GLOBAL AND LOCAL COVERT VISUAL ATTENTION: EVIDENCE FROM A BAYESIAN HIDDEN MARKOV MODEL

JOHN LIECHTY

PENNSYLVANIA STATE UNIVERSITY

RIK PIETERS

UNIVERSITY OF TILBURG

MICHEL WEDEL

UNIVERSITY OF MICHIGAN

Psychological, psychophysical and physiological research indicates that people switch between two covert attention states, local and global attention, while visually exploring complex scenes. The focus in the local attention state is on specific aspects and details of the scene, and on examining its content with greater visual detail. The focus in the global attention state is on exploring the informative and perceptually salient areas of the scene, and possibly on integrating the information contained therein. The existence of these two visual attention states, their relative prevalence and sequence in time has remained empirically untested to date. To fill this gap, we develop a psychometric model of visual covert attention that extends recent work on hidden Markov models, and we test it using eye-movement data. The model aims to describe the observed time series of saccades typically collected in eye-movement research by assuming a latent Markov process, indicative of the brain switching between global and local covert attention. We allow subjects to be in either state while exploring a stimulus visually, and to switch between them an arbitrary number of times. We relax the no-memory-property of the Markov chain. The model that we develop is estimated with MCMC methodology and calibrated on eye-movement data collected in a study of consumers' attention to print advertisements in magazines.

Key words: change point models, reversible jump Markov Chain Monte Carlo, variable dimension time-series models, visual attention, covert attention, overt attention, eye-movements, saccades.

1. Visual Attention

Overt visual attention and covert visual attention to stimuli are closely related. Overt visual attention, eye-movements as they are observed in subjects who are exposed to spatial stimuli, and covert visual attention, the mental "spotlight", are under partially overlapping neuronal control, as revealed by a large number of psychophysical and physiological studies, reviewed by Itti and Koch (2001). Shifts of visual attention are changes in the spatial location to which people attend. In naturally occurring situations, eye movements present the most salient form of shifting attention, thereby coupling patterns of overt and covert attention (Findlay & Gilchrist, 1998; Palmer 1999; Wright & Ward 1998). Thus, eye-movements provide reliable information about covert visual attention in such situations, and they have been extensively used as indicators of attention in areas as diverse as reading, scene perception, target search, car driving and advertising processing (e.g., Henderson & Hollingworth, 1999; Just & Carpenter 1987; Rayner, 1998, Underwood & Radach, 1998; Wedel & Pieters 2000).

Apart from several smaller corrective movements, eye movements to stationary stimuli are composed of saccades and fixations. Saccades are rapid, ballistic changes in eye position, during which vision is essentially suppressed. Fixations are the pauses between saccades during which

the eye position is relatively immobile, and during which information extraction takes place. The center for the execution of eye-movements is located in the superior colliculus (SC) in the human brain.

Regular sequences of fixations, called scanpaths, occur during exposure to a visual stimulus. There is evidence that such visual scanning takes place through the interplay of two neural processes, characterized as "winner-takes-all", and "inhibition of return" (e.g., Itti, Koch & Niebur 1998; Klein 1988; Laberge 1998). Both of these operate on stimulus saliency, which seems to be encoded in multiple regions of the visual system, in particular in the SC, and in the posterior parietal cortex (PPC; Itti & Koch, 2001). Winner-takes-all neural activity ensures that in case of competition for attention, the location of highest salience receives all attention. After processing, that attended location is then transiently inhibited, thus preventing subjects from continually rechecking it. The attention spotlight then focuses on the next salient location, through winner-takes-all again. This process, applied repeatedly, generates a scanpath of fixations and saccades across the stimulus.

1.1. Local and Global Visual Attention

Whereas previously scanpaths had been considered in a deterministic manner, Ellis and Smith (1985) proposed that scanpaths are best described by a stochastic process, where the position of a fixation depends on the position of the previous fixation, according to a first-order Markov process (see also Viviani 1990). Pieters, Rosbergen, and Wedel (1999) recently empirically confirmed scanpaths to follow a first-order Markov process. This Markov process presupposes a uniform scanpath during exposure to a stimulus, that is, a single-state of covert visual attention.

Yet, there is reason to believe that people switch between two different covert states of visual attention, while exploring visually complex stimuli, which should lead to distinct scanpaths of eye-movements (e.g., Antes 1974; Groner 1988; Lévy-Schoen 1981; Pomplun, 1998; Zangemeister, Sherman & Stark 1995). We postulate that these two covert states concern respectively local and global visual attention. In local visual attention, stimuli are explored in detail by extracting information from specific and adjacent locations, and it is characterized by shorter saccades. In global visual attention, stimuli are explored to identify locations to extract information, and possibly to integrate the information from various locations in the stimulus. It is characterized by longer saccades. It has been suggested that the global attention state is prevalent initially to grasp the general meaning of a stimulus, while the local attention state may prevail later on to fill in details and focus on less informative parts (Antes 1974).

Consistent with this view of two covert attention states, LaBerge (1998) distinguishes orient and resolving mechanisms of attention in target search from multi-element displays. In orienting, people scan a stimulus in search for spatially distributed information, while in resolving they extract information from a specific location, and keep it separate from nearby or overlapping information coming from distractors. Related to this, Posner (e.g., Posner, 1980; Posner & Cohen 1984) distinguishes stages of attention engagement/disengagement and of attention shift. These covert stages could be reflected in a single fixation and saccade, but could also express themselves in the larger patterns of fixations and saccades that comprise scanpaths. In a similar vein, Henderson and Hollingworth (1998) propose framework of scene viewing comprising a state in which attention shifts across the scene to particular regions of interest, and a state in which information is extracted from a specific region in the scene. Information acquisition in the latter state is optimised by re-fixations in the region of interest, which is typically reflected in smaller saccades.

