Psychological Review
1995, Vol. 102, No. 3, 503–532

0033-295X/95/$3.00

Episodic Theory of the Dynamics of Spatial Attention

George Sperling
University of California, Irvine

Erich Weichselgartner
Fraunhofer Society

Previous measurements of visual attention in simple reaction time, choice reaction time and complex discrimination experiments in which attention was purported to move continuously across space are reanalyzed. These data and data from attention gating experiments are quantitatively predicted by an episodic (quantal) theory of spatial attention that proposes instead: (1) visual attention can be resolved into a sequence of discrete attentional acts (episodes), (2) each attentional episode i is defined by its spatial attention function $f_i(x, y)$, (3) the smooth transition at time t_0 between episodes is described by a temporal transition function $G(t - t_0)$, and (4) f and G are space-time separable. In new experiments, which use a concurrent motor reaction time task to control for nonattentional factors, the duration of an attention transition is shown to be independent of the distance traversed and of the presence of interposed visual obstacles.

1. The Episodic Theory of Attentional Facilitation and Three Applications

Much research has buttressed the old view (Helmholtz, 1909; Wundt, 1896/1924) that attention may be distributed differentially over the visual field without the need for an eye movement (e.g., Beck & Ambler, 1973; Cohn & Lasley, 1974; Hoffman & Nelson, 1981; Klein, 1980; LaBerge & Brown, 1989; Posner, Snyder, & Davidson, 1980; Sperling, 1960; Sperling & Melchner, 1978; Wolford & Morrison, 1980). Recent research has focused on the dynamics of attention shifts from one location to another in the visual field (e.g., Eriksen & Murphy, 1987; Jonides, 1980; Posner, 1980; Posner, Nissen, & Ogden, 1978; Remington, 1980; Remington & Pierce, 1984; Shulman, Remington, & McLean, 1979; Sperling & Reeves, 1980; Tsal, 1983).

There are two contending types of theories of visual attention shifts: analog and quantal. The analog theory is exemplified by Shulman et al. (1979) and by Tsal (1983), who have suggested that attention shifts spatially in an analog manner in which all intervening points between the starting and finishing locations are traversed. This analog-motion formulation derives from the work of Shepard and his collaborators (Shepard, 1975; Shepard & Cooper, 1982) on analog processes in mental rotation.

We represent the analog theory by a moving spotlight model. The focus of attention is represented by the area illuminated by the beam of a spotlight. In the illuminated area, information is acquired fast and accurately. The area outside of the beam is illuminated only by dim background illumination; it represents the unattended locations where information is acquired slowly, if at all. In moving from one focus to another, mass, inertia, and friction constrain the spotlight's velocity, so it takes longer to move further.

There are two analog variants of the spotlight analogy which differ according to whether or not the spotlight remains on while it is moving. In the strong variant (Figure 1), the spotlight remains on continuously so that, in movement from an initial to a final position, all intermediate locations are illuminated. In terms of an attention shift from an initial location to a final location, the strong spotlight analogy implies that information (fragments or whole units) from intermediate locations will automatically be acquired. In the weak form of the spotlight theory, the beam is turned off during movement.

According to a quantum theory of spatial attention, attention jumps have a fixed, quantized jumping time independent of the distance jumped. Intermediate points between the initial and final points are not attended. In terms of luminance analogies, the quantal theory would be modeled by a system of spotlights that remain fixed during the period of a trial. In a model of a shift of attention, the initial position of attention is represented by one spotlight that turns off while another spotlight, representing the final position, is being turned on. The switch over from one light to the other need not be instantaneous. However, any illumination that happens to fall on locations between the initial and final positions is due to broadly focused beams or to background illumination rather than to a sweep of a light beam across the intermediate space.

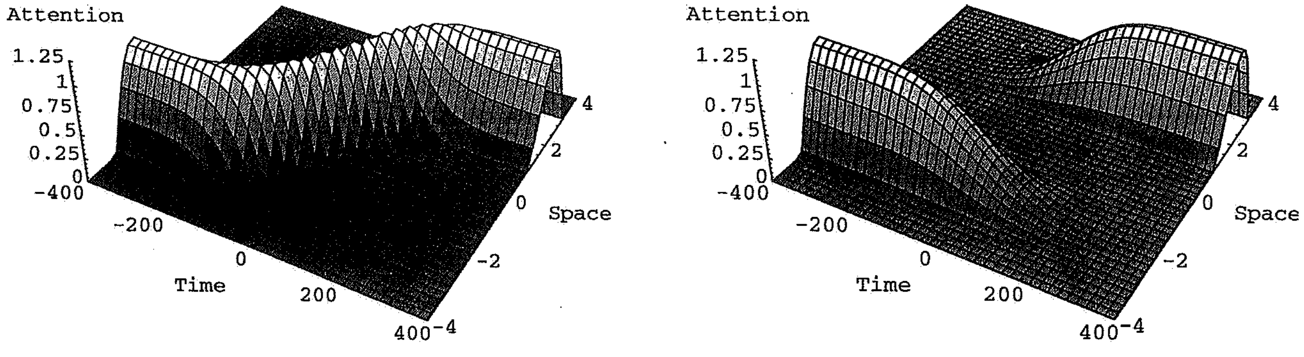
Reeves and Sperling (1986) provided a formal gating model for the process by which attention turns on and off at a particular location, but they did not deal with the time it takes for attention to traverse a given distance. Remington and Pierce (1984) concluded from their data that the time it takes attention to move does not vary with distance, but they did not formulate a formal model. The threefold purpose of this article is to formulate the competing theories more precisely, to reevaluate the

---

George Sperling, Department of Cognitive Sciences and Institute for Mathematical Behavioral Sciences, University of California, Irvine; Erich Weichselgartner, Fraunhofer Society, Munich, Germany.

The experimental work was supported by Air Force Office of Scientific Research, Life Science Directorate, Visual Information Processing Program; the theoretical work and preparation of the article were supported by the Office of Naval Research, Cognitive and Neural Sciences Division. We thank Adam Reeves for helpful discussions during the planning of the experiments.

Correspondence concerning this article should be addressed to George Sperling, Department of Cognitive Science, University of California, Irvine, California 92717. Electronic mail may be sent via Internet to sperling@uci.edu.



Analog

Quantal

Figure 1. Analog and quantal (episodic) versions of the moving spotlight model of attention. The amount of illumination (attention) varies as a function of space ((x, y)) and time t in milliseconds; a two-dimensional slice in the x, t plane is shown. Height above the plane indicates the amount of illumination received by a point x at the instant t from a spotlight (the amount of attention). In the analog movement, all intermediate points are illuminated (attended) as the spotlight is shifted from -2 to $+2$. In the quantal (episodic) shift of attention, the spotlight pointed at location -2 is extinguished and, simultaneously, the spotlight at location $+2$ is turned on. Because extinction and onset take a measurable amount of time, there is a brief period when the spotlights partially illuminate both locations simultaneously.

evidence that has been proposed previously to bear on these theories, and to present new data that can discriminate between the theories.

The quantal theory is reformulated here as an episodic theory to better capture the notion that it is not merely a theory of visual attention shifts but a more general theory of the attentional control of mental processing. After outlining the episodic theory, we first review representative experiments from the two of the most important paradigms that have been used to reach the analog conclusions. The difficulties exposed in these paradigms, and our alternative interpretations of the results, are important in motivating the more complex attention reaction time (ART) procedure that we use to study attentional shifts.

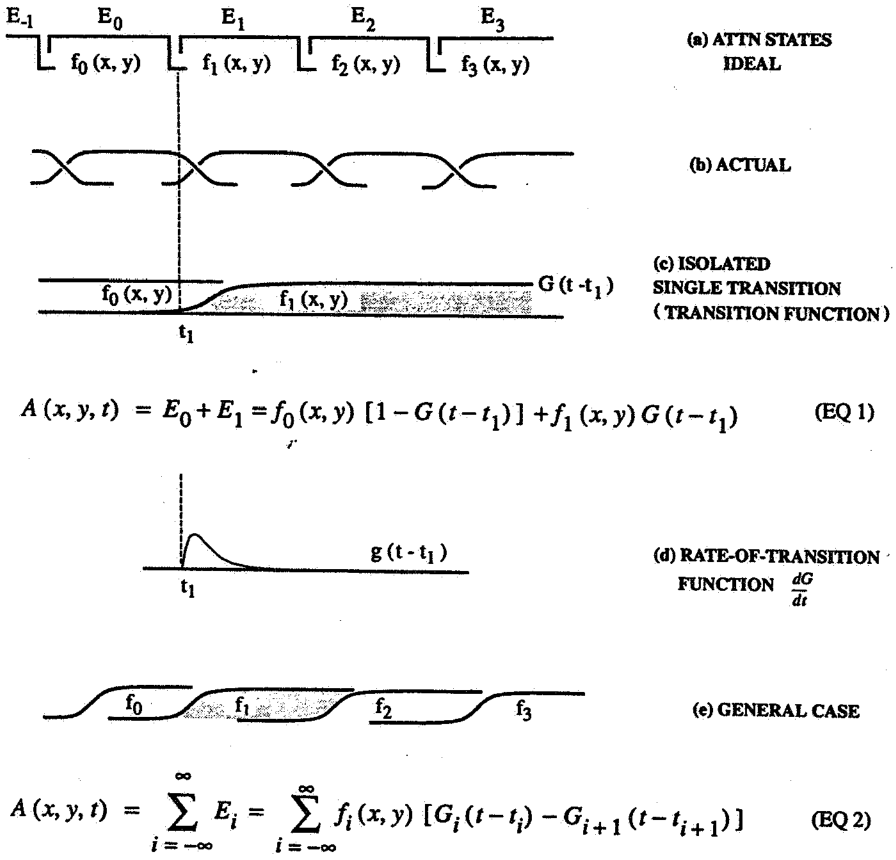
Episodic Theory of Attentional Facilitation

Definition: Attentional state. In the context of experiments that deal with the allocation of attention to locations in visual space, a state of attention is the modifiable component of an observer's location-dependent readiness to respond to stimuli. In the episodic theory, a state of attention is embodied in an episode E that is a function of spatial location x and time t. In general, x is a two-dimensional or even a three-dimensional position vector. As a practical matter, however, in nearly all of the experiments discussed here, attentional variations have been considered in only one (usually the horizontal) dimension. Attention is assumed to affect accuracy or reaction time (RT), or whatever measure is taken as an indicator of performance. All other factors that affect the response measure are lumped together as a residual component. Thus, attention emerges as a theoretical construct in data analysis as that component of performance that varies with "attentional" instructions and similar manipulations.

A sequence of episodes: The spotlight analogy elaborated. The basic idea of the episodic theory is encapsulated by analogy to a number of fixed spotlights that illuminate a theater stage. Information is available from each area of the stage in proportion to the amount of illumination it receives. Each spotlight points at a given location and cannot easily be adjusted. The amount of light that a spotlight i produces at a point x, y on the stage is $f_i(x, y)$. The spotlights may differ in the size, shape, and location of the area they cover and in the total amount of light they produce. The critical constraints of the episodic theory are that, at any one moment, only one spotlight may be turned on and that, while it is turned on, it cannot be moved.

At certain instants in time t_i, the stage manager switches power from one spotlight to another. (For the brief sequences considered here, it is convenient to use the same index i to designate an episode E_i, the episode's spatial distribution function f_i, and the episode's starting time t_i.) The time course of turning on illumination to a spotlight i is $G(t - t_i)$. If the spotlight were turned on instantly, $G(t - t_i)$ would be a unit step function. The power is switched instantly, but, because the spotlights are very hot, the illumination fades out over a period of a few tenths of a second at the turned-off location and, because it takes a few tenths of a second to heat a spotlight, the illumination turns on at the new location with a time course that is perfectly complementary to the time course of turning off. The transition function from one spotlight to the next $G(t - t_i)$ is modeled as a gamma function (Figure 2c) that rises monotonically and continuously from 0 to 1.

An episode E_i is defined as a continuous period during which a particular spotlight produces illumination. It is characterized by the area of the stage that is illuminated and by the function of time that describes the spotlight's activity. An attentional theory is analogous to a script that describes exactly where each spotlight is focused and when it is on during a performance. In the next section, we develop a simple mathematical expression that describes the amount of light falling at each point x, y on



$$A(x, y, t) = E_0 + E_1 = f_0(x, y)\,[1 - G(t - t_1)] + f_1(x, y)\,G(t - t_1) \qquad \text{(EQ 1)}$$

$$A(x, y, t) = \sum_{i=-\infty}^{\infty} E_i = \sum_{i=-\infty}^{\infty} f_i(x, y)\,[G_i(t - t_i) - G_{i+1}(t - t_{i+1})] \qquad \text{(EQ 2)}$$

Figure 2. The episodic theory of attention. Panel a: A succession of ideal attention episodes. Each episode E_i is characterized by a spatial attention function $f_i(x, y)$ and the instants of its onset and termination. Panel b: Actual attention episodes. These do not begin and end instantaneously. Panel c: An isolated single transition at time t_1 from an initial episode E_0 with spatial attention function $f_0(x, y)$ to a second episode, E_1 with $f_1(x, y)$. The temporal function characterizing the transition is $G(t - t_1)$. Equation 1 gives the instantaneous spatial distribution of attention $A(x, y, t)$ as a mixture of f_0 and f_1, where $G(t - t_1)$ controls the mixture proportions. Panel d: The density function $g(t - t_1)$ representing the rate of attention shifting corresponding to the transition function $G(t - t_1)$ in Panel b. Panel e: The general case, the sum of all of the individual transitions, illustrated graphically and described by Equation 2. In the general case, the transition function $G_i(t - t_i)$ may be different for different transitions i. See Appendix B (Glossary) for symbol definitions.

the stage as a function of time t. This is the mathematical formulation of the episodic theory.

The spotlight analogy also incorporates the effects of practice. During rehearsals of a stage production, the positions and timings of the spotlights are adjusted to better accommodate the action. This is quite analogous to the early trials in a psychological experiment during which observers are presumed to optimize their attentional strategy for the experimental task.

An analog (continuous) theory of attention. In the strong form of the analog theory, attention is represented by a single spotlight that is always on. A lighting assistant moves the spotlight to follow the action on the stage and occasionally makes adjustments to its degree of focus. The process in the analog theory is quite different from that of the episodic theory. Nevertheless, a sequence of closely spaced stationary spotlights turned on in rapid succession can approximate a continuous movement to any desired degree of accuracy provided that the number of spotlights is large enough and that the time between successive episodes is small enough. This is a stepwise approximation to a continuous function. From a practical point of view, it is worthwhile to draw a distinction between an episodic and an analog theory only when the spatial steps in the episodic theory represent a large fraction of the total distance to be traversed. This is the case in all of the examples considered here.

Attention and eye movements. The distinction between episodic and continuous theories of attention is profoundly analogous to the distinction between saccadic and smooth-pursuit eye movements. Indeed, spatial attention normally is mediated by saccadic eye movements. Thus, it is probable that the control mechanisms of spatial attention coevolved with the control mechanisms of eye movements. For example, saccadic eye movements do not usually occur faster than about three or four per second. This corresponds well with the timings of the attentional episodes considered here (during which, of course, the eyes remained stationary). In contrast to saccadic eye movements, humans cannot make smooth-pursuit eye movements in a static visual field. Smooth pursuit requires a fixation target that moves not too quickly and with an apparently smooth, predictable trajectory. Khurana and Kowler (1987) showed that spatial attention in smooth pursuit is inextricably linked to the target that drives the smooth pursuit. It might well be that, even when the eyes are stationary, an observer can attentionally track a predictable, slowly moving target (e.g., Cavanagh, 1992). The episodic theory of attention being put forward here simply asserts that, in the absence of a smoothly moving target, attentional movements are generally like saccadic eye movements in the sense of switching quickly between locations and then dwelling for a long time relative to the switching time. There is one important difference: Because of the inertia of the eye, larger eye movements take longer than small ones. Attentional movements are not so constrained.

A mathematical formulation of episodes. The purpose of this section is to provide a formal description of the illumination falling on each point x, y of the stage as a function of time t for a given spotlight sequence. Figure 2 illustrates the basic notions inherent in the episodic theory. The sequence E_{-1}, E_0, E_1, E_2, E_3 of Figure 2a represents a sequence of spotlight (or attentional) episodes that began in the past and is continuing into the future. Each episode E_i is characterized by a spatial facilitation function $f_i(x, y)$ (corresponding to the spatial distribution of illumination in the spotlight analogy) and by the times t_i, t_{i+1} that mark the starting and ending points of episode E_i. In the ideal case illustrated in Figure 2a, the beginning and ending of an episode are instantaneous. In the real case (Figure 2b), spotlights turn on with a time course described by the function $G(t-t_i)$ and turn off with the complementary time course.

As a practical matter, the absolute beginning and final ending times of real attentional episodes are awkward to measure directly. Therefore, the approach taken here is to define a theoretical sequence of episodes and to optimize its parameters so that the theoretical performance best matches observed performance. It is shown that what is most critical to prediction is not the actual starting times of episodes but the midpoints of the attentional transitions. This is because the transition function is equivalent to a cumulative probability function. Its derivative is equivalent to a probability density function. The midpoint of the transition is therefore the median of the underlying probability density function. What matters is the central tendency of the density function (e.g., the median), not its tails, which describe the start and finish of the transition.

Figure 2c illustrates an isolated single transition $G(t-t_1)$ beginning at t_1. The clear area to the left and above $G(t-t_1)$ indicates the reign of the initial attentional episode E_0 which is characterized by $f_0(x, y)$. The shaded area at the right and below $G(t-t_1)$ indicates the reign of the post-transition attentional episode E_1, characterized by $f_1(x, y)$. The weighted contribution of E_0 and E_1 to the net attentional state as a function of time is given by Equation 1 of Figure 2. The derivative of the transition function $G(t)$ is $g(t)$ in Figure 2d, the probability density function mentioned earlier; it describes the rate of transition.

In general, there is a long succession of attentional episodes, not merely the two episodes of a single transition. Each attentional episode E_i turns on with time course $G_i(t-t_i)$ and turns off with time course $G_{i+1}(t-t_{i+1})$. In the present development, a single $G(t-t_i)$ transition function is sufficient to describe all of the attention transitions within a trial. However, in the general case, the transition functions themselves may differ, depending on the particular attentional episodes involved in the transition and on other factors. The subscript i is used to indicate that all transitions need not be identical and that different G_i might be used to describe them. The general attention equation (Equation 2 of Figure 2) asserts that attention as a function of space and time $A(x, y, t)$ is simply the sum of all of the episodes during the interval of interest.

