigm in which observers reported a row of letters, indicated by a tonal attention cue, from a stream of 3×3 letter arrays. The primary data, the time of occurrence of the arrays from which each letter was reported, showed that attention moved simultaneously to all letters of a cued row, with no evidence, whatsoever, of left-to-right or other scan patterns. Analysis of the joint occurrences of two responses (e.g., top left and top center) in terms of the arrays from which they were drawn showed small positive correlations ($<.10$).

In Experiment 2, observers viewed brief exposures of 3×3 letter arrays, and a tonal cue directed them to report just one row. This conventional partial-report paradigm measured the persistence of visual information (iconic memory). The combined use of partial report and poststimulus masking (to terminate persistence) provided an additional measure of iconic decay.

Experiment 3 estimated perceptual acuity for the letter arrays in several contexts.

Altogether, the three experiments measured 486 and 420 primary data points for 2 observers, plus 972 and 504 data points in contingency matrices that were summarized in 27 and 14 correlation coefficients, respectively, from tens of thousands of responses.

A formal model used both the primary data and the correlations to derive the shape of a true attention window plus the mean and variance of the cue interpretation time, which together determined the attention-gating function. Because the dynamics of attention switching had been derived from Experiment 1, and measures of perceptual acuity from Experiment 3, it was possible to derive a pure measure of the time constant α of exponential iconic decay, uncontaminated by attention switching for Experiment 2 ($\alpha = 171$ and 150 ms for the 2 observers).

The model estimated six parameters to deal with cue detection and interpretation, attentional selection, and recall. Fifteen additional parameters were needed to deal with perceptual acuity, stimulus differences between experiments, observers' optional strategies, and a practice effect. The model accounted for 91% and 88% of the variance of the primary data for the 2 observers, respectively.

The differences in processing of stimuli at different but nearby locations, and the variations of these differences with changes in presentation conditions and paradigm, were so large, especially in Experiments 2 and 3, that location differences must be taken into account. A theory that dealt only with aggregate data, and treated all positions equally, might not be correct for any position looked at individually.

Conclusions

About 0.15 s after an auditory cue to attend to a visual area, an attention window opens for 0.2 s (possibly longer in partial-report procedures) to simultaneously admit information into visual short-term memory from all the to-be-attended locations. The duration and variability of the cue interpretation time and the time course of the attention window are all well-defined, and all are independent of the distance of the attention shift to the new locations.

References

- Ahmad, S., & Omohundro, S. (1991). *Efficient visual search: A connectionist solution* (Tech. Rep. No. tr-91-040). Berkeley: University of California, Berkeley, International Computer Science Institute.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal*, 40, 309–328.
- Blaser, E., Sperling, G., & Lu, Z.-L. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences, USA*, 96, 11681–11686.
- Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs, NJ: Prentice Hall.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523–547.
- Burt, P. (1988). Attention mechanisms for vision in a dynamic world. In *Ninth International Conference on Pattern Recognition* (pp. 977–987). Los Angeles: IEEE Computer Society.
- Chastain, G. (1991). Time-course of location changes of visual attention. *Bulletin of the Psychonomic Society*, 29, 425–428.
- Cheal, M., & Lyon, D. R. (1989). Attention effects on form discrimination at different eccentricities. *Quarterly Journal of Experimental Psychology*, 41A, 719–746.
- Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *Quarterly Journal of Experimental Psychology*, 43A, 859–880.
- Cheal, M., & Lyon, D. R. (1994). Allocation of attention in texture segregation, visual search, and location-precuing paradigms. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 47A, 49–70.
- Cheal, M., Lyon, D. R., & Gottlob, L. R. (1994). A framework for understanding the allocation of attention in location-precued discrimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 47A, 699–739.
- Colegate, R., Hoffman, J. E., & Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics*, 14, 217–224.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27, 183–228.
- Dosher, B. A., & McElree, B. (1992). Memory search: Retrieval processes in short term and long term recognition. In L. R. Squire (Ed.), *Encyclopedia of learning and memory* (pp. 398–406). New York: MacMillan.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 188–202.
- Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology*, 12, 75–96.
- Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 778–793.
- Egly, R., & Homa, D. (1991). Reallocation of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 142–159.
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, 12, 201–204.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, 12, 155–160.
- Eriksen, C. W., & Hoffman, J. E. (1974). Selective attention: Noise suppression or signal enhancement. *Bulletin of the Psychonomic Society*, 4, 587–589.
- Eriksen, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics*, 42, 299–305.
- Eriksen, C. W., & Webb, J. M. (1989). Shifting of attentional focus within and about a visual display. *Perception & Psychophysics*, 18, 175–183.
- Eriksen, C. W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 583–597.
- Farrell, J. E., Pavel, M., & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. *Vision Research*, 30, 921–936.
- Fisher, D. L. (1982). Limited-channel models of automatic detection: Capacity and scanning in visual search. *Psychological Review*, 89, 662–692.
- Gegenfurtner, K., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 845–866.
- Gray, H. J., & Isaacs, A. (1975). *A new dictionary of physics*. London: Longman.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Grindley, G. C., & Townsend, V. (1968). Voluntary attention in peripheral vision and its effects on acuity and differential thresholds. *Quarterly Journal of Experimental Psychology*, 20, 11–19.
- Hall, P., & Gegenfurtner, K. (1988). *Run-time library, the HIPL picture processing software* [Software manual]. New York: New York University, Department of Psychology, Human Information Processing Laboratory.
- Hoffman, J. E. (1975). Hierarchical stages in the processing of visual information. *Perception & Psychophysics*, 18, 348–354.
- Hughes, H. C., & Zimba, L. D. (1985). Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 409–430.
- Hughes, H. C., & Zimba, L. D. (1987). Natural boundaries for the spatial spread of directed visual attention. *Neuropsychologia*, 25, 5–18.
- Irwin, D. E., & Brown, J. S. (1987). Tests of a model of informational persistence. *Canadian Journal of Psychology*, 41, 317–338.
- Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology*, 34, 103–112.

- Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society*, 21, 247–250.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.
- Kaufman, J. (1978). *Visual repetition detection*. Unpublished doctoral dissertation, New York University, New York.
- Khurana, B., & Kowler, E. (1987). Shared attentional control of smooth eye movement and perception. *Vision Research*, 27, 1603–1618.
- Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. S. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 259–276). Hillsdale, NJ: Erlbaum.
- Klein, R. M. (1994). Perceptual-motor expectancies interact with covert visual orienting under conditions of endogenous but not exogenous control. *Canadian Journal of Experimental Psychology*, 48, 167–181.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- LaBerge, D., & Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional effect. *Perception & Psychophysics*, 40, 188–200.
- LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, 96, 101–124.
- Lappin, J. S., & Uttal, W. R. (1976). Does prior knowledge facilitate the detection of visual targets in random noise? *Perception & Psychophysics*, 20, 367–374.
- Loftus, G. R., Busey, T. A., & Senders, J. W. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics*, 54, 535–554.
- Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 530–549.
- Loftus, G. R., & Hogden, J. (1988). Picture perception: Information extraction and phenomenological persistence. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 139–191). New York: Academic Press.
- Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin*, 88, 785–820.
- Lu, Z.-L., & Sperling, G. (1995). Attention-generated motion perception. *Nature*, 377, 237–239.
- Lu, Z.-L., & Sperling, G. (2001). Sensitive calibration and measurement procedures based on the amplification principle in motion perception. *Vision Research*, 40, 2355–2374.
- Lyon, D. L. (1987). *How quickly can attention affect form perception?* (Tech. Rep. No. AFHRL-TR-87-28). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.
- Lyon, D. R. (1990). Large and rapid improvement in form discrimination accuracy following a location precue. *Acta Psychologica*, 73, 69–82.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118, 346–373.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Mozer, M. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press.
- Muller, H. J. (1994). Qualitative differences in response bias from spatial cueing. *Canadian Journal of Experimental Psychology*, 48, 251–260.
- Muller, H. J., & Findlay, J. M. (1987). Sensitivity and criterion effects in the spatial cuing of visual attention. *Perception & Psychophysics*, 42, 383–399.
- Murphy, T. D., & Eriksen, C. W. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Psychological Research*, 42, 576–586.
- Musseler, J. (1994). Position-dependent and position-independent attention shifts: Evidence against the spotlight and premotor assumption of visual focussing. *Psychological Research*, 56, 251–260.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44–64.
- Pashler, H. (1991). Shifting visual attention and selective motor responses: Distinct attentional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1023–1040.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16, 283–290.
- Podgory, P., & Shepard, R. N. (1983). Distribution of visual attention over space. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 380–393.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance* (Vol. 10, pp. 531–556). Hillsdale, NJ: Erlbaum.
- Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. J. Pick & I. J. Saltzman (Eds.), *Modes of perception* (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.
- Possamai, C., & Bonnel, A. (1991). Early modulation of visual input: Constant versus varied cuing. *Bulletin of the Psychonomic Society*, 29, 323–326.
- Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 261–282.
- Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, 93, 180–206.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303–317). Cambridge, MA: MIT Press; and New York: Wiley.
- Remington, R. W., & Pierce, L. (1984). Moving attention: Evidence of time-invariant shifts of visual selective attention. *Perception & Psychophysics*, 35, 393–399.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19, 1736–1753.
- Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25, 31–40.
- Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1995). Space and selective attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance* (Vol. 15, pp. 231–265). Cambridge, MA: MIT Press.
- Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 7, 191–216.
- Sagi, D., & Julesz, B. (1985). Fast noninertial shifts of attention. *Spatial Vision*, 2, 141–149.
- SAS Institute. (1985). *SAS user's guide: Statistics, Version 5 Edition* [Software manual]. Cary, NC: Author.
- Scarborough, D. L. (1972). Memory for brief visual displays of symbols. *Cognitive Psychology*, 3, 408–429.
- Shaw, M. L. (1978). A capacity allocation model for reaction time. *Journal*