Shape of the transition function, G. The most familiar cumulative probability function is the normal function. The normal function is not an appropriate transition function because it is not causal; it begins at $-\infty$. Probably the simplest reasonable causal function is the exponentially limited growth function, $1-e^{-(t-t_i)}$, $t \geq t_i$. The problem with exponentially limited growth is that, at the start of the transition t_i, the maximum transition rate is achieved instantly, that is, the transition rate, $e^{-(t-t_i)}$, $t \geq t_i$, is maximal at $t = t_i$. The absence of any buildup time is unrealistic, so the simple exponentially limited growth function would fail to fit most of the data considered herein.

An exponentially limited growth function corresponds to the signal processing action of a single RC stage (e.g., Sperling, 1964), and it approximates a single synapse. The next simplest function is represented by the succession of two identical RC stages, and this second-order gamma function is a default function used for the examples of this article:

$$G(t-\tau_i) = \begin{cases} 0 & t-\tau_i < 0 \\ \displaystyle\int_0^{t-\tau_i} ue^{-u}du & t-\tau_i \geq 0. \end{cases} \qquad (3)$$

In Equation 3, ue^{-u} is the rate-of-transition function (Figure 2d), and u is the variable of integration (time). We use Greek symbols to distinguish model parameters from stimulus parameters which are indicated by Roman symbols. The starting time of the ith transition is designated as τ_i on the assumption that it is a parameter to be estimated from data. The range of integration begins at τ_i and continues to the moment $t-\tau_i$. For the simpler (one-stage) exponentially limited growth, the rate-of-transition function would be e^{-u} instead of ue^{-u}.

Paradigms That Require Three Attention Episodes: Two Examples

Figure 2d illustrates the attentional transition between two attentional episodes. Although two episodes may suffice to describe some attentional experiments, most require at least three. Figure 3 illustrates three consecutive attentional episodes with the transitions described in Equation 3. Episode E_0 represents the initial state of attention as the trial unfolds. During the trial, a cue to shift attention is received at time t_a. Generally, cue interpretation is not instantaneous, so the attention shift begins later at time t_1. Episode E_1 represents the new state of attention initiated by the cue (i.e., the state of attention appropriate to performing the task indicated by the cue). As the end of the trial approaches, the observer again changes his or her

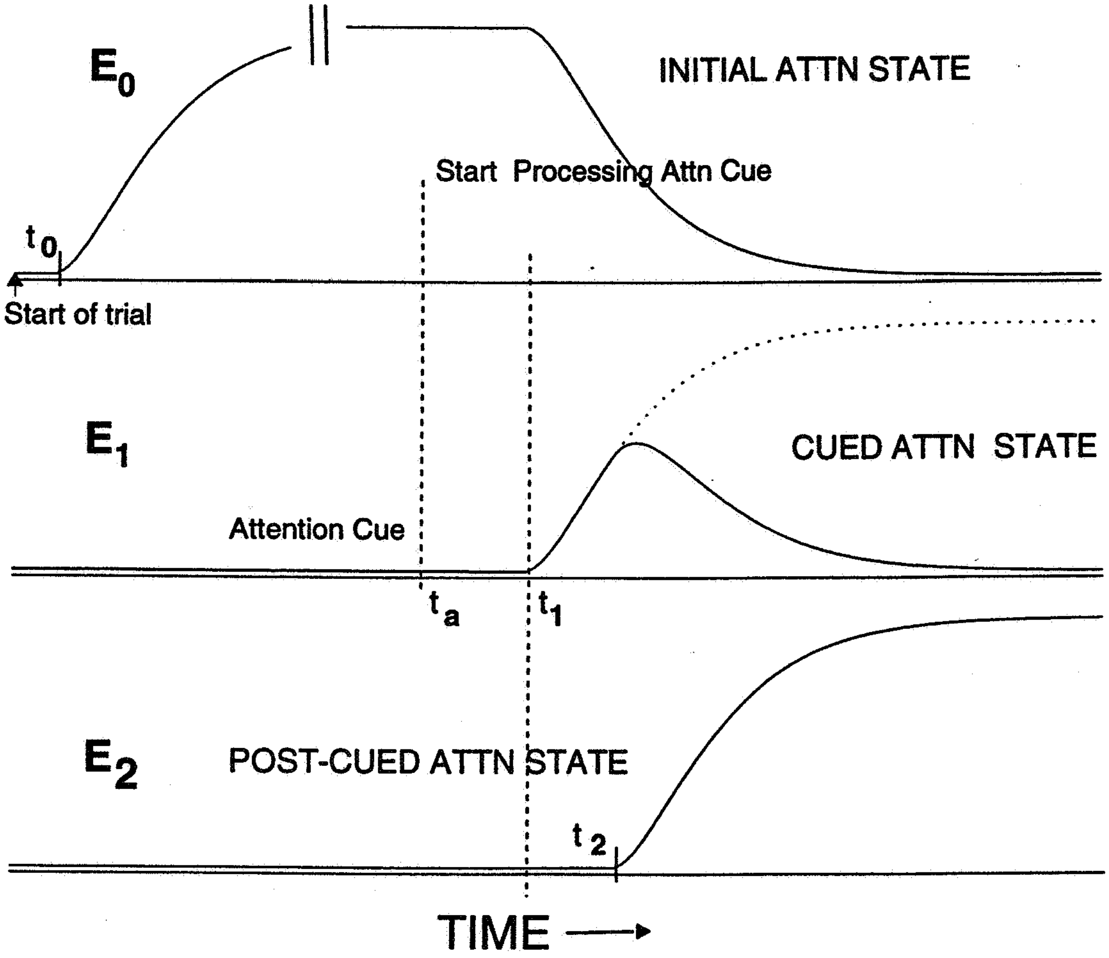


Figure 3. A sequence of three attention episodes. For specificity, the three consecutive episodes of an attention gating experiment are considered. Episode E_0 is the search for the target—the cue to shift attention—at location x_0. The target occurs at time t_a. On detection, the target and its immediate surround are entered into memory; episode E_0 continues until a shift of attention can be initiated at time t_1. Then, in episode E_1, items from the to-be-reported stream at location x_1 are entered into memory. The attention transition function $G(t-t_1)$ describes the changeover at time t_1 from episode E_0 to episode E_1 (Equation 1 of Figure 2). To avoid overloading memory, input to memory is shut off by the self-initiated episode E_2, which follows E_1 so quickly that E_1 never reaches its asymptotic maximum. Had the transition $G(t-t_2)$ not occurred, episode E_1 would have followed the monotonically increasing dotted light line rather than the rising and falling course indicated by the heavy line.

attentional state; the postcued-attention state is episode E_2. This is a generic sequence; we consider two specific paradigms.

(1) Attention-Directed Reaction Time

In attention-directed reaction time experiments, an observer must respond as quickly as possible to a target stimulus. For example, a cue stimulus directs the observer's attention to a particular peripheral location where a target stimulus is most likely to occur. E_0 represents the initial state of readiness to receive the attention cue. E_1 is the new state of peripheral attention in response to the attention cue. Normally, the target occurs during episode E_1; although there are subsequent attentional episodes, they usually are not relevant to the data being collected. On the other hand, if the target has not already occurred at time t_2, the observer assumes that this is a catch trial and reduces his or her readiness to respond. The third episode, E_2, represents this state of reduced readiness by instantiating a spatial function $f_3(x, y)$ that has smaller values than $f_2(x, y)$.

(2) Attention Gating

In a typical attention gating experiment, the observer monitors one location x_0 for a cue that, when it occurs, directs his or her attention to a second location x_1 at which a rapid stream of items is being displayed. The observer's task is to report the earliest possible item at location x_1 after detecting the cue at location x_0. In this paradigm, episode E_0 represents the state of attending to Location 1 and searching for the cue. When the cue occurs at time t_a, it initiates a shift of attention to location x_1. Episode E_1 represents the state of attending to location x_1 and gating items from location x_1 into memory. If episode E_1 were to continue for too long, an excessive number of items would be entered into memory. E_2 represents a self-initiated episode of ignoring the remaining input items to avoid memory overload.

Figure 3 illustrates that episode E_1, which is of primary interest in attention experiments, may be quite short lived when E_2 impinges closely on it. In this respect, the sequence of attentional episodes is analogous to the sequence of states of the human vocal tract during speech. Although the vocal articulators may be programmed to reach target positions for each phoneme in the speech output, speech is so rapid that the target positions characteristic of the production of a phoneme in isolation are seldom reached in dynamic speech. Similarly, in the attention experiments that we consider here, the critical attention episodes (e.g., E_1 of Figure 3) typically are superseded before they reach their asymptote.

Consider the time course of the attention episode E_1 in Figure 3, defined by the difference between identical onset and terminal transitions $G(t-t_1)-G(t-t_2)$. As the interval $\Delta t = t_1 - t_2 \rightarrow 0$, $G(t-t_1)-G(t-t_2) \rightarrow \Delta t g(t-t_1)$. The time course of any such briefly held attention episode is approximated by the derivative $g(t)$ of the transition function. Once the interval Δt between successive attentional episodes is brief, further shortening will not significantly change the time course of attention but will merely reduce the total amount of attention available to the abbreviated episode. This suggests that observers would have little incentive to reduce Δt to its minimum achievable value.

1. *Attentional Theory of Go/No-Go Reaction Time Facilitation*

To develop an empirically testable formulation of the episodic theory of attention, one needs to consider specific observable consequences of attention. Here we concentrate on experiments that give an observer a cue to shift attention to a particular location in space and that measure the observer's reaction time to a visual stimulus that usually occurs at the cued location but occasionally occurs at an unexpected location.

In predicting reaction times to visual stimuli located at various locations in space, one needs to consider components of reaction time other than merely attentional states. There are invariant location-dependent components of reaction time, such as the differential sensitivity of different locations in the retinal field, and there are motoric and other nonattentional processes that contribute to the total reaction time. All of these are lumped together in a single term $k(x)$ that represents the time taken by these components as a function of spatial location x; $k(x)$ is independent of time and instructions.

Temporal and Spatial Attention Multiplicatively Facilitate Motor Reaction Times

Residual and modifiable components of reaction time. In an experiment involving shifts of visual attention, the motor reaction time (MRT) is a function of two variables: location x and time t. There are two critical stimulus events: A cue stimulus that instructs the observer where to attend occurs at time t_a and a target stimulus to which the observer must respond occurs at time t_s. The interval from t_a to t_s is the foreperiod, $t_s - t_a$.

In this section, rather than using an a priori function to describe $G(t-t_i)$ (e.g., the gamma function of Equation 3), we derive the empirical shape of the transition function $G_B(t-t_1)$ directly from that data of a go/no-go reaction-time experiment. (The subscript B is used to indicate the possibility—under investigation here—that the transition function G_B, which applies to basic [i.e., go/no-go] reaction time experiments, may differ in other paradigms.) A latent period before the start of an attention transition would be incorporated directly into the derived shape of $G_B(t-t_1)$, so we assume, without loss of generality, that the attention transition from E_0 to E_1 starts instantaneously with cue arrival (i.e., $t_1 = t_a$).

To generate predictions of reaction times as a function of x and t, one first considers the shortest available foreperiod. In the experiments considered in this section, the shortest foreperiod occurs when the target stimulus occurs simultaneous with the cue stimulus. We conceptualize the mean motor reaction time MRT′ in response to simultaneous cue and target stimuli at t_a as the sum of two terms, k and f, that represent the residual and the modifiable components of the reaction time:

$$MRT'(t_a, x) = k(x) - f_0(x). \tag{4a}$$

The prime symbol is used to indicate a predicted quantity. The sign of f is such that positive values of attention correspond to facilitation—a reduction in predicted reaction time. The subscript 0 indicates the attention episode before and during the arrival of the cue at t_a.

The predicted mean asymptotic reaction time to a stimulus displayed at time t_∞ long after the attentional cue has been displayed and the attention shift is complete is

$$MRT'(t_\infty, x) = k(x) - f_1(x). \tag{4b}$$

The time course between t_a and time t of the attention transition is described by the transition function $G_B(t-t_a)$, $0 \le G_B(t-t_a) \le 1$, which determines the relative proportions of f_0 and f_1 that contribute to the reaction time:

$$MRT'(t,x) = k(x) - [1 - G_B(t-t_a)]f_0(x) - [G_B(t-t_a)]f_1(x). \tag{4c}$$

Note that Equation 4c is essentially of the same form as Equation 1 of Figure 2. We refer to the transition function $G_B(t-t_a)$ interchangeably as the *time course of the attention shift* and the *time course of facilitation*, depending on whether the emphasis is on the assumed attentional process or on its effect on reaction time.

Spatial facilitation. Reaction time facilitation is the difference between the reaction time at time t and location x (Equation 4c) and the reference reaction time at time t_a at the same location x (Equation 4a):

$$MRT'(t,x) - MRT'(t_a,x) = -G_B(t-t_a)[f_1(x) - f_0(x)]. \tag{5}$$

The paradigms under consideration here allow one to solve for only three of the four variables $f_0(x), f_1(x), k(x)$, and $G_B(t - t_a)$, so we define spatial facilitation $f_{01}(x) = f_1(x) - f_0(x)$ and substitute in Equation 5 to obtain

$$MRT'(t, x) - MRT'(t_a, x) = -f_{01}(x)G_B(t - t_a). \quad (6)$$

The left-hand terms of Equation 6 are both observable; they define reaction time facilitation, the improvement of $MRT(t, x)$ relative to the baseline $MRT(t_a, x)$. Equation 6 reasserts that spatial facilitation $f_{01}(x)$ and the temporal attention-shift function combine multiplicatively to reduce the predicted mean reaction time. Positive facilitation is our main concern here. Negative facilitation—an increase in reaction time—would occur if f_{01} were negative and $-f_{01}$ positive. Note that the time course of facilitation $G_B(t - t_a)$ is assumed to be a function only of t and that spatial facilitation is a function only of x; the variables are separable. We derive these functions in relation to the experiment that we consider next.

Go/No-Go Reaction Times to One of N Alternative Stimuli

Shulman et al.'s (1979) experiment appears to provide the most provocative evidence for analog attention shifts among the studies inspired by Posner (Posner et al., 1978). We chose it to demonstrate that even data regarded (without an explicit theory) as yielding the strongest support for an analog theory are actually predicted very accurately by the quantal, episodic theory.

Shulman et al. (1979) measured motor reaction time (MRT) to the onset of a point of light at one of four possible spatial locations in a multistimulus, single-response task. The locations were arranged horizontally and symmetrically across the visual field with the near lights ±8 deg of visual angle from fixation and the far lights ±18 deg from fixation. On each trial, an attentional cue (a left- or right-pointing arrow) was presented at the fixation point at time t_a before the light onset at time t. The cue indicated that the far light on the pointed-to side would be presented with probability $p = 0.7$ and that each other light would be presented with probability $p = 0.1$. On detecting any light, the observer was to press the same key as quickly as possible. Randomly, 15% of the trials were blanks; that is, no light occurred and the observers had to withhold their responses.

Space-time separable predictions. Before considering Shulman et al.'s (1979) interpretation of their own experiment, we analyze it in terms of the quantal theory, which differs substantially from their analysis. Parameter estimation is simplified by converting the multiplicative formulation of Equation 6 to an additive formulation by multiplying through by -1 to make the two sides positive and taking logarithms ($G_B \geq 0$ by definition, and $f > 0$ for the usual case of positive facilitation). Logarithms yield

$$\log[MRT'(t_a, x) - MRT'(t, x)] = \log[f_{01}(x)] + \log[G_B(t - t_a)]. \quad (7)$$

The left side of Equation 7 represents the logarithm of spatiotemporal facilitation, and the right side asserts that this logarithm is equal to the sum of independent spatial and temporal functions.

Parameter estimation results are given in Figure 4, which shows (a) the empirically derived temporal attention-shift function $G_B(t - t_a)$, (b) the derived spatial facilitory function $f_{01}(x)$, (c) their combination (Equation 6), and (d) Shulman et al.'s (1979) data.[1] This account predicts more than 95% of the variance of the data; obviously, it is quite good. What does it mean?

The attention-shift function $G_B(t - t_a)$ starts off like a cumulative probability function of t but reaches a maximum at $t - t_a = 350$ ms, the optimal foreperiod for all cue conditions (Figure 4a). This indicates that, although the attention cue initiates the change from attention episode E_0 to E_1, a new process—the return of attention in the direction of the initial episode E_0—is evident 350 ms after the cue.

Spatial facilitation $f_{01}(x)$ is assumed to be a function only of x, the location of the presented light relative to the cued location. The constraints placed on $f_{01}(x)$ by the data show that $f_{01}(x)$ is (a) greatest for the near light adjacent to the cued far light, (b) intermediate for the cued far light and for the near light on the opposite side, and (c) least for the far light opposite the cued light (Figure 4b). Simply stated, the average observer represented by these data did not move attention sufficiently far laterally to maximize performance at the cued location and apparently also moved attention in the direction opposite to the cue. To determine whether this chosen strategy is indeed the optimal available strategy or simply a miscalculation by observers with insufficient practice, one would have to induce even larger attention shifts to compare their consequences with those of the observed attention shift.

In summary, the spatial and temporal attention functions derived by applying the quantal theory to Shulman et al.'s (1979) data are quite reasonable, and the theory makes excellent quantitative predictions. Shulman et al.'s interpretation of their data and the pro and con arguments are reviewed in Appendix A.

Three-Dimensional Representations of RT Facilitation

Figure 5 shows a three-dimensional representation of the reaction time (RT) facilitation of Shulman et al.'s (1979) data and of the predictions derived from our analysis. The precue state of attention that existed before the attentional cue is represented by the ground plane. The ordinate of Figure 5 represents the difference in attention between the postcue and precue states (Equation 6) rather than attention itself, because the precue state of attention was not determined independently of other factors that influence reaction time. Figure 5b shows the predictions of the episodic attention theory (i.e., no space–time interactions), as described in Equation 4c. Except for small random fluctuations in the data graph, the data (Figure 5a) and theory (Figure 5b) representations are indistinguishable.

In a space–time separable theory, the ridges of the RT facilitation function must be parallel to the space x and time t coordinates, as indeed they are in Figure 5. (Note that whether the

---

[1] The estimation of parameters from Equation 7 was performed in collaboration with Cathryn J. Downing.

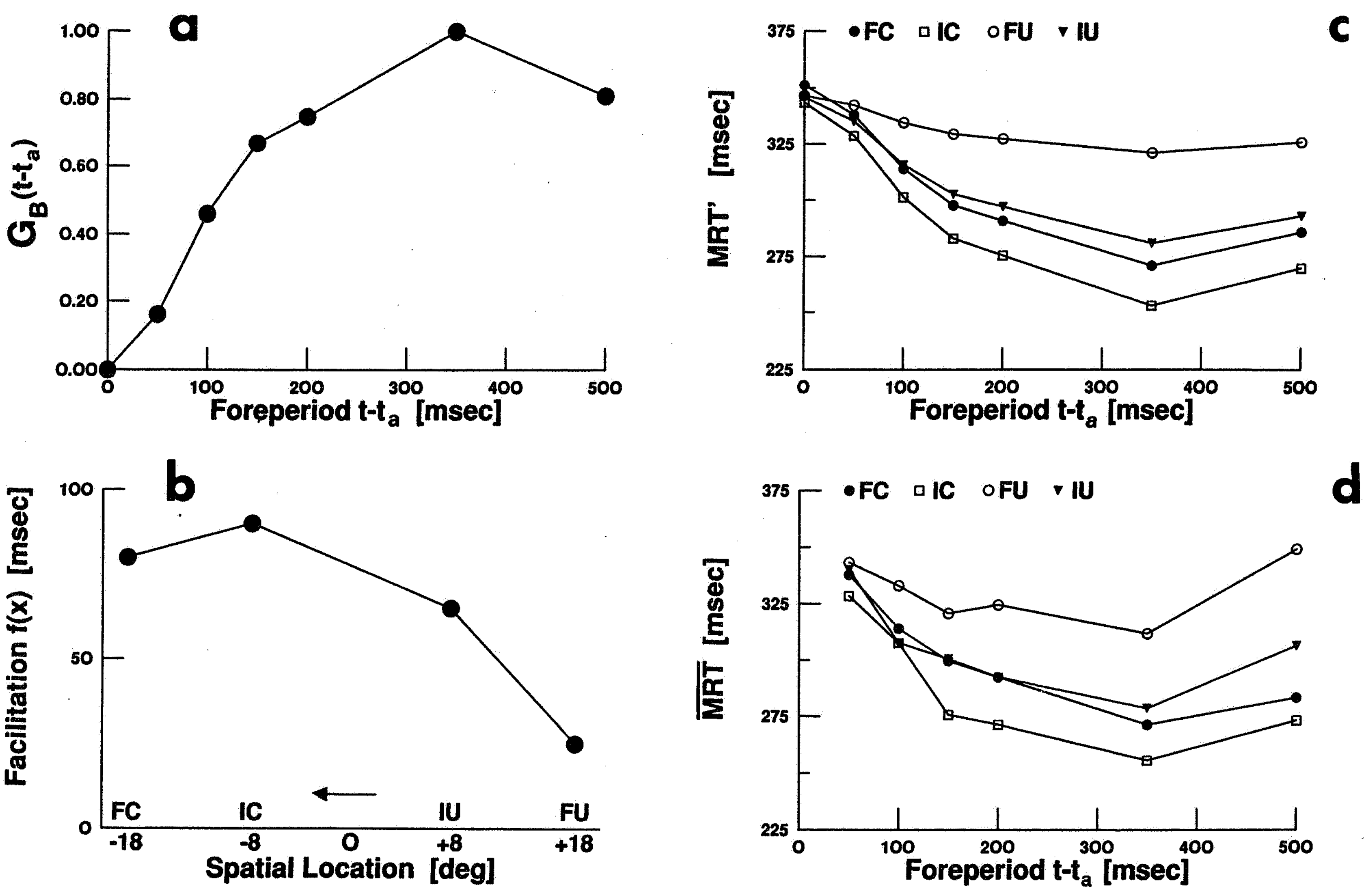


Figure 4. Predictions made by the episodic theory of attention shifts of Shulman et al.'s (1979) multi-stimulus, single-response reaction time experiment. Panel a: The temporal attention shift function $G_B(t - t_a)$ as a function of foreperiod $t - t_a$, where t_a is the time of the warning stimulus and t is the time of the test stimulus. Panel b: The spatial facilitation function $f_{01}(x)$ as a function of spatial location in degrees of visual angle; $f_{01}(x)$ is the average facilitation for trials with attention directed leftward and rightward. The cued direction is indicated by IC and FC. IC is the intermediate location on the cued side; FC indicates the far, cued location; IU and FU are the intermediate and far locations, respectively, on the uncued side. Zero indicates the location of visual fixation; the arrow indicates the side to which attention was directed. Panel c: Predicted motor response times $\text{MRT}'(t, x) = \overline{\text{MRT}}(t_a, x) - f_{01}(x)G_B(t - t_a)$ as a function of foreperiod $t - t_a$ with location x as the curve parameter. Panel d: Mean MRTs observed by Shulman et al. (p. 524). In essence, the subtle differences between the observed IC curve (Panel d) and the predicted IC curve (Panel c) at foreperiods of 150 and 200 ms form the basis of their analog theory (see text for details).

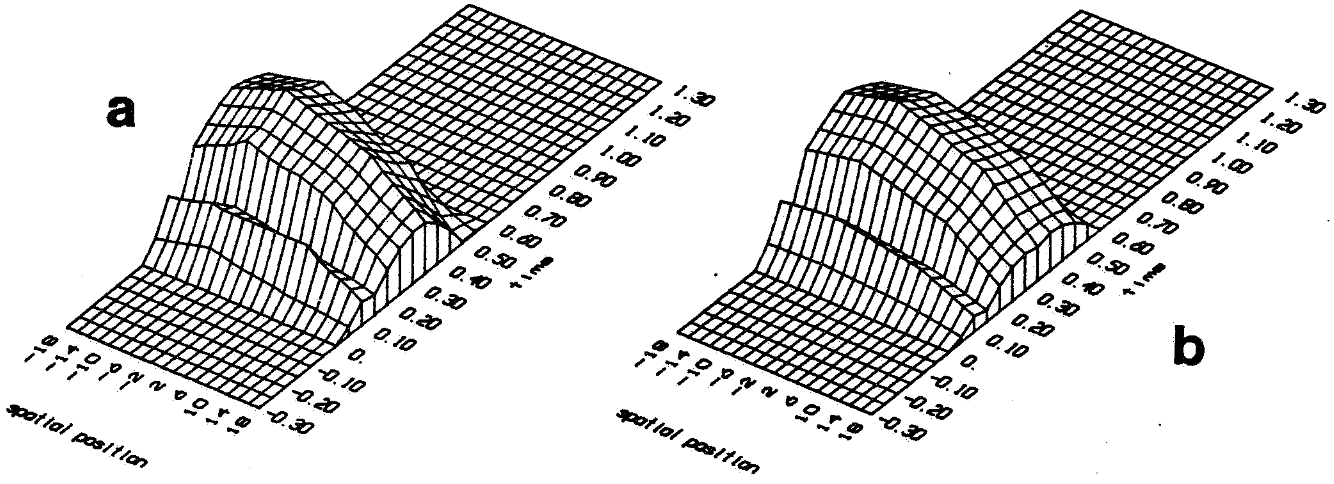


Figure 5. Space–time representation of attentional facilitation in a spatial cued go/no-go experiment. The initial state of attention is taken as the zero baseline. After a cue to shift attention from location 0 to location −18, a stimulus is presented at a spatiotemporal position x, t. The height above the ground plane $f_{01}(x, t)$ represents the speedup of reaction time relative to the baseline condition. Panel a: Data of Shulman et al. (1979). Panel b: Predictions of the space–time separable theory of Equation 6 in the text. At the scale shown here, data and theory do not differ discriminably.

main ridge appears to be parallel to the x-axis or to the t-axis depends on the arbitrary relative scale of these axes.) However, in a moving spotlight analog theory of attention shifts, the main ridge must be slanted, as in the middle section of Figure 1a, because x and t are correlated during movement. Visual inspection of the data, graphed as in Figure 5a, offers no suggestion of an analog attention shift but strongly suggests space–time separability, as we found. What do these observations mean in terms of conventional analyses of reaction time facilitation?

Facilitation incorporates a combination of spatial and temporal components of attention. Although foreperiod facilitation usually is regarded as a purely temporal phenomenon resulting primarily from reduction in temporal uncertainty, there is no purely temporal facilitation in the theory proposed here; rather, there is only spatiotemporal facilitation. This is because the temporal component of attention describes only *when* facilitation occurs; the spatial component of attention is needed to describe *what* stimulus–response combinations actually become facilitated. If one is to predict MRT (t, x), both spatial and temporal components must be specified.

The large existing body of literature on foreperiods is concerned primarily with the efficiency of a warning or attentional cue in informing the observer *when* the test stimulus will occur. The precue used in the many Posner-inspired studies, of which Shulman et al. (1979) is representative, is informative not only about when a test stimulus will occur but also about where the test stimulus will occur. The appropriate generalization of purely temporal facilitation to this case is space–time separable facilitation. Separability means that functions for spatial and for temporal facilitation combine independently, as they obviously do in Shulman et al.'s data. There was no reason a priori why an analog theory might not have described RT facilitation, and it might still be appropriate for yet-to-be discovered tasks.

A Third Attentional State in RT Facilitation

We propose that the apparent decline of facilitation $G_B(t - t_a)$ for Shulman et al.'s (1979) longest foreperiod (largest t) is not an intrinsic property of long foreperiods but is a consequence of the following peculiarity in the experimental procedure. In Shulman et al.'s go/no-go experiment, on each trial on which a stimulus was presented the stimulus was preceded by one of seven preselected foreperiods. In addition, 15% of trials were catch trials on which no stimulus was presented. Ordinarily, from the foreperiod aging distribution, it would be expected that the reaction time after the seventh (longest) foreperiod would be the quickest because, once the sixth foreperiod has been exceeded, the onset time of the last stimulus is completely determined. Because the longest foreperiod occurs one seventh of the 85% of the time that stimuli are presented (12%), and because 15% of the trials are catch trials, the conditional probability of a catch trial rises from 15% before the first elapsed foreperiod to more than 56% after the sixth lapsed foreperiod. Practiced observers (like ideal detectors) are extremely sensitive to such decreasing conditional probabilities of stimulus occurrence, and they reduce their sensitivity accordingly. Thereby, they increase their reaction time as the likelihood of a stimulus diminishes (see Luce, 1986; Sperling & Dosher, 1986). Indeed, in a second experiment, Shulman et al. (1979) used a different, longer distribution of foreperiods, and the longest foreperiod again suffered a relative decline in reaction time although this foreperiod was now much longer than in their first experiment.

According to the episodic theory of attention, an observer takes into account changing stimulus probabilities by initiating the gradual onset of a third attention episode, in which the observer is more cautious and therefore slower than in the second episode. In the episodic theory, this is represented by a spatial function $f_3(x, y)$ that has smaller values than $f_2(x, y)$. The apparently simple go/no-go RT procedure requires three consecutive attentional episodes to account for the data.

2. *Attentional Theory of Choice Reaction Time Facilitation*

Vocal RTs in a Spatially Cued Letter Discrimination Task

Of the choice reaction time studies, Tsal's (1983) study of the effect of attention on the speed of letter discrimination in a choice reaction time paradigm apparently makes the strongest argument for continuous shifts of attention. Tsal measured vocal (naming) choice reaction times (VRTs). His targets were the letters *O* or *X* in Experiment 1 and *C* or *X* and *D* or *O* in Experiment 2. The target could appear randomly at one of two locations ±x deg of visual angle on either side of fixation. In different sessions, x was 4 deg, 8 deg, and 12 deg. The test stimulus was delivered at time t. An attentional cue (a light flash adjacent to the target location) had been delivered earlier at time t_a to indicate the target location to the observer. At each stimulus location, Tsal estimated the longest foreperiod $\tau_c(x) = t - t_a$ that yielded significantly less than the asymptotic maximum VRT facilitation (see Figure 6A). The critical foreperiod $\tau_c(x) - \tau_a$ increased with x at a rate of approximately 32 ms per 4 deg, so Tsal concluded that attention takes 8 ms to traverse a degree.

There are two kinds of problems with Tsal's (1983) study: technical and logical. An elementary technical problem is that Tsal's τ_c, which estimates the point of departure of a curve from its asymptote, is statistically biased because it depends on the amount of data available. A second technical problem stems from the fact that discrimination difficulty increases as viewing angle becomes more peripheral, and it is important to unconfound the effects of difficulty from the effects of peripheral viewing on VRTs. To resolve this issue, in Experiment 2, Tsal compared a difficult discrimination (*D* vs. *O*) at 4 deg with an easy discrimination (*C* vs. *X*) at 8 deg. He found quite similar VRTs as a function of foreperiod $t - t_a$, although he estimated slightly different τ_c for the two conditions. Because these are, in fact, the only unconfounded data of the experiment, Tsal's argument about attentional shifts ultimately rests entirely on the reliability of these small differences between very similar sets of VRTs and on the interpretation of the differences in terms of the latency of an attentional shift.

A third technical problem in Tsal's (1983) study is that he neglected to measure VRTs for targets at unexpected locations. Indeed, there is every reason to expect inferior performance at unexpected locations. But, unless VRTs are demonstrated to suffer at unexpected locations relative to expected ones, one does not know for sure that attention actually moves in the direction of the cue. There conceivably could be only temporal alerting without any spatial shift of attention [e.g., a flat spatial facilitation function $f(x)$ in terms of the analysis of Equations 4a–4c].

An Attentional Preparation Function for Choice RTs, G_C

The interpretive problems in Tsal's (1983) experiment arose because he did not provide a formal argument to justify why there should be any relation of τ_c to attention-shift time. We do so here, as follows. An attentional cue that indicates the location of a to-be-presented peripheral test stimulus yields VRT facilitation because it allows observers to prepare for the test by directing their attention—and, when there is enough time, their gaze—to the indicated location of the letter stimulus. Before the attentional cue, the observer is in a neutral, alert attention state represented by episode E_0. After the cue, there is a transition to a new attention state E_1 in which attention is directed at the side indicated by the cue. The rate at which the observer changes attention episodes is $g_C(t - t_a)$ (Figure 2d), where t is time, t_a is the onset time of the attentional cue, and the subscript C indicates choice reaction time. Assume that $g_C(t - t_a) = 0$ for $t < t_a$ and that $g_C(t - t_a) \geq 0$ for $t \geq t_a$. The cumulative attention $G_C(t - t_a)$ at the end of the foreperiod determines facilitation. Cumulative preparation (attention switching) is the transition function G_C:

$$G_C(t - t_a) = \int_{-\infty}^{t} g_C(t' - t_a)dt' = \int_{t_a}^{t} g_C(t' - t_a)dt'. \quad (8)$$

Our simple assumption is that, at the moment of stimulus onset t, the observer's VRT is facilitated by an amount proportional to the cumulative preparation $G_C(t - t_a)$; that is,

$$\mathrm{VRT}(t - t_a, x) = \mathrm{VRT}(0, x) - G_C(t - t_a). \quad (9)$$

Equation 9 is basically of the same form as Equation 4c except that, because spatial information is unavailable, the spatial terms $f_0(x) = 0$ and $f_1(x) = 1$.

Equations 8 and 9 can be used to derive the rate of preparation $g_C(t - t_a)$ from the data in Figure 6A. The partial derivative of the observed VRTs (Figure 6A) with respect to time gives the rate of preparation $g_C(t - t_a)$, $\partial \mathrm{VRT}(t - t_a, x)/\partial t = -g_C(t - t_a)$; this is shown in Figure 6B.

To illustrate how $g_C(t - t_a)$ yields VRT facilitation, two sample foreperiods are shown in Figure 6. Figure 6C illustrates a long foreperiod that yields much VRT facilitation because it gives the observer a long time to complete the mental preparation. A short foreperiod yields little VRT facilitation because the short time between the attentional cue and stimulus allows little preparation (Figure 6D). All of the points in Figures 6B, 6C, and 6D are derived from the observed data points in Figure 6A and the two assumed asymptotes. Each line segment in the VRT(t_a) data is labeled to facilitate the reader's understanding, and the label is carried by the derived points in Figures 6B, 6C, and 6D. The cumulative preparation function $G_C(t - t_a)$ is not shown separately because it is identical, except for a vertical reflection, to the VRT(t_a) shown in Figure 6A.

It is clear from Figure 6C that the earliest foreperiod at which an attentional cue has less than its maximal effect (Tsal's [1983] τ_c) is a measure of when preparation is asymptotically complete. It is not a measure of when preparation was initiated or of the mean preparation time, except under strong and unlikely assumptions about the invariance under different conditions of the distribution of $g_C(t - t_a)$.