- of *Experimental Psychology: Human Perception and Performance*, 4, 586–598.
- Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, 14, 353–409.
- Shaw, M. L., & Shaw, P. (1977). Optimal allocation of cognitive resources to spatial locations. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 201–211.
- Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Perception and motivation* (pp. 739–781). New York: Wiley.
- Shih, S. (1996). Evidence for quantal attention shifts. *Investigative Ophthalmology and Visual Science*, 37(4), S532.
- Shih, S., & Sperling, G. (1994). Using cluster analysis to discover and characterize covert strategies. *Psychological Science*, 5(3), 175–178.
- Shih, S., & Sperling, G. (1995). A model of selective attention in early visual processing. *Investigative Ophthalmology and Visual Science*, 36(4), S857.
- Shih, S., & Sperling, G. (1996). The time course of covert visual attention shift. *International Journal of Psychology*, 31(3/4), 109.
- 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.
- Shulman, G. L., Wilson, J., & Sheehy, J. B. (1985). Spatial determinants of the distribution of attention. *Perception & Psychophysics*, 37, 59–65.
- 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. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica*, 27, 285–292.
- 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., & Doshier, B. A. (1986). Strategy and optimization in human information processing. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (Vol. 1, pp. 2-1 to 2-65). New York: Wiley.
- Sperling, G., & Kaufman, J. (1978, August). *Three kinds of visual short-term memory*. Talk presented at the Eighth International Symposium on Attention and Performance, Princeton, NJ.
- Sperling, G., & Melchner, M. J. (1978). Visual search, visual attention, and the attention operating characteristic. In J. Requin (Ed.), *Attention and performance* (Vol. 7, pp. 675–686). Hillsdale, NJ: Erlbaum.
- 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* (Vol. 8, pp. 347–360). New York: Academic Press.
- Sperling, G., & Shih, S. (1997). Measuring and modeling selective attention in early visual processing. *Abstracts of the Psychonomic Society*, 2, 18.
- Sperling, G., & Shih, S. (1998). A mathematical theory of iconic memory and attention. *Journal of Mathematical Psychology*, 42, 507–508.
- Sperling, G., & Spelmann, R. G. (1970). Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of human memory* (pp. 149–202). New York: Academic Press.
- Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, 102, 503–532.
- Tsal, Y. (1983). Movement of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194–214.
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78, 507–545.
- Van der Heijden, A. H., Schreuder, R., & Wolters, G. (1985). Enhancing single-item recognition accuracy by cueing spatial locations in vision. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 37A, 427–434.
- van Santen, J. P. H., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A: Optics and Image Science*, 1, 451–473.
- Weichselgartner, E., & Sperling, G. (1987, November 6). Dynamics of automatic and controlled visual attention. *Science*, 238, 778–780.
- Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, 8, 521–527.
- Wright, R. D., & Ward, L. M. (1994). Shifts of visual attention: An historical and methodological overview. *Canadian Journal of Experimental Psychology*, 48, 151–166.
- Yantis, S. (1988). On analog movements of visual attention. *Perception & Psychophysics*, 43, 203–206.

Appendix A

Analyses of Row and Column Confusions

On each trial in Experiment 1, each response letter (corresponding to a particular cued row R and response column C) is matched to every stimulus letter (row r , column c) in the set of six critical arrays, indexed by t , to yield $p_{R,C}(r, c, t)$, the probability of a match. Because each stimulus letter is chosen randomly and independently, the probability of a match by chance is $1/21$ (the number of different stimulus letters). A standard correction for chance guessing was applied to $p_{R,C}(r, c, t)$ to estimate the true probability of a match, $p_{R,C}(r, c, t)$. $p_{R,C}(r, c, t)$ is summed over the six critical arrays to yield $p_{R,C}(r, c)$, the estimated probability that a response R, C truly matches a stimulus r, c within the critical arrays. We analyzed the $9 \text{ stimulus locations} \times 9 \text{ responses} \times 2 \text{ observers} \times 2 \text{ SOAs} = 324$ guessing-corrected probabilities plus an additional cluster of less-well-practiced responses for observer S.S.

Table A1 shows that observers indeed reported letters of the cued row in their correct column position. The average probability of a report in the matching row and column is .83 and .73 for the 2 observers, respectively; the average probability of a row confusion is .023. The estimated true probability of most row or column confusions is very close to zero. Note that an unbiased estimate of zero probability is equally often positive and negative, and this is nearly the case in Table A1.

To test the statistical significance of row and column confusions, a chi-square test was performed for each cell of Table A1. The chi-square value was based on the difference of observed and expected frequency of matches of each response R, C to each stimulus r, c, t ; these chi-square

values were added over the six critical arrays, yielding six degrees of freedom (see Table A1).

Diagonal confusions. For both observers, all the recall probabilities for each cued row and column ($2 \text{ observers} \times 2 \text{ SOAs} \times 3 \text{ cued rows} \times 3 \text{ columns} = 36 \text{ cells}$) were very significantly above the chance level ($p \ll .00001$). Of the 288 other cells in Table A1, 72 represent row confusions, 72 column confusions, and 144 both row and column confusions. Of the 208 row and column confusion cells, two have chi-square values with $p < .01$; this is what would be expected by chance. We conclude that there are no significant diagonal confusions.

Row confusions. Of the 72 cells that represent row confusions in Table A1, there is 1 with $p < .01$ for observer S.S. and there are 3 for J.S. The 3 cells represent more row confusions than would be expected by chance ($p < .001$). Although this is clearly statistically significant, it represents a small fraction of cells and the actual values are not large (average for observer J.S. = .062), which—for these 3 worst cells—is less than 10% confusions).

Column confusions. There are 6 cells of column confusions with $p < .01$ in Table A1, but 3 of these cells have fewer rather than more than the expected number of column confusions. On the other hand, there are 11 cells with $p < .001$; these have statistically significantly many more confusions than could have occurred by chance. Seven of the 11 cells occur when the middle row is cued. By far, the largest number of column confusions occurred when observer J.S. thought he was reporting the middle letter of the middle row.