Essentially, τ_c estimates a point of the right-hand tail of $g_C(t - t_a)$. Estimating properties of tails of distributions is tricky. If the variance of the preparation time function $g_C(t - t_a)$ were to vary with peripheral location (e.g., as a result of increased

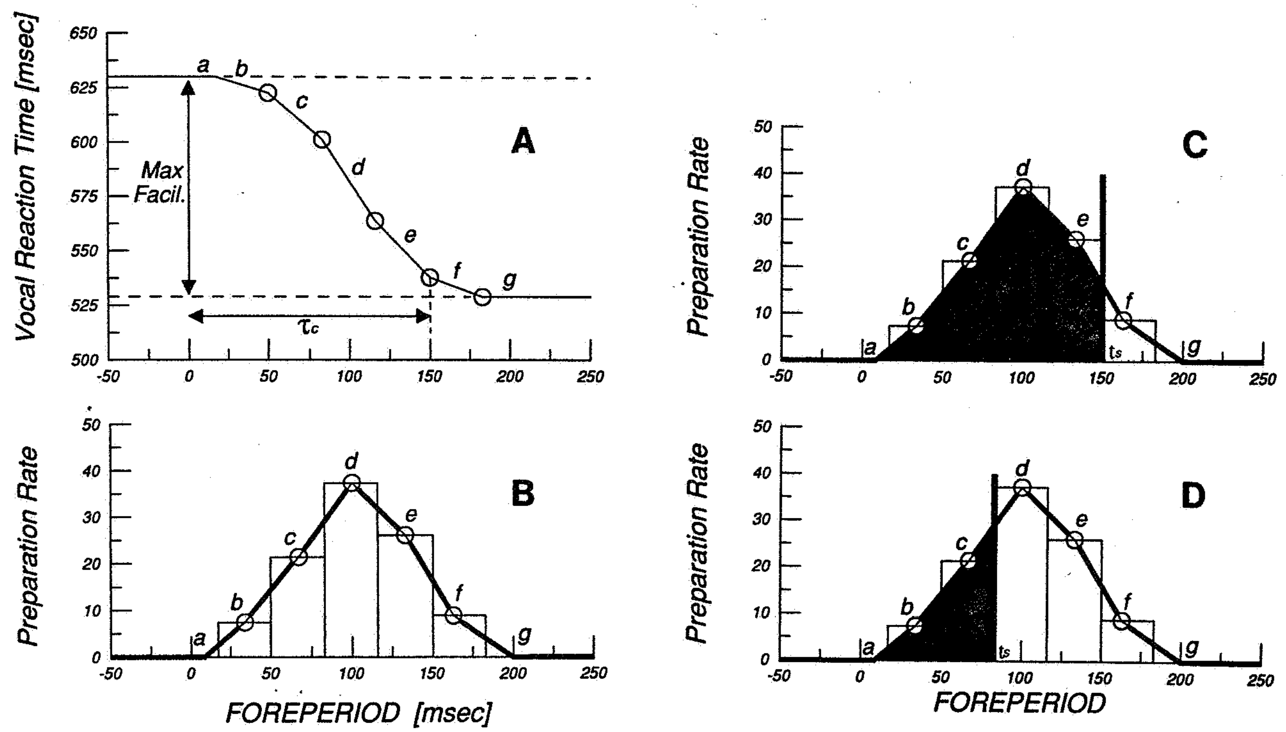


Figure 6. Analysis of vocal reaction times (VRTs) according to the episodic attention shift theory. Panel A: Observed VRT as a function of foreperiod duration (t, time of the stimulus, minus t_a, time of the warning cue; cf. Figure 4d). The stimulus is a letter that occurs 12 deg to the left or right of fixation. Open circles indicate the data from Tsal (1983, Figure 1). Tsal defined the critical time τ_c as the earliest t at which VRT(t) departs significantly from its lower asymptote. The transition function $G_C(t - t_a)$ is not shown separately because it is simply the top-to-bottom mirror reflection of Panel A (cf. Figure 4a). Panel B: The attentional rate of preparation function $g_C(t - t_a)$, the derivative of $G_C(t - t_a)$ with respect to t. The abscissa is time after the warning cue $t - t_a$; the ordinate is the instantaneous rate of preparation $g_C(t - t_a)$. The letter above each horizontal line segment indicates that its height equals the slope of the corresponding segment in Panel A. The segmented curve estimates the underlying attention rate of preparation function. Panel C: Attentional preparation within a trial and τ_c. Coordinates are as in Panel B. The shaded area under g_C represents $G_C(t - t_a)$, the cumulative preparation during the foreperiod $t - t_a$ (i.e., before the test stimulus arrives). The foreperiod illustrated is the critical duration τ_c shown in Panel A. Panel D: Attentional preparation function for a shorter foreperiod ($t - t_s$) than in Panel C. The preparation functions in Panels C and D have been aligned with respect to the warning cue t_a so that the abscissas for all the panels represent the same passage of time in the experiment.

variance between observers in how they prepare for peripheral stimuli), then estimates of τ_c would also increase, even though the mean of $g_C(t - t_a)$ might remain constant or even decrease. Furthermore, because the particular measure that Tsal (1983) used depends on the statistical significance of the difference between VRTs at neighboring foreperiods, it is statistically biased. It confounds the amount of data collected, the particular values of t at which VRTs are sampled, and the reliability of these data with the obtained τ_c estimate itself. An estimate of a basic parameter of $g_C(t - t_a)$, such as its median (the time at which, on average, half of the total facilitation has been achieved), would have been a much more valuable statistic than τ_c in arriving at tenable conclusions about attentional preparation. However, the data that are available fail to provide evidence for analog attention shifts.

In terms of the conceptualization of Equation 9 and Figure 6, Tsal's (1983) statistic τ_c is inappropriate. Because his data apparently were collected with this particular statistic in mind, the data themselves are inadequate for more informative computations. However, more complete data about VRTs as a function of foreperiod could define the preparation function $g_C(t - t_a)$, which would be quite useful. It is an intriguing possibility that the preparation function $G_C(t - t_a)$ for choice reaction time (proposed here for Tsal's data) and attention-shift functions from other paradigms represent similar attentional dynamics (we discuss this issue later).

3. *Attentional Theory of Facilitation of Discrimination Accuracy*

Spatially Cued Pattern Discrimination

Discrimination Accuracy as the dependent variable. Up to this point, two kinds of attention experiments have been analyzed. In go/no-go reaction-time experiments, the effect of selective attention was to speed the go reaction time to particular stimuli. The experiments were conducted so that the proportion of trials with false alarms (premature responses) was small and remained constant across attentional conditions. In the choice reaction time experiments, the effect of selective attention was to speed particular classes of responses. Again, a high level of discrimination accuracy was maintained in all conditions to facilitate interpretation of results. Here we consider an experiment representative of a third class of attention experiments: discrimination experiments. Discrimination experiments usually are conducted with stimulus conditions that maintain response accuracy at levels far below 100% so that the effect of attention in selectively improving accuracy for certain classes of stimuli can easily be observed. Response latency usually is not measured.

Lyon (1987): Procedure and results. Lyon (1987) studied attentional facilitation of discrimination in a procedure that was representative of those of similar investigations (e.g., Bashinski & Bacharach, 1980; Shaw, 1982; Shaw & Shaw, 1977) but provided a richer data set. The task of the observers was to discriminate the orientation of a small test pattern, that is, to judge whether the pattern pointed up, left, down, or right. Four patterns in randomly chosen orientations were flashed simultaneously at four locations (north, west, south, and east) symmetrically disposed around the fixation point. The location of the target pattern was indicated by an attentional cue, and, after a variable foreperiod, the test pattern was presented for one of three exposure durations. All test patterns were followed immediately by a visual noise masking stimulus (Sperling, 1963).

Attentional facilitation was provided by the cue that informed the observer, with 100% reliability, of the location to be tested. Lyon (1987) tried both central and peripheral cues; after practice, both types of cues yielded equivalent performance. The time from the occurrence of the attentional cue to the test flash (the foreperiod) was varied randomly from 17 to 217 ms, and the exposure duration was 34, 50, or 84 ms in a crossed design. Results indicated that accuracy was near chance (25% correct responses) when the foreperiod and exposure durations were short and that accuracy increased to an asymptotic level (this level depended on exposure duration) as the foreperiod increased beyond about 100 ms.[2]

Attentional Preparation for a Discrimination Task, G_D

The theory of attentional facilitation in discrimination experiments is essentially similar to the theory of reaction time facilitation. Lyon (1987) made no distinctions between different spatial locations. The initial spatial function $f_0(x)$ is always 0.25 to represent chance guessing. The cued spatial function $f_1(x)$ is assumed to be 1.0, also independent of x, representing perfect performance when time is unlimited. Therefore, the attentional transition function $G_D(t - t_a)$ describes the growth of attentional facilitation from zero to one. Because the $f_i(x)$ are so simple in this theory, we develop it simply in terms of G_D. The subscript D indicates discrimination, and, in this context, the attention transition function is simply an attentional preparation function.

It is reasonable to expect that the attention preparation function G_D will implicitly account for the effects of exposure duration. However, because there is no upper limit to exposure duration (even though accuracy, by definition, reaches a ceiling at 100%), a monotonically increasing resource-performance characteristic (Norman & Bobrow, 1975) $M^+(D)$ is needed to relate the effective exposure duration D to response accuracy. Because data are unavailable for those stimulus conditions that would most discriminate between alternative formulations, our aim here is to generate reasonable predictions with a very small number of parameters (three) rather than to make precise predictions, as in the previous examples.

Attentional preparation $G_D(t - t_a - \tau_1)$ is assumed to be zero not only before the attentional cue at time t_a but also for an additional fixed latent period τ_1 after the attentional cue (the cue interpretation time). (In the previous examples, the entire shape of the attention transition function was estimated from the data; a latent period would have been inherent in the shape itself. In this example, a standard functional form, gamma, is used, and it requires explicit recognition of a possible nonzero cue interpretation time.) From $t_a + \tau_1$ onward, $G_D(t - t_a - \tau_1)$ is the same function described in Equation 3 and Figure 3, a second-order gamma function with a time constant τ_2 (that determines its slope) and an asymptote of 1.0 (see Equation 10a below):

$$G_D(t - t_a - \tau_1) = \begin{cases} 0 & t < t_a + \tau_1 \\ \dfrac{1}{\tau_2}\displaystyle\int_0^t \frac{(t - t_a - \tau_1)}{\tau_2} e^{-(t-t_a-\tau_1)/\tau_2} dt = 1 - \left(1 + \frac{(t - t_a - \tau_1)}{\tau_2}\right) e^{-(t-t_a-\tau_1)/\tau_2}, & t \geq t_a + \tau_1. \end{cases} \qquad (10a)$$

A stimulus that begins at time t_1 and is terminated by a poststimulus mask at t_2 has an actual stimulus duration of $t_2 - t_1$. Attentional preparation determines the information gained from a stimulus. It is simplest to express the information gained in terms of an effective duration D. D equals the actual duration when attentional preparation is complete; otherwise, D is smaller. As did Reeves and Sperling (1986), we assume that the effective duration D is the integrated product of the attention preparation function and the stimulus availability functions. In the case of a stimulus that is available only between t_1 and t_2, this computation of effective duration reduces to the integral of $G_D(t - t_a - \tau_1)$ between t_1 and t_2,

$$D = \int_{t_1}^{t_2} G_D(t - t_a - \tau_1)dt, \qquad (10b)$$

the shaded area under $G_D(t - t_a - \tau_1)$ in Figure 7a. The time of the warning cue is t_a. When attentional preparation is asymp-

---

[2] Lyon (1990) transformed the data from probability correct to a d' based on a four-alternative forced-choice model. We prefer to consider the raw data without an intervening d' transformation.

totically complete for large t, $G_D(t - t_a - \tau_1) \approx 1$, and effective stimulus duration equals real duration. When the stimulus is replaced by the poststimulus mask before the warning cue, $G_D(t - t_a - \tau_1) = 0$, and the effective duration is zero. Obviously, an effective duration of 0 is a simplification that can be valid only when observers make no significant observations before receiving a cue.

The resource performance characteristic $M^+(y)$ is still needed to relate effective duration (in milliseconds, with a range from 0 to ∞) to probability correct (a pure number ranging, in Lyon's [1987] experiment, from 0.25 to 1). M^+ is required, a priori, to be monotonically increasing and S shaped because, for small signals (small effective durations), accuracy typically is proportional not to signal amplitude directly but to signal power (Carlson & Klopfenstein, 1985; Nachmias & Sansbury, 1974), and, for large signals, accuracy asymptotes to a constant less than or equal to 1.0. Perhaps the simplest such function is the same kind of cumulative second-order gamma function used to describe attentional preparation in Equation 10a, except that here it is used to describe the nonlinear conversion of light flux to probability correct:

$$M^+(D) = 0.25 + 0.75\frac{1}{\tau_3^2}\int_0^D te^{-(t/\tau_3)}dt$$

$$= 0.25 + 0.75\left[1 - \left(1 + \frac{D}{\tau_3}\right)e^{-(D/\tau_3)}\right]. \tag{10c}$$

The constant τ_3 determines the rate at which probability grows as a function of effective duration. Figure 7b shows $M^+(t)$ and data from Lyon (1987). Because effective duration asymptotically approaches actual duration for large t, the experimental data that determine M^+ are asymptotic performances at long foreperiods. The asymptotic data shown (p) are the observed percentage correct of target identifications (p_{obs}), in which chance guessing yields an expected $p_{\text{obs}} = 0.25$.

To predict all of Lyon's (1987) data (accuracy as a function of foreperiod and durations), we estimated three parameters: the cue interpretation time τ_1 (13.8 ms), the rate of attentional preparation τ_2 (21.4 ms), and τ_3 (22.0 ms), which describes the conversion of effective duration (in milliseconds) to accuracy (probability correct). These predictions, together with Lyon's data, are shown in Figure 7c. The theory makes predictions not only for the range of conditions investigated by Lyon but for a much wider range of conditions as well. Although the theoretical predictions capture the essential features of Lyon's data, the

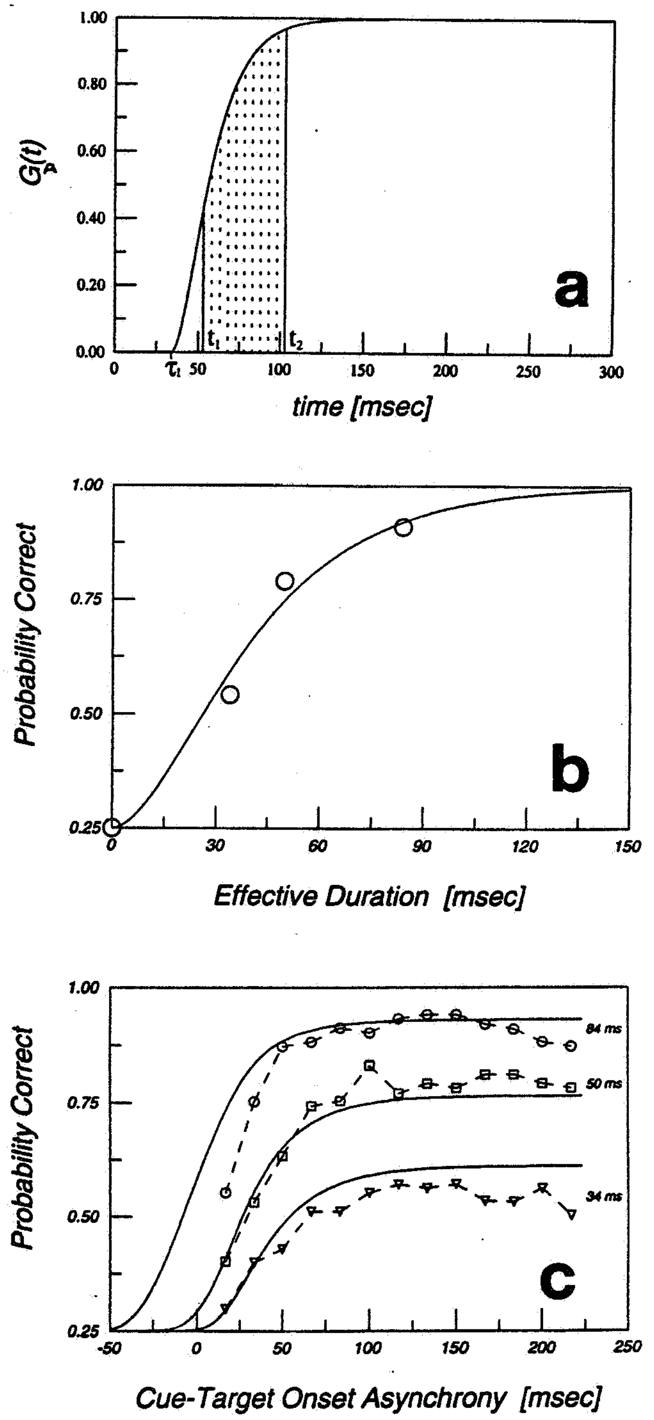


Figure 7 (opposite). Application of an attentional preparedness theory to discrimination accuracy. Panel a: Attentional preparedness $G_D(t - t_a - \tau_1)$ as a function of time $t - t_a$ after the attentional warning cue. The stimulus begins at time t_1 and ends at time t_2 when a poststimulus mask is turned on. The cross-hatched area under $G_D(t - t_a - \tau_1)$ is the effective duration, which represents the information gained from the display (the integrated product of stimulus availability multiplied by attentional preparedness). Parameter τ_1 (13 ms) is the cue interpretation time; τ_2 is the time constant of $G_D(t)$ (related to its slope, Equation 10a). Panel b: Probability correct versus effective duration (information gained). Asmooth, one-parameter monotonic function $M^+(y)$ converts effective duration into probability correct. Data are estimated from Lyon (1987, asymptotic values in Figure 6, p. 13); 0.25 is the chance level of performance. Panel c: Probability correct versus foreperiod (time from cue onset to target onset). The data points connected by dashed lines represent the three target exposure durations studied by Lyon (1987). The solid curves are the predictions $M^+(y)$ of the three-parameter attentional preparation model described in the text.

predictions are by no means a perfect fit to the data. More precise prediction would require more than three parameters. That is, to make predictions comparable in accuracy to those made for the go/no-go and choice RT experiments discussed earlier, one would need to estimate not merely the parameters of theoretical functions but the functions themselves. This, in turn, would require a wider range of data than Lyon collected to provide an adequate test bed. We conclude that extremely simple assumptions capture many essential features of attentional preparation in a discrimination experiment but that more data are needed to proceed further.

2. Two Experiments to Demonstrate Quantal Attention Shifts

The Attention Reaction Time (ART) Paradigm

Up to this point, we have demonstrated that, in go/no-go and choice reaction time experiments, as in a discrimination experiment, dependent measures (reaction time and accuracy) can be simply related to attentive processes with an explicit, formal theory. When such a theory was provided, the evidence for analog attention shifts vanished. To determine whether spatial attention shifts are analog (as in the continuous moving spotlight analogy) or quantal (as in the spotlight switching analogy), we propose a more complex procedure specifically tailored to this issue. The basic notion is analogous to that of a bubble or cloud chamber used for measuring the trajectory of physical particles. Space and time are populated with a large number of items; the particular ones that are absorbed into memory (like collision-produced droplets or bubbles) define the trajectory of attention.

The particular procedure used here was based on the ART paradigm of Sperling and Reeves (1980); this paradigm involves an explicit, computational theory (Reeves & Sperling, 1986), and it has some variants that will enable strong tests of analog versus quantal theories of shifts of attention. In the ART paradigm, the observer initially maintains attention on a peripherally located target stream. Detection is indexed by the accuracy of reporting a target embedded in the stream and by the latency of pressing a key (a motor reaction time, MRT) contingent on detection of the target. Attention is controlled by presenting the target-containing stream so rapidly that observers report they must pay full attention to the stream to detect the target. In Variation 1 of the Sperling and Reeves procedure (Procedure 1), the target was the letter *C*, and the distractor stimuli were other letters randomly chosen. In Variation 2 (Procedure 2), the target was the numeral 0, and the distractor stimuli were other numerals randomly chosen.

In the ART paradigm, the target acts both as a cue for the motor response and as a cue to shift attention from the peripheral stream of letters to a foveal stream of numerals. Each numeral falls on top of the previous one, so that the visual persistence of each is limited by the arrival of the next. The observer attempts to report the earliest four numerals simultaneous with and following the target. The time from target onset to the first reported numeral defines the ART. The ART includes the times for detecting the target, shifting attention, and identifying the first numeral (Sperling & Reeves, 1980). The four reported numerals are used to assess visual memory for the numeral stream (Reeves & Sperling, 1986; Sperling & Reeves, 1983).

The motor task of pressing a key and the cognitive task of shifting attention are concurrent (Sperling, 1984). Both follow detection of the target, and they produce quite similar latencies (Sperling & Reeves, 1980). Experiment 1 addresses the preliminary question of whether these two tasks interfere with one another.

Experiment 2 addresses the question of whether attention shifts depend on the distance traversed.[3] The distance between a peripheral target-containing stream and the central to-be-attended stream was varied, and both ARTs and MRTs were measured. The to-be-attended stream, to which attention was shifted, was always at fixation and thereby remained constant as the size of the required attention shift was varied. The MRT was used to index detection latency for the peripheral target. This provided an essential control for changes in target detectability that occur when the target is moved to the periphery to create stimulus conditions for large attention shifts.

An additional question in Experiment 2 is whether other stimuli, located between the central fixation point and the peripheral stimulus to which the observer attended, have an influence on the velocity of attention shifts. A stream of noise characters was placed 4 deg to the left of the fixation point and 3 deg to the right of the letter stream. We placed the streams no closer than that to avoid lateral interference (such as the slowing of MRTs to a target letter by noise characters within 1 deg reported by B. A. Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1972; see review by Bouma, 1978). The purpose of the interposed stream of noise characters was to provide a potential obstacle between the two stimulus streams. If attention passed across that region of visual space in analog fashion, it might be slowed, or there might be an intrusion of irrelevant characters or features from the obstacle stream into focal attention.

Experiment 1: Measuring Both ART and MRT in Concurrent Tasks

Practice in Concurrent and Isolated Tasks

We initially tested the assumption, important for Experiment 2, that MRT and ART can be used independently to assess components of processing. This would be the case either if the ART and MRT tasks had no influence on each other or if the interference were constant across experimental conditions.

In addition, we were interested in studying the effects of practice, anticipating that several sessions might be needed before an asymptotic level of interference (or noninterference) was obtained. We therefore studied the performance of a naive observer (DP) who performed both ART and MRT tasks either concurrently, as outlined earlier, or individually as isolated control tasks. In each session, four blocks of 32 trials were run. First, the concurrent task was run; then, in alternating order, the two isolated control tasks were run; and, finally, the concurrent task was run again. It was decided—on the basis of prior experiments involving the other observer EW—to run DP with

---

[3] A preliminary report of Experiments 1 and 2 is contained in Weichselgartner, Sperling, and Reeves (1985).

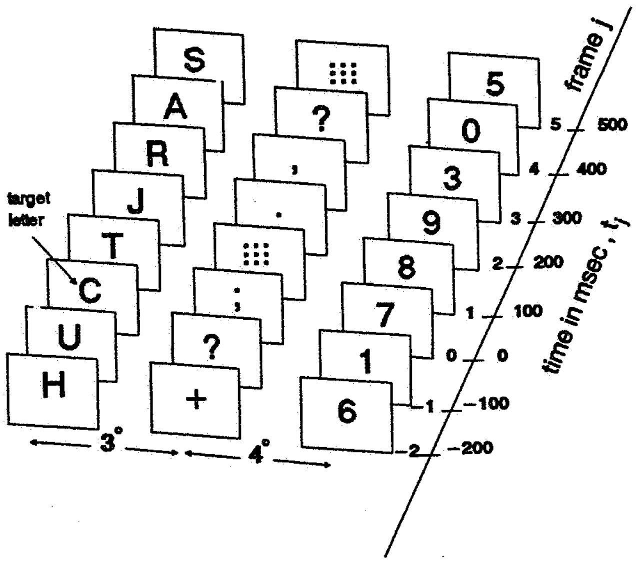


Figure 8. Illustration of the attention reaction time procedure. On the face of a cathode-ray tube, 25 frames—composed of letters and numerals in Experiment 1 and letters, numerals, and special characters in Experiment 2a—appear in serial order. The presentation rate of the letter target-containing stream (and of the special characters) is 6.67 frames per second. The presentation rate of the next to-be-attended numeral stream is 10 frames per second. The frame exposure is 18 ms. With equal probability, 7 to 14 frames precede the target letter *C*, and the observer's task is to attempt to report the four earliest occurring numerals simultaneous with and subsequent to the target. To disambiguate the analysis of the response sequence, numerals in frames 0 to 9 are all different.

a letter frequency of 6.67 characters per second and a digit frequency of 10 characters per second.

Experiment 1 was run for 19 sessions, but one block (32 trials) of concurrent tasks in Session 18 was lost. We thus had 2,400 trials [(19 × 128) − 32] involving DP available for analysis (608 trials for each of the two isolated control tasks and 1,184 trials for the concurrent task).

Method

Observers. One observer was a paid graduate student (DP), and the second observer was Erich Weichselgartner (EW), who had been involved in 1,500 trials with the paradigm before Experiment 1. DP was initially naive. Both observers had normal vision.

Attention Reaction Time (ART) procedure. From the observer's point of view, the task consisted of the following components: target detection, manual reaction-time response concurrently with a shift of visual attention, and retention of and subsequent recall of four items from visual short-term memory.

The ART procedure, illustrated in Figure 8, required the observer to fixate on a stream of numerals presented on a cathode-ray tube (CRT). The numerals appeared one after the other, in the same location, at a constant rate of 10 per second. To the left of the numerals was a stream of letters at a rate of 6.67 per second. Embedded in the letter stream was the target letter *C*. Explicitly, the task of the observer was to detect the target letter in peripheral vision while maintaining fixation on the numeral stream and then to report the earliest possible numeral. Implicitly, the ART task required the observer to pay attention to the letter stream until the target was detected and then to shift attention as quickly as possible to the numeral stream.

The numeral stream was arranged so that the 10 numerals that followed target onset were all different. These 10 numerals are called the critical sequence, and the positions in the numeral stream that they occupy are called the critical set of positions. The time delay between target onset and onset of the first reported numeral is a measure of the time needed to shift attention from the target to the numeral stream (it also includes other components). Sperling and Reeves (1980) called this time the ART for the trial.

Motor Reaction Time (MRT) procedure. On target detection, the observer also had to release a response key; this motor task became quite automatic with practice. These trials yielded a distribution of MRTs. We defined a critical interval extending from target onset to 800 ms after target onset. Trials with motor responses outside the critical interval represented anticipations or misses and were excluded from analysis (see Sperling & Reeves, 1980, p. 349).

Apparatus. The stimuli were presented on a Hewlett Packard 1310A CRT with fast P4 white phosphor. The cathode-ray tube was driven by a Digital Equipment Corporation PDP-11/34 computer via a specially designed display interface and a software system developed for real-time vision experiments (Melchner & Sperling, 1980). The resolution of the display was 1,024 × 1,024 pixels. Each pixel could be intensified individually.

Stimuli. The characters used in Experiment 1 consisted of pixels, all of which were intensified in less than 1 ms and embedded in a 10 × 7 array. Characters consisted of uppercase letters A–Z, the numerals 0–9, and six special characters (· + , : ? and #). Each alphanumeric character subtended a visual angle of approximately 1.32 deg vertically and 0.1 to 0.92 deg horizontally at the viewing distance of 0.75 m.

Each character was refreshed (presented) five times at 4-ms intervals. The screen was blank between characters. The background illumination of the display was 3.5 cd/m^2. Because of their similarity to numerals, the letters *B, I, O, Q, S,* and *Z* were deleted from the character set. The character *C* was chosen as the target, and the remaining 19 letters served as distractors.

Trials. A trial began with the message "Ready" on the CRT, together with a fixation dot. When the observer was confident of his fixation on the dot and prepared to detect the target at the location where it would ultimately appear, he pressed a response button. The numeral stream then began at the fixation point, and the letter stream began simultaneously 2 deg to the left of the numerals. The target letter (*C*) appeared randomly and with equal probability at any position from Positions 8–15 in the letter stream. Immediately on detecting the target, the observer pressed the response button under the right index finger and simultaneously shifted attention to the numeral stream. The observer attempted to report the earliest possible numeral (ideally the numeral that was simultaneous with the target), as well as the following 3 numerals. The numeral stream stopped when 2 s had elapsed after the motor response and at least 13 numerals had occurred after target onset. The observer then entered the 4 remembered numerals onto a keyboard of 10 buttons labeled 0 through 9.

Control experiment for intrusion errors. In the experiments reported here, the critical sequence of 10 all-different numerals began simultaneously with the onset of the target. The optimal temporal placement of the critical sequence is important because every response numeral is indexed according to its position within the critical sequence, although it may also occur elsewhere in the trial sequence. A control experiment was devised to check whether observers would report numerals from outside the critical sequence. Three observers (2 of whom did not participate in the present study) were run with the procedure outlined earlier, not knowing that one numeral randomly was omitted from the critical sequence. If intrusions from numerals outside the critical sequence occur, one would expect the observers to report numerals that are not members of the critical sequence. None of the observers reported a not-presented numeral as a first response (in 200 trials). Only the first responses were used here (to determine ARTs); Responses 2, 3, and 4 can be used to determine other properties of short-term memory. Previous work (Reeves, 1977; Sperling & Reeves, 1980) had shown that the first of four responses (when four were required) was statistically indistinguishable from a single response (when only one was required).

Feedback. After the responses, an appropriate feedback message was shown on the CRT. The observer first received MRT feedback ("early" if the response button was released before target onset, "slow" if the MRT was more than 400 ms or less than 800 ms, and "late" if the MRT was more than 800 ms). The feedback display was terminated by a button press. If the motor response of the observer occurred within 400 ms after target onset, no MRT feedback was given.

Feedback about the four reported numerals appeared next. The observer was presented with the four numerals that he had typed onto the pushbutton board. With the following button press, the observer could either exclude the trial from analysis (by pressing Button 9, if he had typed in a wrong numeral) or acknowledge the trial (by pressing any other button). Then the observer was presented with a display of the critical sequence (as a row of numerals), below which the response sequence was shown (also as a row of numerals). The observer was instructed to use the information about the critical sequence and the MRT feedback to improve subsequent responses. The next button press produced the "Ready" message and initiated a new trial.

Results

The results of Experiment 1 are shown in Figure 9a. There was little variation in DP's MRTs over the 19 practice sessions. However, there seemed to be a slight trend for ARTs to become faster with practice. A single-variable analysis of variance (ANOVA) partially confirmed these impressions: Practice did not speed up MRTs significantly in either condition, nor did practice speed up the ARTs in the concurrent condition. However, ARTs in the isolated condition significantly decreased over sessions, $F(18, 589) = 2.77, p < .01$. The overall finding of stability in MRTs over sessions of fixed mapping (the target was always *C*) is typical (Schneider & Shiffrin, 1977).

In contrast to the small influence of practice on DP's reaction times, the type of task (isolated control vs. concurrent) consistently had a larger effect. This was found with a large-sample *t* test for differences between means (Hays, 1981). DP's mean MRTs and mean ARTs were both significantly faster in the isolated control task than in the concurrent task: MRT, $t(1790) = 3.07, p < .01$; ART, $t(1790) = 2.80, p < .01$.

For EW, task interference was measured after (not as part of) extended practice. Three sessions of three blocks of 40 trials were run; the order of tasks was counterbalanced across sessions. EW's mean MRTs and mean ARTs were significantly faster for the isolated control task than for the concurrent task: MRT, $t(238) = 1.94, p < .05$; ART, $t(238) = 2.13, p < .05$. EW's reaction times (960 trials) are illustrated on the right side of Figure 9a.

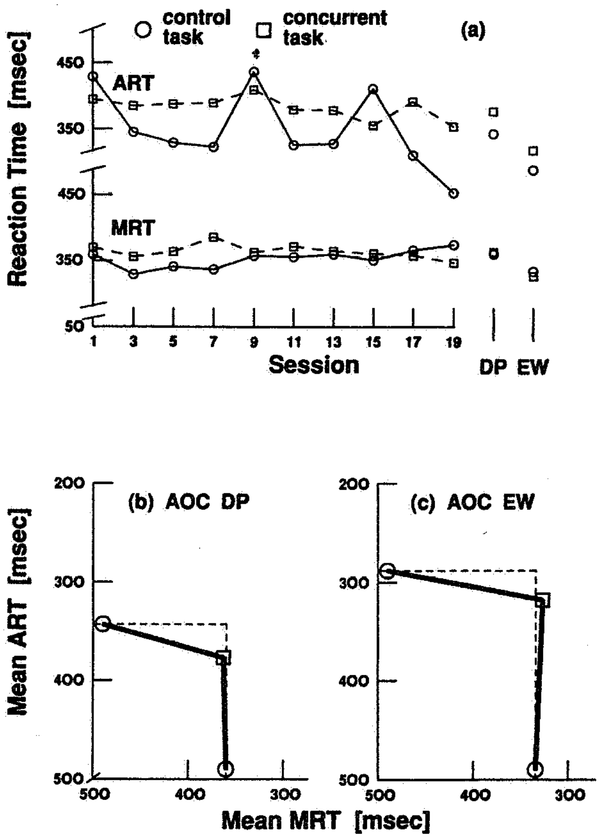


Figure 9. Effects of concurrent tasks on motor reaction time (MRT) and attention reaction time (ART) in Experiment 1. Panel a: Effect of practice. Mean ARTs (top) and MRTs (bottom) for observer DP in pairs of 19 consecutive sessions showing isolated control and concurrent conditions. On the right are average MRTs and ARTs over the last 10 sessions for DP, along with the final performance of observer EW in a similar experiment. Panel b: Operating space representation of the control and dual task performance for observer DP. The heavy line represents the attention operating characteristic (AOC). Mean reaction times for the isolated control tasks (open circles) are indicated on the abscissa (MRT) and ordinate (ART); open squares represent performance in the concurrent task. Panel c: AOC for observer EW.

Discussion

MRT is independent of concurrent ART but ART is slowed by MRT. With practice, the MRTs shown in Figure 9 became increasingly independent of the presence or absence of the concurrent ART task. Differences between isolated and concurrent MRTs were seen only in the earlier sessions. For the second half of DP's sessions (Sessions 10 to 19), the overall MRT difference between isolated and concurrent conditions was only 3 ms and was statistically insignificant. The summary data for DP (Sessions 11–19) are shown at the extreme right of Figure 9a. EW was already highly practiced in these procedures; his summary data in Figure 9 were based on 960 trials. Unlike the MRTs, the ARTs shown in Figure 9 showed consistent effects of the MRT task throughout the experiment. Requiring the MRT task slowed ART by about 30 ms for both observers.

Attention operating characteristics. In Figures 9b and 9c, the data shown on the right side of Figure 9a are graphed as operating space representations of isolated control- and dual-task performance (Navon & Gopher, 1979; Norman & Bobrow, 1975; Sperling, 1984; Sperling & Melchner, 1976, 1978). The RTs for the isolated control tasks are plotted on their individual axes; the RTs for the concurrent tasks are centered in the graph. The connected lines are attention operating characteristics (Sperling & Melchner, 1978).

Good performance is represented up and to the right in Figures 9b and 9c. That is, a dual-task deficit would move data points downward (increase in ART) and leftward (increase in MRT) in relation to the independence point at the intersection of the dotted lines perpendicular to the isolated control-task RTs. It is obvious that MRTs are not impaired by a concurrent task but that there is an ART loss.

It is known that, within a condition, MRTs and ARTs are correlated (Sperling & Reeves, 1980). That means that MRTs and ARTs are correlated measurements of the same detection processes, because subsequent processes differ for ARTs and MRTs. Here we found that, after considerable practice, MRTs were about the same whether or not the attention shift was required. Therefore, MRT—alone or with a concurrent ART task—can be used to index the target detection component of performance independently of the attentional manipulation. In Experiment 2, we measured distance-dependent ARTs and subtracted out the distance-dependent component of target detection. The residual was the true latency of attention shifts ASL(x) as a function of distance x.

Experiment 2: Effects of Distance and Visual Obstacles on Attention Shifts

Distance

We now approach the main question of whether attention shift latencies depend on the distance "traveled." The empirical question to be answered is as follows: What is the effect of distance between target and numeral streams on ART and MRT? We studied this by (a) having the observer fixate on the numerals at all times, so that the difficulty of numeral detection was kept constant, and (b) varying distance by varying the eccentricity x of the target stream. Due to lower acuity in the periphery, we expected the latency $\tau_d(x)$ of the internal process of detecting a target to increase with x. We also expected, correspondingly, that the probability of correct detection $P_c(x)$ would fall with increasing x. Let MRT(x, P_c), ART(x, P_c), and $\tau_d(x, P_c)$ be motor reaction time, attention reaction time, and target detection latency as functions of eccentricity x and P_c. Assume an additive model of processing time in which ARTs and MRTs share a common detection stage (Sperling & Reeves, 1980) but differ in subsequent stages. Specifically, the ART has a component stage that reflects the distance-dependent attention-shift latency ASL(x). Thus,

$$\mathrm{MRT}(x, P_c) = \tau_d(x, P_c) + \mathrm{MRT}_{\mathrm{res}} \qquad (11a)$$

and

$$\mathrm{ART}(x, P_c) = \tau_d(x, P_c) + \mathrm{ASL}(x) + \mathrm{ART}_{\mathrm{res}}, \qquad (11b)$$

where $\mathrm{MRT}_{\mathrm{res}}$ and $\mathrm{ART}_{\mathrm{res}}$ are the residual MRTs and ARTs and ASL(x) is the component of the attention shift whose latency depends on the distance shifted. In line with the results of Experiment 1, we assume that MRT is uninfluenced by the ART task. We also assume that the effect of the MRT task on ART does not vary with x.

We assume that detection time as a function of x and P_c is the same for ARTs and MRTs; thus, the difference between Equations 11b and 11a,

$$\mathrm{ART}(x, P_c) - \mathrm{MRT}(x, P_c) = \mathrm{ASL}(x) + \mathrm{Constant}, \qquad (11c)$$

provides a measure of the attention shift time ASL(x). If distance had no effect on attention shifting latency, then ASL(x) would be a constant independent of x. Alternatively, an ASL(x) that increased with x would yield estimates of the time for attention to move.

Method: Experiment 2a. Initial Observations

Observers. The same observers were used as in Experiment 1. EW was highly familiar with the task and target (the letter *C*; about 9,000 prior trials), and DP had, at this point, participated in about 3,000 trials.