(Appendixes continue)

Table A1
 Experiment 1: Guessing-Corrected Probability (Summed Over Six Critical Arrays) That the Letter in a Stimulus Row and Column Matches a Response Letter

Cued row (100-ms SOA) [150-ms SOA] ^a	Response column	Stimulus row	Stimulus column, 100-ms SOA			Stimulus column, 150-ms SOA		
			Left	Center	Right	Left	Center	Right
Observer S.S.								
Top (1,326) [1,146]	Left	Top	.899****	.024	.008	.923****	.013	.012
		Middle	.024	.022	-.015	.055	.021	.012
		Bottom	.002	-.004	-.005	.019	-.002	-.010
	Center	Top	.017	.828****	-.006*	.004	.885****	.017
		Middle	.024	.037*	.014	-.017	.030	-.007
		Bottom	.008	.001	-.040	.001	-.006	-.005
	Right	Top	.004	.033	.549****	.034	.028	.626****
		Middle	.020	.018	.041	-.022	-.012	.027
		Bottom	.045	.025	.005	.047*	-.022	.007
Middle (1,322) [1,128]	Left	Top	.038	-.013	.004	.008	.0155	-.015
		Middle	.900****	.011	-.009	.893****	.047**	-.002
		Bottom	.000	-.005	.007	-.001	-.020	-.011
	Center	Top	.007	.000	.001	.025	.010	.010
		Middle	.035*	.818****	.023*	.048	.905****	.048**
		Bottom	-.028	-.019	-.001	-.011	.002	.011
	Right	Top	.003	.019	.015	.019	-.013	.044
		Middle	.027	.080**	.535****	.045*	.098**	.648****
		Bottom	.003	.001	.026	-.008	.002	.023
Bottom (1,328) [1,126]	Left	Top	.006	.013	-.000	-.011	.007	.039
		Middle	.023	.006	-.006	.050	-.003	.008
		Bottom	.939****	-.006	-.000	.933****	.045	.019
	Center	Top	-.017	-.008	-.008	.001	.028	-.032
		Middle	.027	.026	.017	-.001	-.004	.014
		Bottom	.012	.851****	-.007	.026	.852****	.024
	Right	Top	-.003	-.032	.035	.018	.014	.017
		Middle	.035	-.000	.001	.002	.007	-.004
		Bottom	.014	.026	.539****	.012	.056**	.608****
Observer J.S.								
Top (844) [680]	Left	Top	.731****	-.013*	-.001	.719****	-.024	-.016
		Middle	-.016	-.031	-.031	-.000	-.017	.003
		Bottom	.020	-.028	-.033	-.005	.003	.001
	Center	Top	.067**	.537****	.009	.098**	.545****	.017
		Middle	.070	.072*	-.010	-.002	.018	.012
		Bottom	-.019	.001	-.033	-.016	-.013	-.013
	Right	Top	.010	.012	.476****	-.010	.012	.614****
		Middle	.007	.016	.041*	.049	.029	.037
		Bottom	.019	.005	.028	.010	.029	.027
Middle (859) [659]	Left	Top	.041	-.031	.007	-.007	.012	-.013
		Middle	.706****	.003	-.024	.796****	.022	-.020
		Bottom	-.007	.015	-.008	-.020	-.013	.009
	Center	Top	.025*	-.004	-.010	-.026	.025	.001
		Middle	.193**	.281****	.018	.221**	.353****	.129**
		Bottom	.004	.023	.015	-.010	.001	.009
	Right	Top	.022	.041	-.031	-.008	-.016	.001
		Middle	-.035*	.025	.591****	.016	.033	.697****
		Bottom	.024	-.015	.039	.035	-.028	.009
Bottom (846) [695]	Left	Top	.014	-.006	.023	-.025	-.011	-.013
		Middle	.023	.014	.022	.007	.011	.017
		Bottom	.643****	.000	-.031	.706****	.025	.013
	Center	Top	.022	.005	-.032	-.027	-.022	.014
		Middle	-.001	.028	.012	-.005	.011	-.002
		Bottom	.057	.574****	-.027	.070	.635****	-.016
	Right	Top	.010	-.052	.005	.011	-.001	.017
		Middle	-.002	.008	.075*	.017	.014	-.002
		Bottom	.004	.005	.506****	.031	.037**	.590****

Note. Values in bold italic represent the stimulus positions the observer attempted to report. Boldface indicates the requested stimulus positions. Chi-square tests were used to determine significance of difference from zero. SOA = stimulus onset asynchrony.
^a Column values in parentheses are the number of trials for the 100-ms-SOA condition. Column values in brackets are the number of trials for the 150-ms-SOA condition.
 * $p < .01$. ** $p < .001$. **** $p \ll .000001$.