Procedure. The procedure was identical to that described in Experiment 1, except that eccentricity was varied. Four different conditions (three different digit stream to letter stream distances and the obstacle condition) were presented in a mixed-list design. DP participated in 1,500 trials in eight sessions, and EW participated in 1,000 trials in nine sessions. The distances between the digit stream and letter stream were 7 deg (with and without the interposed noise stream), 4 deg, and 2 deg. The conditions were randomized over trials, each condition occurring with a probability of .25 on any given trial. As in Experiment 1, the letter stream was presented at 6.67 characters per second, and the numeral stream was presented at 10 characters per second. In the obstacle condition, the letter and digit streams were separated by 7 deg, and a stream of noise was inserted between them. The noise stream consisted of special characters (· + , : and ?) presented at a rate of 6.67 per second. The noise, a "dynamic obstacle," was 4 deg to the left of the digit stream and 3 deg to the right of the letter stream.

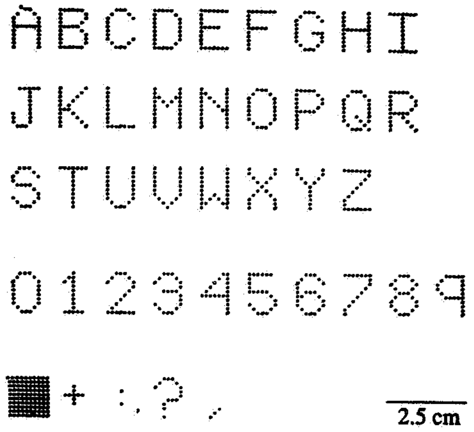
Method: Experiment 2b. Replication

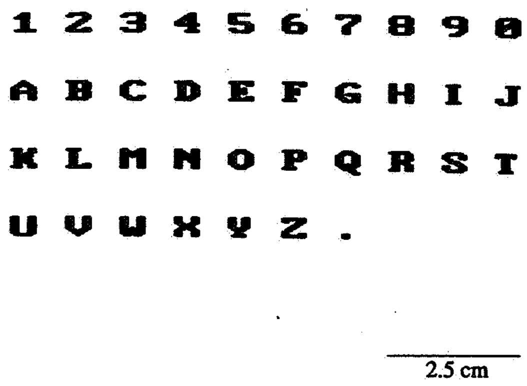
There was a potential for the results of Experiment 2a to be controversial, so we reran the same basic procedure as in that experiment but with many minor variations to determine whether the main results could be replicated. Experiment 2b was run in a different country with observers who spoke a different language. The characters composing the target and test streams were switched; numerals replaced letters, and vice versa. A completely different display system was used with a brighter background and smaller, differently shaped characters that necessitated other changes in procedure.

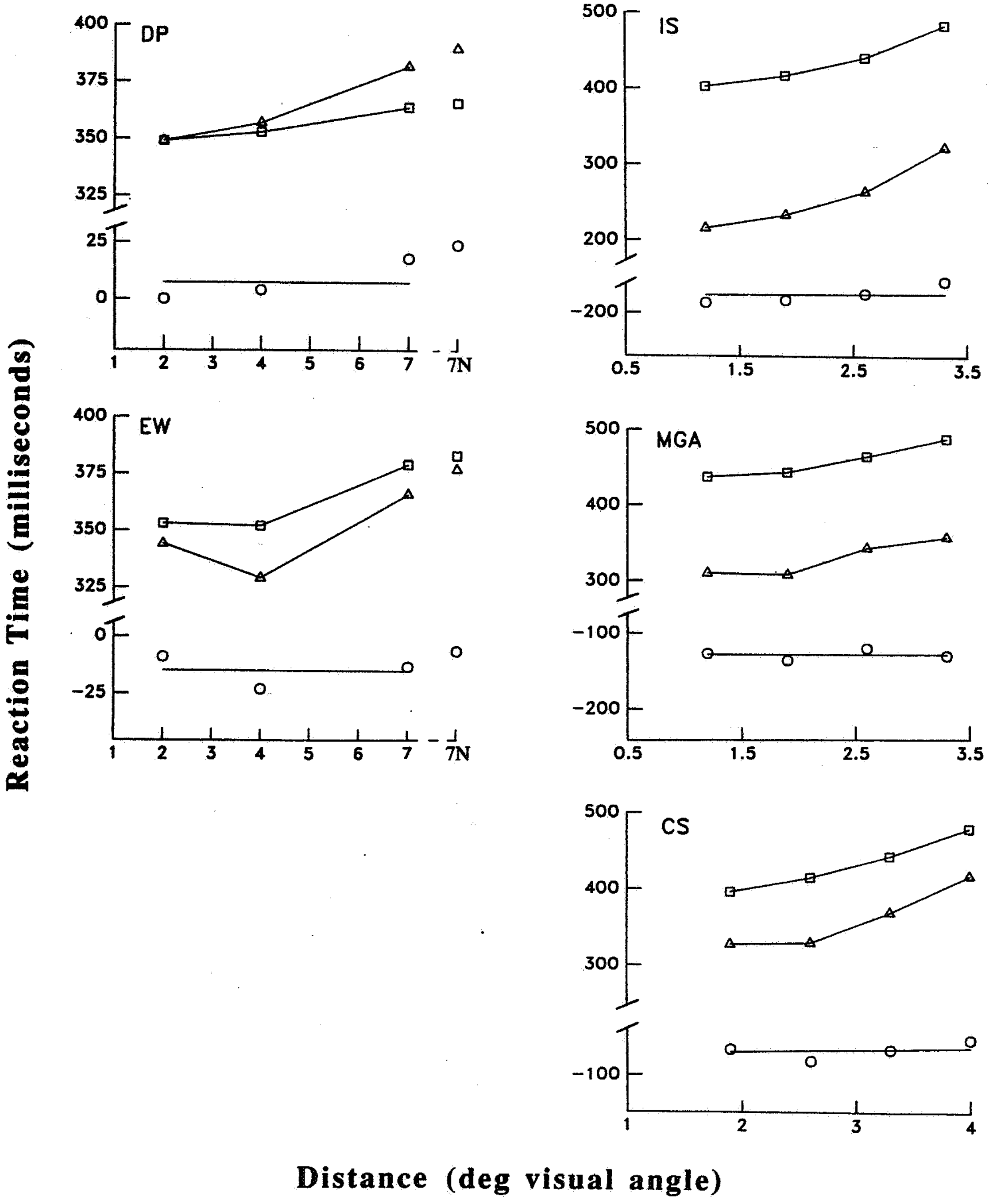
Observers. Three members of the Regensburg (Germany) University Psychology Department observer pool participated in Experiment 2b (CS, IS, and MGA). All three had been involved in 2,000 practice trials before Experiment 2b and had normal or corrected-to-normal vision.

Apparatus. The stimuli were presented on a video raster monitor with a fast white phosphor driven by an Olivetti M24 microcomputer. The computer software allowed precise control over stimuli and responses.

Stimuli. The characters were generated with the computer's 40 × 25 text-mode video driver. Characters consisted of uppercase letters A–Z, the numerals 0–9, and the punctuation mark; the characters are shown on the upper right side of Figure 10. Each alphanumeric character subtended a visual angle of approximately 0.5 deg vertically and 0.1–0.6 deg horizontally at the viewing distance of 0.5 m. Each character


ABCDEFGHI
JKLMNOPQR
STUVWXYZ
0123456789
+ :,?,
2.5 cm


1 2 3 4 5 6 7 8 9 0
A B C D E F G H I J
K L M N O P Q R S T
U V W X Y Z .
2.5 cm


Reaction Time (milliseconds)
Distance (deg visual angle)
DP
400
375
350
325
25
0
1
2
3
4
5
6
7
7N
IS
500
400
300
200
-200
0.5
1.5
2.5
3.5
EW
400
375
350
325
0
-25
1
2
3
4
5
6
7
7N
MGA
500
400
300
-100
-200
0.5
1.5
2.5
3.5
CS
500
400
300
-100
1
2
3
4

was presented for 18 ms. The screen was blank between characters. The background illumination of the display was 10.0 Lux.

Procedure. The procedure was the same as that used in Experiment 2a, except that the letter stream and numeral stream were reversed. Thus, the numeral stream was the target stream, and the letter stream was the test stream. This extended the critical sequence to all stimulus positions (compare the control experiment described in the *Method* section of Experiment 1). The numeral 0 was chosen as the target, and the remaining nine numerals served as distractors. Because all 26 letters of the alphabet were used for the test stream, each numeral except for the target was, on average, repeated 2.5 times. No two identical numerals followed one another in succession. Immediately on detecting the target numeral 0, the observer was required to retain the earliest possible letter from the test stream.

Presentation rates and distances were chosen on the basis of the results of the practice trials. Because viewing conditions were more difficult than in Experiment 2a, presentation rates were generally slower and distances narrower. The presentation rates were adjusted for all observers such that their MRTs were roughly equivalent. Presentation rates for the numeral stream and letter stream were identical (6.67 characters per second for observer CS, 5.0 characters per second for IS, and 3.33 characters per second for MGA). The distances between the digit stream and letter stream were 1.2 deg, 1.9 deg, 2.6 deg, and 3.3 deg for IS and MGA and 1.9 deg, 2.6 deg, 3.3 deg, and 4.0 deg for CS.

Table 1

Experiment 2a: Attention Reaction Times (ART), Motor Reaction Times (MRT), Their Correlations (r), and Differences (DIF = ART−MRT) for Three Distances Between Digit Streams and Letter Streams and for the Obstacle Condition

Distance (degrees)	*n*	*ART*	*SD*	*MRT*	*SD*	*DIF*	*SD*	*r*
				Observer DP				
2	345	349	160	349	85	0	158	0.29
4	340	357	153	353	93	4	163	0.19
7	360	382	139	364	107	18	142	0.35
7 + obstacle	355	390	168	366	96	24	166	0.31
				Observer EW				
2	236	344	148	353	107	−9	156	0.28
4	267	329	127	352	100	−23	134	0.32
7	253	366	158	379	126	−13	159	0.39
7 + obstacle	244	377	149	383	120	−6	133	0.53

Note. Reaction times and standard deviations (SD) are given in milliseconds. The number of trials (*n*) was determined by a random number generator that assigned equal probability of occurrence to each distance condition.

Results

Main effects. Figure 10 shows ART and MRT as a function of the required attention shift distances (x). In 25 of 26 instances, both ART and MRT increased slightly with x. Nevertheless, the difference between ART and MRT and, thus, the attention shift latency ($ASL(x)$ of Equation 11c) remained approximately the same.

This pattern of results was statistically confirmed by a single-variable (eccentricity) ANOVA. For all observers, the $ART(x, P_c)$ increased significantly with eccentricity x: CS, $F(3, 473) = 21.01, p < .01$; DP, $F(2, 1142) = 5.31, p < .01$; EW, $F(2, 753) = 4.44, p < .05$; IS, $F(3, 475) = 28.54, p < .01$; and MGA, $F(3, 950) = 9.90, p < .01$. $MRT(x, P_c)$ increased nearly significantly for DP, $F(2, 1142) = 2.49, p = .083$, and significantly for the other observers: CS, $F(3, 473) = 38.21, p < .01$; EW, $F(2, 753) = 4.71, p < .01$; IS, $F(3, 475) = 42.27, p < .01$; and MGA, $F(3, 950) = 22.39, p < .01$.

ART-MRT difference (interaction). Although there possibly is a tendency for the ART−MRT difference to be slightly greater with more peripheral targets, the results of the ANOVA indicate that the differences between $ART(x)$ and $MRT(x)$ were not statistically significant for any observer. ARTs, MRTs, their differences, and their correlations for the three eccentricities and the obstacle condition are listed in Table 1. The results of the obstacle condition are shown on the right side of Figures 10a and 10b. For both observers in Experiment 2a, a large-sample t test for differences between means revealed no significant differences between the 7 deg condition alone and the 7 deg plus obstacle condition. This result is in accord with previous research (Flowers & Wilcox, 1982; Murphy & Eriksen, 1987). ARTs, MRTs, their differences, and their correlations for the four eccentricities used in Experiment 2b are listed in Table 2.

Discussion: Three Possible Methodological Artifacts Are Excluded

Our results show no statistically significant effects of eccentricity x on the differences between ART and MRT. Insofar as these both have the component of target detection latency (D) in common, and insofar as certain methodological issues are addressed, ART−MRT difference represents a pure measure of the variation of attention shift time with distance uncontaminated by processing capability at different eccentricities. Before one can conclude that distance has no effect on attention shift time, one must first ask whether this result was genuine or artifactual. We consider three possible artifacts—guessing, parallel processing of the two streams, and reconstruction from visual memory—and reject each in turn.

Guessing. Observers might have guessed in some of their responses, and if the guessing rate had varied with eccentricity, our estimates of ART would have been contaminated. We can discount guessing because guesses did not occur in the control experiment (see the *Method* section). Moreover, the ART dis-

Figure 10 (opposite). Top: Stimuli used in Experiments 1 and 2a (left) and Experiment 2b (right). The stimuli were white on a black background. Bottom: Attention reaction times (triangles), motor reaction times (squares), and differences (ART−MRT; circles) as a function of attention shift distance (distance from target stream to the next-to-be-attended stream); 7N indicates the noise condition. The ART−MRT differences are expected to be constant, not necessarily zero, for each observer; straight horizontal lines are drawn through their means.

Table 2

Experiment 2b: Attention Reaction Times (ART), Motor Reaction Times (MRT), Their Correlations (r), and Differences (DIF = ART−MRT) for the Four Distances Between Digit Streams and Letter Streams

Distance (degrees)	*n*	*ART*	*SD*	*MRT*	*SD*	*DIF*	*SD*	*r*
Observer IS								
1.2	118	215	91	402	41	−187	83	0.41
1.9	119	232	78	416	53	−184	67	0.53
2.6	121	262	93	439	52	−176	79	0.53
3.3	121	320	114	481	79	−160	98	0.53
Observer MGA								
1.2	240	310	102	437	54	−127	98	0.33
1.9	238	308	113	443	66	−136	98	0.51
2.6	236	343	131	464	67	−120	108	0.57
3.3	240	357	136	487	101	−130	134	0.39
Observer CS								
1.9	117	327	102	396	43	−69	88	0.51
2.6	119	329	83	415	48	−86	80	0.34
3.3	121	369	116	443	65	−73	108	0.40
4.0	120	417	101	479	90	−62	99	0.47

Note. Reaction times and standard deviations (*SD*) are given in milliseconds. The number of trials (*n*) was determined by a random number generator that assigned equal probability of occurrence to each distance condition.

tribution was extremely narrow in Experiment 1. If observers had guessed the first numeral they reported, then such occasionally guessed numerals would have been randomly related to positions in the critical sequence and the ART distribution would have been broader. In fact, because numerals that had been presented in late positions of the critical sequence were never reported in the first response, such guessing never occurred.

Parallel processing. Observers might have processed both letter and numeral streams in parallel, avoiding the need for a shift of attention. We reject the notion of such parallel processing of both streams because observers virtually never reported numerals that had been presented before or simultaneously with the target and only rarely reported the numeral presented immediately after the target. The great majority of reports were of numerals presented 200 ms to 500 ms after the target. Had the observers been able to process both streams in parallel, they would have been able to report earlier occurring numerals. Indeed, observers were instructed to report the first numerals they could simultaneous with or after the target (not the first numeral after a shift of attention), and they were continuously reinforced for earliness of reported numerals throughout the experiment. But observers could not process the numeral stream earlier without missing the target in the letter stream. The task, not merely the instructions, demanded a shift of attention (Sperling & Reeves, 1980).

Reconstruction from visual memory. We can also reject the explanation in terms of visual memory, although this is a little more complex. Suppose that, contrary to instructions, observers were biased to report numerals that they knew had occurred late. Such a bias would mean that measurement of ART, the time from target onset to the onset of the first numeral reported, would be overestimated. In theory, the constancy of attention shift latency ASL(x) = ART(x) − MRT(x) over x, which we took as evidence for a fixed attention shift time, might occur if, in fact, ASL(x) did increase with x; however, observers deliberately compensated for longer attention shifts at larger eccentricities by decreasing such a bias for later occurring numerals. That is, in some conditions, observers selectively moved forward their reports from visual memory.

In fact, we have no reason to postulate such a bias and can reject it with the data at hand. If observers are to use a visual memory of the sequence of numerals to advantage and preferentially report numerals they know to occur later, they must know the order in which numerals actually occurred. However, an important finding in this work is that the order of the numerals is only poorly known (Reeves & Sperling, 1986). If observers attempt to use a memory with such poor order information, they will increase the variability of the distribution of the first report but only marginally influence its mean (i.e., the ART). An important finding relevant to this conjecture is that order scores in Experiment 2 did not vary with distance, so any effect on the mean would be independent of distance.[4]

Conclusion: Attention Shifts Are Quantal

Having rejected these possible methodological artifacts, we now take the constancy of attention shift latency ASL(x) with eccentricity x to mean that the latencies of attention shifts from periphery to fovea do not vary significantly with the distance to be shifted. Either attention shifts discretely, or it shifts in analog fashion but too fast to measure with the present technique. A shift time as slow as that suggested by Tsal (1983), 8 ms per degree, would have been easily detected and can be ruled out. If the obstacle condition had influenced ART or MRT, one might have concluded that attention was influenced by a position between the two attended positions. Because the obstacle had no effect, and because ASL(x) was constant, the most parsimonious interpretation is that attention shifted discretely between the two streams.

In the present study, attention moved from periphery to fovea, and peripheral acuity losses were estimated by the slowdown of MRTs to peripheral stimuli. In a spatially cued discrimination paradigm similar to that of Lyon (1987; described earlier), Cheal and Lyon (1989) measured attentional preparation for peripheral targets located at eccentricities of 2, 6, and 10 deg. Attention moved from fovea to the periphery. The authors compensated for peripheral acuity losses by increasing the size of peripheral targets. They found the same rate of attentional

---

[4] To illustrate, we computed the probability P_{co} of a correctly ordered pair (the probability that any pair of numerals was given in the response in the same order as it had been presented). These probabilities were, for the 2 deg, 4 deg, 7 deg, and obstacle conditions, respectively, 0.65, 0.65, 0.64, and 0.65 for DP and 0.71, 0.76, 0.72, and 0.75 for EW. The mean of 0.65 for DP accords with the results of other observers at this numeral rate; EW was slightly more accurate than previous observers (Reeves & Sperling, 1986).

preparation independent of the distance to the periphery. Their study and the present study explored two directions of attentional movement (to the fovea or to the periphery), two methods of compensation for peripheral acuity loss, and two different paradigms but came to precisely the same conclusion: The dynamics of attention shifts are independent of the distance traversed.

3. Episodic Theory of Attention: Mechanisms

Space–Time Separable Attentional Episodes (Saccadic Attention Shifts)

We briefly recapitulate the main concepts put forward in the introduction. Spatial attention consists of successive attentional states that we call episodes. In this respect, episodes are like the fixations between saccadic eye movements. Each episode is described by separable space and time functions. The spatial function defines the attentional state. The temporal transition function describes the changeover from one episode to the next. Because spatial attention is assumed to consist of a sequence of distinct spatial states, this is a quantal theory of spatiotemporal attention. That is, a priori, any spatial state might be possible, as might any temporal transition between successive states. In spite of their overlap, however, successive attentional states represent well-defined discrete events. According to this theory, the observer is always in an attentional state with respect to the visual environment, just as the eyes are always pointed in a direction in space. The difference is that, even in saccadic eye movements, the eyes are physically constrained to move over intermediate points between two fixations, whereas attention simply fades out at one location as it fades in at the next.

The episodic theory is applied to attention experiments by using a sequence of two or three consecutive attentional states to characterize the observer's behavior. Here we first elaborate the theory to illustrate how it applies to the attention gating paradigm, and then we recapitulate how the theory applies the three other paradigms analyzed in this article in terms of the specific sequences of attentional states. Finally, we show how it applies to a wider range of attentional phenomena.

Theory of Attention Gating in the Attention Gating Paradigm

In the attention gating experiments described earlier, the observer initially searched for a target that appeared a few degrees to one side of fixation and then attempted to recall the simultaneous numeral and the next three numerals from a stream at the point of fixation. The problem for the observer in the attention gating paradigm is that the number of stimulus items greatly exceeds the number that can be retained in visual short-term memory. Attention restricts the flow of items to memory by functioning like a gate to admit items only during a brief interval (Reeves & Sperling, 1986; Sperling & Reeves, 1980).

The attention gate. The attention gating theory involves two conceptually different kinds of inputs. The first kind of input is the to-be-processed information itself. Stimuli are represented by their luminance as a function of space and time. Sensory processing consists of converting the luminance representation of each stimulus item i into a vector of features $\mathbf{b}_i(t - t_i)$ whose magnitude (strength) is a function of time. For the present discussion, it is sufficient to consider just the one-dimensional strength $b_i(t - t_i)$ rather than the multidimensional stimulus vector $\mathbf{b}_i(t - t_i)$. After exposure termination, $b_i(t - t_i)$ decays exponentially (sensory persistence). Sensory persistence is terminated abruptly by the onset of a subsequent superimposed item. The location in time and space of sensory information that is admitted to further processing—visual short-term memory, in this instance—is determined by the attention gate. Opening of the gate is initiated by the second kind of input, the attentional cue. Closing of the gate is self-initiated. Information that passes the gate is integrated, stored in memory, perturbed by noise, and sequentially ordered for report (see later discussion and Figure 11).

Sequence of episodes in an attention gating task. Conceptualization of the attention gating paradigm in terms of a sequence of attentional episodes focuses on the control of the attentional gate (see Figure 3). The initial episode E_0 is initiated at the beginning of the trial at time t_0. The initial episode consists of directing spatial attention to a peripherally located character stream and a state of search for the target letter. The target—the attentional cue—occurs at time t_a. During episode E_0, the neural processing pathways are set to detect a target when it occurs (i.e., by suitable comparisons with memory representations), to automatically admit the target to short-term visual memory, and to initiate the next episode. In principle, entry of the target into short-term visual memory and initiation of a new attentional state are separate events. In the attention gating paradigm, however, target detection is assumed to immediately initiate a new episode E_1. Episode E_1 consists of the actual opening of the attention gate and the admission of items from the to-be-attended stream to memory. Because the rapid stream of to-be-attended items would quickly overflow memory capacity, a self-initiated gate closure begins at time t_2 very shortly after gate opening. Gate closure represents the onset of a third episode E_2 in which stimulus items are excluded from memory. The subsequent processes of extracting remembered items from memory and reproducing items for report involve later attentional episodes; these episodes are not under consideration here.

Time course of the attention gate. In the episodic theory, the time course of the attention gate is equivalent to the time course of the first episode (Figure 3b). The time course of any episode i is defined by Equation 2 of Figure 2; in particular, the time course of the first episode is simply

$$E_1(x, t) = f_1(x)[G_1(t - t_1) - G_2(t - t_2)], \qquad (12)$$

where $f_1(x_{\text{target}}) = 1$ represents an open gate at the target location x_{target}. Assume that G_i is the gamma function of Equation 1. Assume further that the duration $t_2 - t_1$ of the first episode is brief (e.g., less than τ, the time constant of G_i). Then

$$E_1(x_{\text{target}}, t) \approx \frac{t - t_1}{\tau} e^{(t-t_1)/\tau}. \qquad (13)$$

The functional form of Equation 13 exactly matches the dynamics of gate opening and closing (the gating attention func-

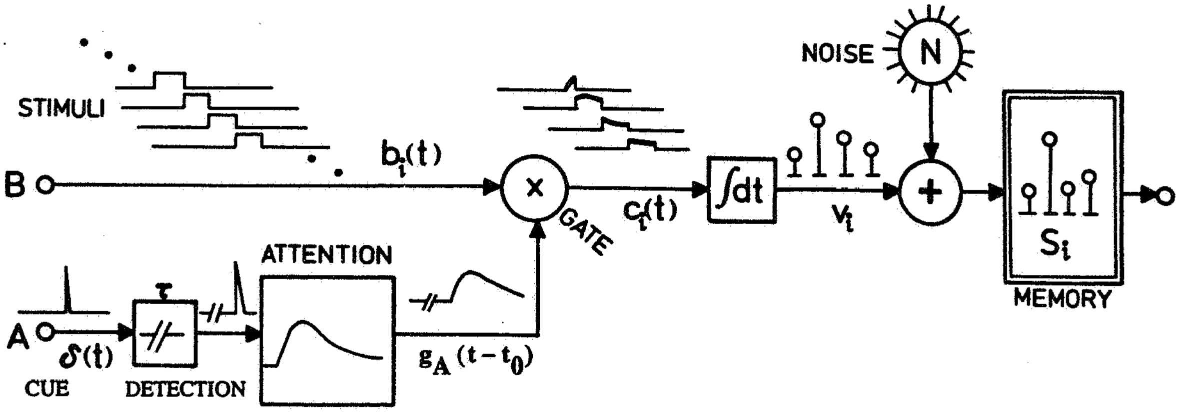


Figure 11. Attention gating model. When an attention cue $\delta(t)$ arrives at input A, it is detected and interpreted after a delay of τ. The detected cue activates the attention gate at time t_0, and it opens with time course $g_A(t - t_0)$. Stimuli $b_i(t)$ in the next-to-be-attended input, B, pass through the attention gate (X) with output strength $g_A(t - t_0) \times b_i(t)$. The instantaneous item strength is cumulated ($\int$). The strength of the stimulus ultimately stored in memory includes noise N. Stimuli from memory are output as responses in decreasing order of strength.

tion $g_A(t - t_1)$ proposed by Reeves and Sperling, 1986, to account for their large data set). These authors found a good functional description for an attentional gate; the present account derives this description from more elementary assumptions about consecutive episodes.

From attention gate to response. The episodic theory deals primarily with the dynamics of the attentional spotlight; a complete theory must also account for subsequent processing. The instantaneous flow of information through the attentional gate is given by the product of the attention function g_A multiplied by the strength of sensory information: $g_A(t-t_0)b_i(t-t_i)$. The cumulated, central strength v_i of item i is given by

$$v_i = \int_{t_0}^{\infty} g_A(t-t_0)b_i(t-t_i)dt. \qquad (14)$$

Random variation is represented by an internal noise source N that adds to cumulated strength. The net cumulated strengths $s_i = v_i + N$ determine the order of report, the strongest item being reported first and subsequent items being reported in decreasing order of strength. These relations are illustrated in Figure 11.

How to derive an empirical attention gating function $g_A(t - t_a)$. To relate the attention gating model to data, mean strength v_i is derived from the items reported by the observer. Similar estimates of v_i are obtained either from the frequency of reporting an item i or from the placements of i within the reports in which it occurs. Visual availability (for rapid display rates) is derived by assuming that an item is available during the interval beginning with its onset and ending with the onset of the following item. Once mean strength v_i and availability $b_i(t - t_i)$ are known, $g_A(t - t_0)$ can be derived from Equation 14.

Location versus process. In a stage performance, it is useful to distinguish between where in space a spotlight is pointing and the dramatic action that is occurring at that location. Similarly, in attending to a peripheral location, searching for a target, detecting a target, and recording the target in memory are actions that occur under the spotlight of attention; they do not necessarily change the location or configuration of the spotlight. In the experiments under consideration here, it has been assumed that movements of the spotlight of attention are associated with changes in the associated processes (the action under the spotlight). On the other hand, it seems reasonable to assume that several actions (such as searching, detecting, and recording a target in memory) may take place within the same episode.

Spatiotemporal representation of visual memory in the attention gating paradigm. In attention gating experiments, observers remember not only the numerals to which they shift attention but the target (attentional cue) whose detection initiated the shift. The dynamics of the automatic attentional episode that records the target itself were described by Weichselgartner and Sperling (1987). They observed two peaks in probability of reporting items as a function of time. The first was confined to the target itself and occasionally included a temporally adjacent item. The second peak, which was broader and centered around items 300 to 400 ms after the target, was identical to the attention gating function of Reeves and Sperling (1986).

Figure 12 shows a three-dimensional representation of memory contents in the attention gating paradigm. The horizontal axes are space x and time t. The height of the attention function represents the rate at which information is entered into memory from the corresponding x, t location. The two peaks overlie the x-locations of the two streams viewed by the observers and each peak is centered on the moment in time during which access to visual short-term memory is maximal at that location. Because a stream of noise characters that occurred between the two stimulus streams did not affect performance, attention (memory access) has been assigned the value zero at these intermediate locations.

Memory input intervals. The memory input interval is the

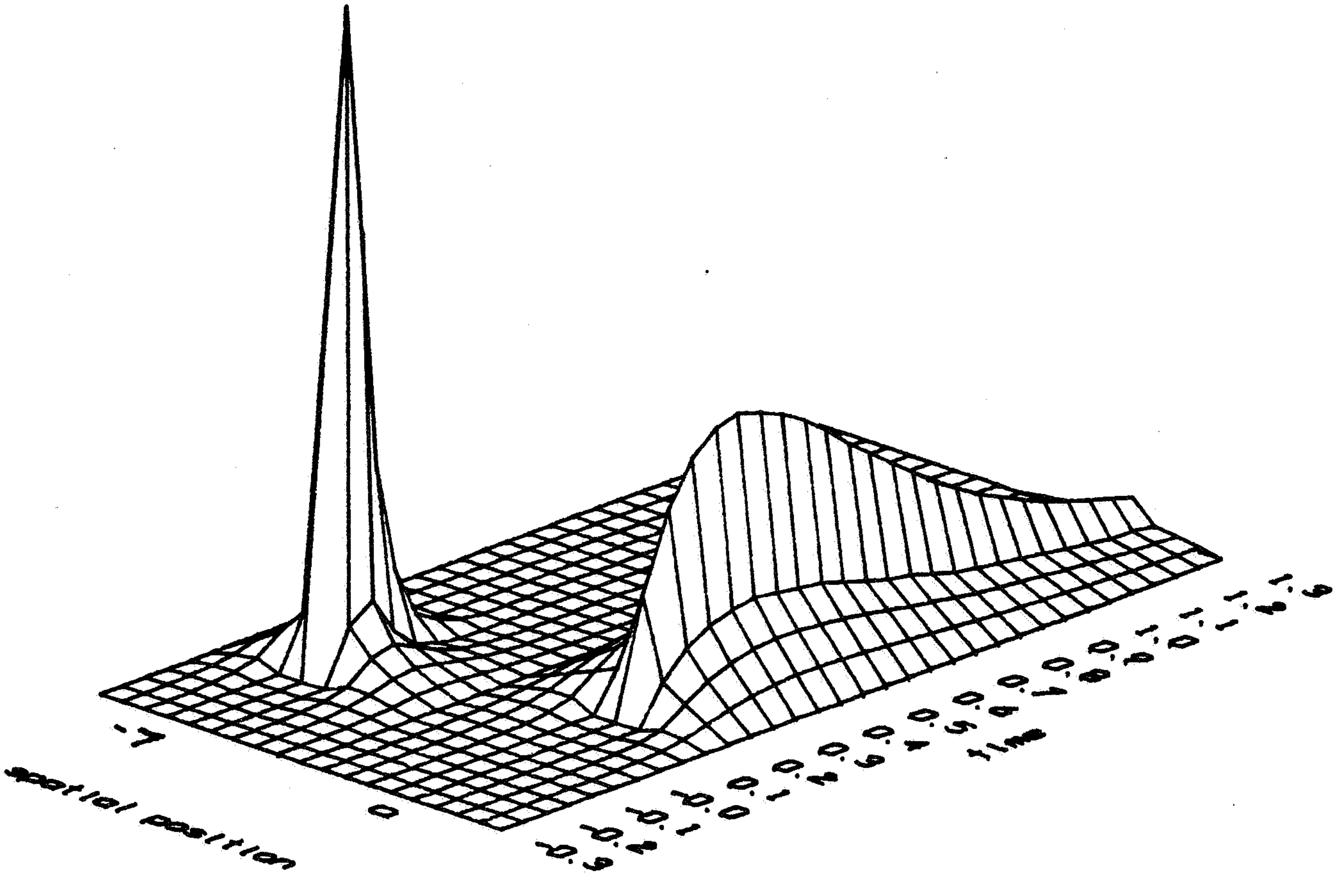


Figure 12. Three-dimensional space–time representation of the contents of memory in the attention reaction time paradigm. Height represents the amount of memory access as a function of spatial position x and time t. Initially, spatial attention is centered on the target stream (−7 deg, at left). Because the temporal attention gate that admits stimuli to memory is closed, the net initial value of attention is zero everywhere. On detection of a target, the target (and, occasionally, some closely adjacent items) are automatically entered into memory (the high peak at left). Then attention is shifted to the next to-be-attended stream (fixation, location 0). The long-duration peak at right represents the rise and fall of attention at location 0, the attention gating function $g_A(t-t_1)$.

period of time during which material from an event enters memory. In attention gating experiments, the input interval is synonymous with the transition-smoothed "duration" of the input episode $g_A(t-t_1)$. The input interval has been measured in numerous attention gating experiments and conditions, principally by Reeves (1977) but also by Reeves and Sperling (1986), Sperling and Reeves (1980), and Weichselgartner and Sperling (1987). A typical result (Sperling and Reeves, 1980) is embodied in the broad peak of Figure 12. Weichselgartner and Sperling (1987) measured the input interval associated with the memory of the target (the attention cue) as well as the to-be-attended stream. Their findings are represented in the narrow peak of Figure 12. They found the input interval of target memory to be much briefer than the input interval of any of the controlled attention gating experiments cited earlier. In a procedure similar to that of Weichselgartner and Sperling (performed for different reasons), Intraub (1985) produced quite similar results. It is obvious that the interval of target access to memory is much more compact than the interval of attention-controlled access, a matter that is reconsidered in the section on microprocesses of attention.

A Universal Attention Function?

Is it possible that the four spatiotemporal attentional functions considered here that describe (a) spatiotemporally cued memory gating $g_A(t-t_0)$, (b) spatiotemporally cued go/no-go reaction time facilitation $G_B(t-t_0)$, (c) spatiotemporally cued choice reaction time facilitation $G_C(t-t_0)$, and (d) spatiotemporally cued accuracy facilitation $G_D(t-t_0)$ all follow the same spatiotemporal time course after the onset t_0 of the alerting cue? Because the individual attentional episodes are space–time separable, one can study the possibly similar temporal components independently of the obviously different spatial components. Figure 13 illustrates the four attention transition functions that have been considered in this article, each scaled to the same

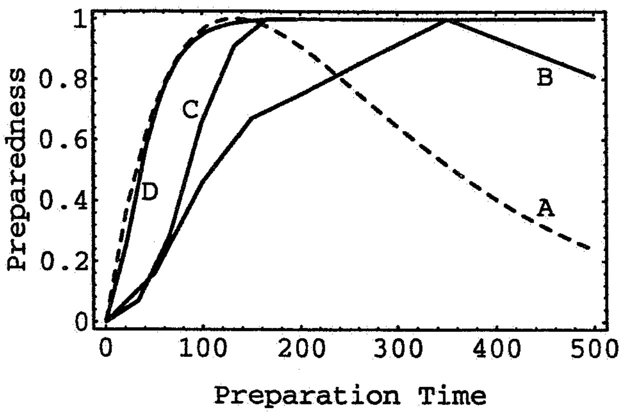


Figure 13. Comparison of four spatial, visual attentional preparation functions showing the fraction of maximal preparedness reached as a function of preparation time. A = attentional memory gating (Reeves & Sperling, 1986); B = go/no-go reaction times (Shulman et al., 1979); C = choice reaction times (Tsal, 1983); D = discrimination accuracy (Lyon, 1987). See text for details.

peak height. Even a cursory glance indicates that these functions are not identical. We consider the paradigms and the derived functions.

Attentional preparation in a discrimination task, G_D. Lyon's (1987) study of spatiotemporal facilitation after a cue in a discrimination task enlarged similar studies (e.g., Bashinski & Bacharach, 1980; Shaw & Shaw, 1977). The interpretation of discrimination experiments is quite straightforward. Initially, attention is disposed symmetrically around fixation E_0. After the cue, attention asymptotically becomes focused on the cued location E_1. The apparent movement of attention from fixation to the cued location results from the changeover from the initial attentional state to the final state rather than from a continuous attentional movement. The attentional transition function G_D derived from Lyon's (1987) experiment is a very fast-rising function, perhaps because the procedure requires neither motor preparation nor subsequent attentional turnoff, both of which might slow the rate of preparation.

Attention in spatially cued choice reaction times, G_C. In these procedures, a spatial location is cued; identification accuracy and choice speed are studied as a function of time and distance of the test stimulus from the cued location. Because of the restricted range of warning intervals studied in the particular experiment analyzed here (Tsal, 1983), only two attentional states are needed to describe the data: an initial state of readiness to perceive warning cues E_0 and the asymptotic state E_1 of peripheral attention (left or right) determined by the cue. This experiment followed and was followed by similar experiments (e.g., Shaw, 1982; see Shiffrin, 1988). Unfortunately, both spatial and temporal data must be obtained in the same experimental conditions to fully exploit the spatiotemporal theory, and such data still are unavailable.