Appendix B

A Signal-to-Noise Theory for Predicting Temporal Contingencies Between Spatial Locations

The theory assumes that, for each response letter (left, middle, right), there was an instant t in time at which the response letter was selected from the stimulus stream. If t was greater than the onset time of frame k and less than the onset of frame $k + 1$, then stimulus letter k was stored in VSTM for subsequent report. Let x and y represent two different response locations chosen from among (left, center, right). Let X and Y be the random variables representing the real-valued times at which the responses in location x and location y are selected from the stimulus stream. Let j represent a trial number. The sampling times on trial j , x_j and y_j , are derived from the distribution of the cue interpretation times and the true attention window as follows.

Let S be a random variable that represents the distribution of cue interpretation times, and let s_j be the sample value on trial i . Let N be a random variable that has a density function equivalent to the true attention window, and let n_j be a random sample on trial i . Then the letter sampling times for locations x and y are

$$x_i = s_i + n'_i, \text{ and} \tag{B1a}$$

$$y_i = s_i + n''_i \tag{B1b}$$

where n' and n'' are independent samples. Equation B1 states that the cue interpretation time s_j affects both locations equally, but each is sampled independently within the true attention window (n'_i, n''_i). To simplify the subsequent development (without any loss of generality), we choose the time axis so that S and N have mean zero. Let σ_S^2 , σ_N^2 , σ_X^2 , and σ_Y^2 , respectively, be the variance of S , N , X , and Y . Note that $\sigma_S^2 = \sigma_N^2$ and

$\sigma_N^2 = \sigma_{\lambda+v}^2$. Because S and N are independent, $\sigma_X^2 = \sigma_Y^2 = \sigma_S^2 + \sigma_N^2$. The Pearson product-moment correlation between X and Y is

$$\begin{aligned} r &= \frac{\sum_{i=1,N} x_i y_i}{N \sigma_x \sigma_y} = \frac{\sum_{i=1,N} (s_i + n'_i)(s_i + n''_i)}{N(\sigma_S^2 + \sigma_N^2)} \\ &= \frac{\sigma_S^2}{\sigma_S^2 + \sigma_N^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_{\lambda+v}^2}. \end{aligned} \tag{B2}$$

Hence,

$$\frac{r}{1-r} = \frac{\sigma_\tau^2}{\sigma_{\lambda+v}^2} = \frac{\sigma_\tau^2}{2\lambda^2 + v^2/12}. \tag{B3}$$

Equation B3 illustrates that the standard deviation of cue interpretation time σ_τ acts like signal and that of the true attention window $\sigma_{\lambda+v}$ acts like noise in classical signal-detection theory. According to Equation B2, these two factors determine the correlations in the contingency tables. In the full model, matters are more complex because memory noise σ_s reduces the values of the correlations to much less than would be expected by the pure signal-to-noise theory. A second consequence of the signal-to-noise theory is that all the correlations would be expected to be the same (between left and center, between center and right, and between left and right). In fact, for the observers, the observed correlations are approximately but not exactly the same. There is a tendency for the left-center correlation to be slightly higher than the other two (see Table 4). These issues are considered in Appendix C.

Appendix C

Temporal Contingencies Between Spatial Locations: Monte Carlo Procedures and Results

The R^2 (percent of variance accounted for) derived for the model (Appendix D) only measures the goodness of the model's performance in simultaneously predicting recall probabilities of all three experiments. The concern here is how well the model predicts the correlations between the arrays of the stimulus stream from which pairs of response letters in the choice attention-gating experiment are chosen, that is, the contingencies described in Tables 3 and 4 of the text.