With respect to the transition function $G_C(t - t_0)$ derived from Tsal's (1983) choice reaction time data (Figure 6A), the first 50 ms of G_C are extrapolated, not measured, because the relevant data were not collected. The slope of the measured component of Tsal's G_C for choice RT preparation is quite similar to Lyon's (1987) G_D preparation for a discrimination task. Both of these functions increase monotonically and are mostly complete within about 100 ms.

RT facilitation in go/no-go tasks, G_B. In reaction time facilitation by spatial cues (Shulman et al., 1979), there are three critical successive attentional episodes: an initial state of attention $E_0(x)$ before the spatial cue, a maximally prepared state for the location indicated by the spatial cue $E_1(t)$, and a more cautious state $E_2(t)$. Although the transition between the initial and the prepared states might have been described by a generic transition function, as in Figure 3 (and Equation 3), there were insufficient data to model the second transition. Therefore, these two transitions were combined in a single temporal function, $G_B(t - t_0)$ (Figure 4a). This alerting function is the temporal component of the space–time facilitation function derived from Shulman et al.'s data (Figure 1a) for go/no-go reaction times. It is the slowest rising of the four functions in Figure 13, and it is nonmonotonic. (The falloff beyond the maximum was explained earlier as caused by anticipation of a catch trial.) It is not clear why it should take longer to reach asymptotic preparation for go/no-go reaction times (G_D) than for choice reactions (G_C). The slow G_D probably represents the circumstances of the particular experiment (e.g., the particular distribution of foreperiods and catch trials or the involvement of relatively unpracticed observers) rather than a characteristic of the go/no-go paradigm.

Attention gating, g_A. Attention gating experiments, such as those of Reeves and Sperling (1986) and the present studies, yield empirically based attention gating functions $\hat{g}_A(t - t_1)$. Deriving the transition function $G(t - t_1)$ requires a few additional steps. It requires making assumptions both about the functional form of G and about the time interval $t_2 - t_1$ between successive episodes and then solving Equation 12 for G. The problem is that a shallow G and a brief $t_2 - t_1$ interval are almost indiscriminable from a steeper G and a longer $t_2 - t_1$ interval. The available data do not suffice to determine G_A. The temporal gating function shown in Figure 13 is g_A from Reeves and Sperling (1986, Table 1, observer AR). This function is typical of many data but is not a true transition function. However, it serves to indicate the kind of temporal dispersion that one finds in attention gating experiments.

The four attentional transition functions of Figure 13 indicate that there is more than a 2:1 range in the rate of attentional transitions (i.e., from less than 100 ms to more than 200 ms). Is this a result of the difference in paradigms or a result of other factors? Unfortunately, there is no obvious relation between the attentional tasks in the different paradigms and the slopes of the transition functions in Figure 13. Furthermore, within a paradigm, there may be large variations of observed attentional transitions resulting from amount of practice, stimulus presentation conditions, range of foreperiods tested, and so forth. Whether different tasks provoke different attention transition functions remains an open question.

Other Paradigms

Iconic memory and partial report experiments. Partial report procedures involve the retrieval of items from very short-term visual memory (iconic memory; Neisser, 1966). In a typ-

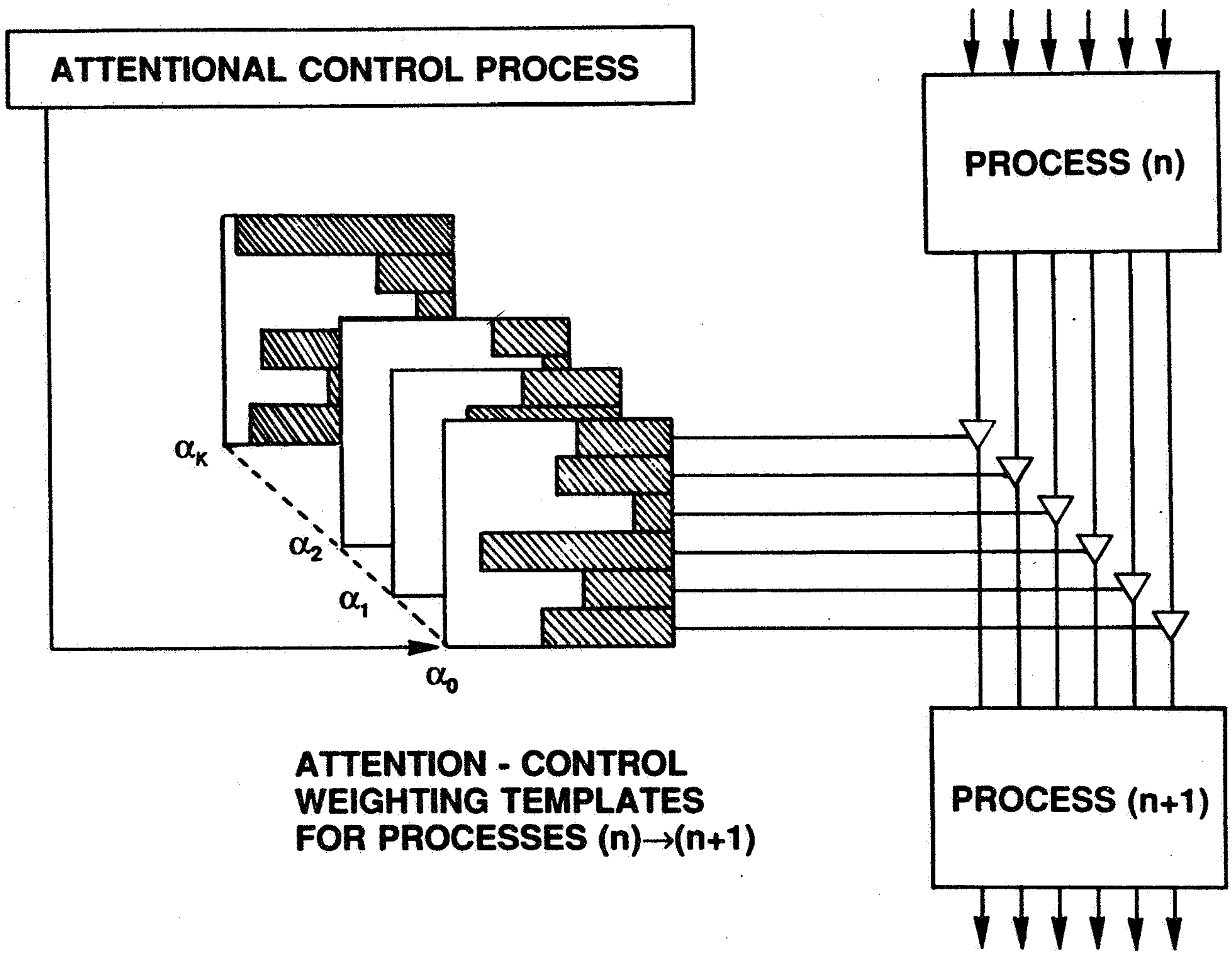


Figure 14. Attention in a neural network. Attentional weighting determines the amount of signal passed by each pathway i between processes (n) and (n + 1). The triangles represent amplifiers whose gain $\alpha(i)$ is controlled by the modulatory attentional process. There is an initial template of weights $\alpha_0(i)$ that may be a set of default weights. An attentional control process implements the changeover (at time t_0) from the initial template $\alpha_0(i)$ to another (previously learned) template $\alpha_k(i)$ with a time course $G(t - t_0)$ (Equation 1).

ical partial report experiment (Sperling, 1960), an array of letters is briefly exposed, and a cue indicating a particular row to be reported is presented before, during, or after the exposure. In terms of attentional processes, the tonal cue directs the attentional gate to the indicated row. Indeed, the results of partial report experiments can be modeled quite accurately by assuming just two attentional episodes (Gegenfurtner & Sperling, 1993). The initial episode represents a default state, with attention directed to the middle row of the display; the tonal cue directs attention to the indicated row. According to Gegenfurtner and Sperling's (1993) analysis, attentional switching occurred instantaneously, a result that is very difficult to reconcile with the present analyses. Earlier accounts, such as that of Rumelhart (1970), also made instantaneous-switching assumptions (see Gegenfurtner & Sperling, 1993, for a review). In these partial report procedures, it is difficult to completely disentangle attentional processes from the rapid decline with time of the legibility of iconic memory and the resulting time-dependent transfer rate of items from iconic memory to more durable storage. Nevertheless, taken at face value, Gegenfurtner and Sperling's (1993) finding would compel one to consider that different tasks might call forth quite different transition functions.

Localization of stimuli flashed during saccadic eye movements. When an extremely brief flash occurs during a saccadic eye movement, the apparent location of the flash may differ considerably from its actual physical location (O'Regan, 1984; Sperling & Speelman, 1965; see reviews by Matin, 1986; Sperling, 1990). Sperling (1990) proposed to account for these localization errors as follows. Under his experimental conditions, as in normal vision, each fixation corresponded to an attentional state. A flash that occurred during a saccade appeared to occur at a location in space that was not exactly appropriate to the position of the eye at the instant of the flash. The perceived location was predicted by weighting the location that would have been appropriate for the presaccadic fixation and the location appropriate for the postsaccadic fixation according to the same attention transition function used here to describe purely attentional movements in the absence of saccades (Equation 1). Localization errors occur because the attentional transition is somewhat slower than the eye movement, and, for some observers, the attentional transition is slightly advanced relative to the saccade. If attentional transitions mirrored saccadic eye movements perfectly then, according to this theory, there would be no localization errors.

During a saccade, the place in the environment to which a flash striking the fovea is referred changes continuously. However, as in the case of purely attentional transitions, the apparently continuous trajectory of the referred foveal location can be derived from a mixture of two states; continuous tracking is not required. A model that essentially incorporates a linear transition function can make quite accurate predictions of localization and mislocalization during four degree saccadic eye movements (Sperling, 1990; Sperling & Speelman, 1965).

Microprocesses of Attention

In the spotlight analogy, it is important to distinguish between the spatial location of the spotlight and the action that is occurring under the spotlight. In stage performances, action and location are closely coordinated, as indeed they are in attentional tasks. In the episodic theory, it has been assumed that the spatial aspects of visual attention are coordinated with the various tasks, such as searching for a target or admitting items to memory. At an algorithmic level, control of the spatial and computational components of attention may be quite similar. At a neural level, they are likely to be carried out by different neurons in different parts of the brain, and the distinction is critical.

At a neural level, we conceive of selective attention as a selective modification of the parameters of ongoing processes of perception, decision, and action. Attention and bias are labels used, sometimes interchangeably, for the processes that assign relative weights to information from parts of the visual field, that assign relative allocation weights to mental processing resources for alternative tasks, that assign relative weights to various response alternatives, that assign relative weights to decision alternatives, and so on (e.g., Mozer, 1988; Schneider & Detweiler, 1987). In this section, we describe the voluntary short-term modulation of processes of perception, recognition, decision, response selection, and so on (Figure 14). Weights themselves describe the state of attention $f(x)$; the process of instantiating a predetermined set of weights determines the temporal dynamics of attention.

Voluntary attention is, by definition, a top-down process. However, the control of weights can be bottom up (e.g., a sensory event can facilitate perception of similar subsequent events) or top down, or mixed. The attentional control process in Figure 14 is deliberately neutral as to the origin of the control (top down or bottom up).

In attention experiments, it is assumed that the templates themselves, the $f_i(x)$, are modifiable between trials or between sessions but not within a trial. During practice, the observer is presumed to learn (a) the optimal set of weights for the interval before the cue that informs about stimulus location, (b) the optimal weights contingent on each cue, and (c) how to quickly instantiate the cue-appropriate weights. Insofar as there are similar dynamics of weight assignment at different processing levels, one may expect similar dynamics for attentional processes, even when they operate at different processing levels (e.g., perceptual, decision, and response levels) and when several operate simultaneously.

There will be space–time separability in attention when the dynamics of the process $G(t)$ of instantiating a particular template of weights $f(x)$ are independent of the particular weights themselves. An important corollary is that the distinction between bias and sensitivity, which is central in the analysis of sensory selection versus decision processes (see review by Sperling & Dosher, 1986), does not imply a corresponding distinction in the kinds of underlying neural processes. The attentional modification of, for example, decision bias and of stimulus sensitivity parameters may involve similar neural processes and neural dynamics; the weights merely operate at different levels of processing.

In the tasks considered here, selective attentional enhancement of processing was contingent on spatiotemporal location. The paradigms themselves required a change of process coincident with the change in location (e.g., from searching or waiting for an attentional cue at the initial location to admitting items to memory or detecting a target at the cued location). In all cases, attentional selection was based on location or, in the case of the gating experiments, both location and time. Indeed, early attentional selection seems to be based primarily on spatial location and not on physical features such as color or size (Shih & Sperling, in press).

Conclusion

In conclusion, the theory of attentional episodes (each episode being described by space–time separable functions) is computationally attractive and gives an excellent account of existing data. It is consistent with plausible assumptions about neural processing.

References

Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics, 28,* 241–248.

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

Bouma, H. (1978). Visual search and reading: Eye movements and functional visual field: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 115–147). Hillsdale, NJ: Erlbaum.

Carlson, C. R., & Klopfenstein, R. W. (1985). Spatial frequency model for hyperacuity. *Journal of the Optical Society of America, Series A, 2,* 1747–1751.

Cavanagh, P. (1992). Attention-based motion perception. *Science, 257,* 1563–1565.

Cheal, M. L., & Lyon, D. (1989). Attention effects on form discrimination at different eccentricities. *Quarterly Journal of Experimental Psychology, 41A,* 719–746.

Cohn, T., & Lasley, D. (1974). Detectability of a luminance increment:

Effect of spatial uncertainty. *Journal of the Optical Society of America, 64,* 1715–1719.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149.

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

Eriksen, C. W., & Murphy, T. D. (1987). Movement of attention across the visual field: A critical look at the evidence. *Perception & Psychophysics, 42,* 299–305.

Flowers, J. H., & Wilcox, N. (1982). The effect of flanking context on visual classification: The joint contributions of interactions at different processing levels. *Perception & Psychophysics, 32,* 581–591.

Gegenfurtner, K., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 845–866.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: CBS.

Helmholtz, H. (1909). *Handbuch der physiologischen Optik* [Handbook of physiological optics] (Vol. 1, 3rd ed.). Leipzig, Germany: Voss.

Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics, 30,* 283–290.

Intraub, H. (1985). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 431–442.

Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology, 34,* 103–112.

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 VIII* (pp. 259–276). Hillsdale, NJ: Erlbaum.

Klemmer, E. T. (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology, 54,* 195–200.

LaBerge, D., & Brown, V. (1989). Theory and measurement of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

Laming, D. R. (1968). *Information theory of choice-reaction times.* New York: Academic Press.

Link, S. W. (1975). The relative judgment theory of two-choice response time. *Journal of Mathematical Psychology, 12,* 114–135.

Link, S., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 12,* 77–105.

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

Lyon, D. L. (1987). *How quickly can attention affect form perception?* (Technical Report AFHRL-TR-87-28). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Lyon, D. L. (1990). Large and rapid improvement in form discrimination accuracy following a location cue. *Acta Psychologica, 73,* 69–82.

Matin, L. (1986). Visual localization and eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (Vol. 1, pp. 1–45). New York: Wiley.

Melchner, M. J., & Sperling, G. (1980). *VEX: A computer system for real-time Vision EXperiments* (Bell Telephone Laboratories Technical Memorandum). Murray Hill, NJ: AT&T Bell Laboratories.

Mozer, M. (1988). *A connectionist model of selective attention in visual perception* (Technical Report CRG-TR-88-4). Departments of Psychology and Computer Science. University of Toronto, Toronto, Ontario, Canada.

Murphy, T. D., & Eriksen, C. W. (1987). Temporal changes in the distribution of attention in the visual field in response to precues. *Perception & Psychophysics, 42,* 576–586.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research, 15,* 1039–1042.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86,* 214–255.

Neisser, U. (1966). *Cognitive psychology.* New York: Appleton-Century-Crofts.

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

Nickerson, R. S. (1967). Expectancy, waiting time and the psychological refractory period. *Acta Psychologica, 27,* 23–34.

Nickerson, R. S., & Burnham, D. (1969). Response times with nonaging foreperiods. *Journal of Experimental Psychology, 79,* 452–457.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.

O'Regan, J. K. (1984). Retinal versus extraretinal influences in flash localization during saccadic eye movements in the presence of a visible background. *Perception & Psychophysics, 36,* 1–14.

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

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. 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.

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

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206.

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

Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics, 35,* 393–399.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7,* 191–216.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. Bower (Ed.), *The psychology of learning and memory* (Vol. 21, pp. 54–118). New York: Academic Press.

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

Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology, 14,* 353–408.

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.

Shepard, R. N. (1975). Form, formation, and transformation of internal representations. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 87–122). Hillsdale, NJ: Erlbaum.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shiffrin, R. M. (1975). The locus and role of attention in memory systems. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 168–193). New York: Academic Press.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology,* pp. 739–811 (2nd ed.). New York: Wiley.

Shiffrin, R. M., McKay, D. P., & Shaffer, W. O. (1976). Attending to forty-nine spatial positions at once. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 14–22.

Shih, S., & Sperling, G. (in press). Is there feature-based stimulus selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance.*

Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 522–526.

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

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (11, Whole No. 498), 1–29.

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

Sperling, G. (1964). Linear theory and the psychophysics of flicker. *Documenta Ophthalmologica, 18,* 3–15.

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. (1990). Comparison of perception in the moving and stationary eye. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 307–351). Amsterdam: Elsevier Biomedical.

Sperling, G., & Dosher, 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. 1–85). New York: Wiley.

Sperling, G., & Melchner, M. J. (1976). Estimating item and order information. *Journal of Mathematical Psychology, 2,* 192–213.

Sperling, G., & Melchner, M. J. (1978). Visual search, visual attention, and the attention operating characteristic. In J. Requin (Ed.), *Attention and performance VII* (pp. 675–686). Hillsdale, NJ: Erlbaum.

Sperling, G., & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 347–360). Hillsdale, NJ: Erlbaum.

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

Sperling, G., & Speelman, R. G. (1965). Visual spatial localization during object motion, apparent object motion, and image motion produced by eye movements. *Journal of the Optical Society of America, 55,* 1576.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 523–530.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238,* 778–780.

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

Welford, A. T. (Ed.). (1980). *Reaction time.* New York: Academic Press.

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

Wundt, W. (1924). *An introduction to psychology* (R. Pinter, Trans.). London: George Allen & Unwin. (Original work published 1896)