For a particular observer and SOA, consider all the responses made in two columns, i, j , selected from (left, center, right). Table 3 illustrates a joint distribution of stimulus frames from which items in response columns were chosen; Table 4 shows the correlations derived from 41 such joint distributions. This appendix shows how predictions of the correlations were derived from the model (by means of Monte Carlo simulations) and presents the predictions, together with the data, in Table C1.

The first step is to find optimum parameters for the model; we consider here only what was relevant for these particular predictions. The model for fitting the primary data is relatively indifferent to the ratio in which the cue interpretation time and the true attention window contribute to the overall variance of the attention-gating function, so we start with these parameters (see Appendix D). The other starting point is a value of signal-to-noise ratio $\sigma_\tau/\sigma_{\lambda+v}$ derived from Equation B3. Because the to-be-predicted correlations were obtained only from Experiment 1, we further fine-tune the model's parameters that were relevant for Experiment 1 on the primary data of Experiment 1 (for this correlation computation only). These slightly

modified parameters and the initial $\sigma_\tau/\sigma_{\lambda+v}$ were used in subsequent Monte Carlo simulations to optimize $\sigma_\tau/\sigma_{\lambda+v}$.

Once optimum parameters had been computed, for each observer, 100 runs of Monte Carlo simulations of Experiment 1 trials were performed for each cluster and each cued row. For each run, the number of Monte Carlo trials was set at the number performed in the actual experiment (see Table C1). At the end of each run of Monte Carlo simulation, three 6×6 contingency tables (left-center, left-right, center-right) were obtained for the particular cued row in the particular cluster. For each predicted 6×6 contingency table, we computed a Pearson correlation indicating the correlation within the predicted table. Hence, for each 6×6 contingency table in the data (see Table 4), we obtained a mean and standard deviation of the model-predicted correlation based on the 100 Monte Carlo runs.

Each observed correlation was then compared with the model-predicted 95% and 99% confidence intervals. The results are summarized in Table C1. Because the number of comparisons was large (14 for J.S. and 27 for S.S.), it is more appropriate to use the results of the 99% confidence interval to assess the model's performance. For J.S., one of the observed correlations fell outside the model-predicted 99% interval. For S.S., one observed correlation fell outside the model-predicted 99% interval, which was associated with the 100-1 cluster, which occurred very early in terms of the overall data collection. Thirty-eight of the 41 model predictions do not differ significantly from the data.

Table C1
*Means and Standard Deviations of Predicted Correlation and Means of Observed Correlation
 (Between Frames From Which Letters Are Reported) as a Function of Contingent Spatial
 Locations, Row, and Cluster*

SOA (cluster)	Row	N	Contingency	Predicted r^a		Observed mean r
				M	SD	
Observer S.S.						
100 (1)	Top	304	Left-center	.1056	.0477	.0827
			Center-right	.0799	.0522	.2074 ^b
			Left-right	.0840	.0564	.0384
100 (1)	Middle	285	Left-center	.0555	.0596	.1350
			Center-right	.0394	.0621	.2253 ^c
			Left-right	.0350	.0592	.0962
100 (1)	Bottom	287	Left-center	.1177	.0467	.0764
			Center-right	.0802	.0597	.0577
			Left-right	.0830	.0517	.0876
100 (2)	Top	1,022	Left-center	.1320	.0250	.0891
			Center-right	.1030	.0286	.0704
			Left-right	.0940	.0273	.0204
100 (2)	Middle	1,037	Left-center	.0725	.0332	.0816
			Center-right	.0628	.0326	.0544
			Left-right	.0573	.0297	.0448
100 (2)	Bottom	1,041	Left-center	.1517	.0247	.1081
			Center-right	.1034	.0290	.0527
			Left-right	.0951	.0282	.0818
Weighted mean						
		7,376		.0833	.0314	.0729
Observer J.S.						
100	Top	844	Left-center	.0883	.0336	.1018
			Center-right	.0764	.0326	.0398
			Left-right	.0984	.0340	.0389
100	Middle	859	Left-right	.0760	.0326	.0952
100	Bottom	846	Left-center	.1233	.0306	.0421 ^c
			Center-right	.0926	.0390	.0870
			Left-right	.0979	.0360	.0283
150	Top	680	Left-center	.0667	.0375	.0252
			Center-right	.0387	.0426	.1120
			Left-right	.0676	.0407	.0511
150	Middle	659	Left-right	.0571	.0428	.1025
150	Bottom	695	Left-center	.0925	.0391	.1326
			Center-right	.0641	.0374	.0106
			Left-right	.0547	.0394	.0192
Weighted mean						
		4,583		.0798	.0367	.0631

Note. SOA = stimulus onset asynchrony.

^a The model's statistics for each 6×6 contingency are based on 100 Monte Carlo simulations, each of N trials. ^b The observed correlation is outside the estimated 95% confidence interval. ^c The observed correlation is outside the estimated 99% confidence interval.

Appendix D

Estimating the Model's Parameters

Optimization of Model Parameters

Minimization of the sum of the squared deviations. For a given set of parameters, the estimates of their values that best account for the data of recall probability were obtained with the optimization program PRAXIS (Brent, 1973). The criterion for optimization was the minimization of the sum of the squared differences between the model's predicted probabilities of letter reports and the observed probabilities. This optimization expression is appropriate for the multidimensional regression and the R^2 that is used below.

Goodness of Fit: R^2 Criterion. The quality of the fit was indicated by the statistic R^2 (Judd & McClelland, 1989), which is the amount of variance in the data that is accounted for by the model corrected for the number of free parameters. Let p_i and q_i , respectively, be the observer's and model's performances (i.e., recall probability); let p be the mean of the observer's performances; and let N and K be the number of data points and free parameters, respectively. The statistic R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1,N} \frac{(p_i - q_i)^2}{N - K}}{\sum_{i=1,N} \frac{(p_i - p)^2}{N - 1}} \quad (D1)$$

This statistic represents the goodness of fit adjusted for the number of free parameters. The value of R^2 is between 0 and 1, with 1 denoting a perfect fit.

Obtaining Parameter Estimates

No analytic solution. The characteristics of the present study and model did not permit an analytic solution for the model parameters because of numerous complexities. For example, the data of Experiment 1 do not

give direct access to a real-valued time at which the stimulus is sampled, only to a reported letter (which spans an interval of 100 or 150 ms). Second, there might be more than one match between response and stimulus letters because letters were sampled with replacement. Third, the theory uses additive ("vertical") noise in memory and temporal uncertainty ("horizontal noise") in the variation in cue interpretation time; these are awkward to combine analytically. Because of such complications, estimating parameters requires a Monte Carlo simulation.

Optimization strategy. Using Monte Carlo simulation to estimate a large set of parameters is not practical. Our optimization strategy consists of several distinct steps.

1. Assume for the moment that the convolution of the distribution of cue interpretation times and the true attention window to generate an attention-gating function is valid. This permits ordinary hill-climbing optimization (computation of a Jacobian matrix). This procedure is used to generate optimal parameters for all the first-order data. The "optimal" analytic parameters yield an overall attention-gating function but do not yield the components (σ_τ , λ , v).

2. Use a Monte Carlo simulation and make a systematic grid search for optimum values of critical parameters to tune the parameters for the first-order data of Experiment 1.

3. Use Monte Carlo simulation and a grid search among the three parameters of the attention window to find optimum values for the components of the attention window (σ_τ , λ , v) for fitting the secondary data (Pearson correlations of Experiment 1).

4. Using these (σ_τ , λ , v) values, use Monte Carlo simulation and a grid search to optimize critical parameters for the primary data of all three experiments. Iterate steps 2, 3, and 4.

Received September 19, 1999

Revision received August 29, 2000

Accepted June 23, 2001 ■

Wanted: Your Old Issues!

As APA continues its efforts to digitize journal issues for the PsycARTICLES database, we are finding that older issues are increasingly unavailable in our inventory. We are turning to our long-time subscribers for assistance. If you would like to donate any back issues toward this effort (preceding 1982), please get in touch with us at journals@apa.org and specify the journal titles, volumes, and issue numbers that you would like us to take off your hands.