In further support of our reasoning, two brain areas appear to be strongly involved in the control of visual attention (see the review of Itti & Koch, 2001). The posterior parietal cortex (PPC) globally directs visual attention towards regions of interest in the scene; the inferotemporal cortex (ITC) is involved in local exploration, recognition and identification of objects in the
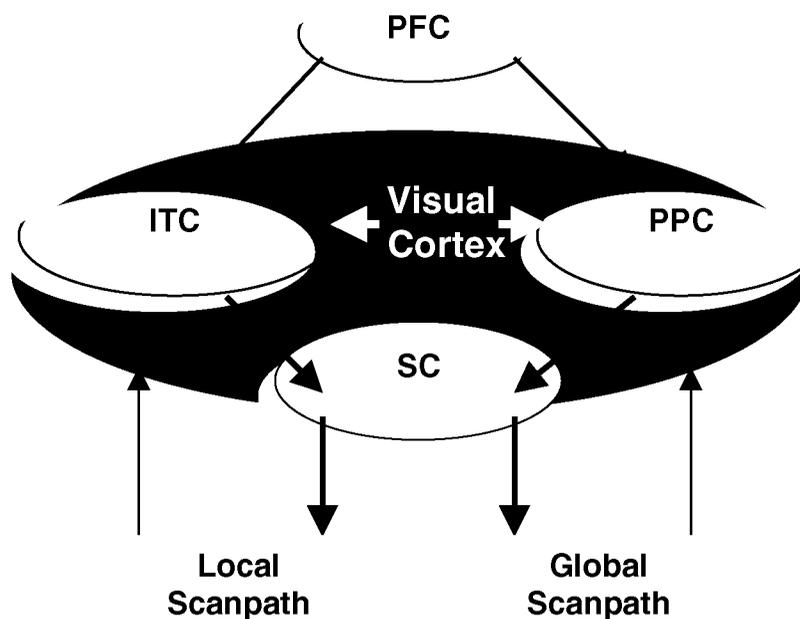
FIGURE 1.
Neural control of overt and covert visual attention (adapted from Itti & Koch, 2001).

scene (see Figure 1). In the ITC-controlled local attention state, the focus is on specific aspects and details of the stimulus, and on examining its content with greater visual detail. In the PPC-controlled global attention state, the focus is on exploring the informative and perceptually salient areas of the stimulus, and possibly on integrating the information contained therein. That is, bottom-up control of visual attention by the PPC would be characterized by global scanpaths, consisting of longer saccades, since the eyes are directed at salient features of or objects in the scene for further exploration. Bottom-up control of visual attention by the ITC, on the other hand, would be characterized by local scanpaths, consisting of shorter saccades, serving to process objects in detail.

Since understanding of heterogeneous stimuli involves both local and global attention, the PPC and ITC need to interact intensively. Such interaction has been shown to be under the control of the prefrontal cortex (PFC), which controls the switching of attention back and forth between local and global attention states through feedback mechanisms (see Figure 1). The PFC, in addition, is also responsible for planning of the execution of eye-movements, the center for which is located in the SC. Thus, the PFC switches control between the PPC and the ITC continuously, and at the same time plans the eye-movements, executed by the SC. Thus, activity in the PPC and ITC is likely to be directly reflected in the eye-movements and scanpaths, and the difference in these attention control systems is likely to be reflected in differences among the scanpaths.

In sum, there is reason to believe that there are (at least) two states of covert visual attention, local and global, with qualitatively different information processing functions, which are expressed in distinct patterns of eye-movements. Dominant presence of both attention states is less likely to occur for relatively homogenous stimuli, as typically used in reading tasks, where the local attention state may prevail (McConkie, 1983; Rayner, 1998). In contrast, visually heterogeneous stimuli such as complex scenes (Yarbus 1967), or stimuli containing multimode information such as instructions with text and pictures (Stolk, Boon & Smulders 1993), and

print advertisements (Pieters, Rosbergen & Wedel, 1999; Rayner, Rotello, Stewart, Keir & Duffy 2001) may give rise to more extensive switching between global and local visual attention states.

Although there is ample theoretical support for the postulated states of local and global visual attention, their existence and relative prevalence over the time course of stimulus exposure has remained empirically untested. Identification of these two attention states may have been hampered by both the absence of sufficiently detailed eye-movement data across larger numbers of subjects and sufficiently heterogeneous stimuli, and the mathematical complexity of representing such a process of switching among different unobserved states of visual attention.

Here, we develop and test a psychometric model that aims at dealing with these issues. It is assumed that the observed eye-movement pattern is an indicator of, that is, imperfectly reflects, the two unobserved attention states. In attention research one is primarily interested in inference on the covert attention process, for example on the time subjects spend in the local and global attention states and the transition probabilities between them. Statistics on these unobserved states cannot be directly derived from the data or simple models that ignore the existence of such latent attention states. Our model serves to identify the covert visual attention process and enables its statistical description. It describes the observed time series of saccades typically collected in eye-movement research and uses a Markov representation of the overt visual attention process (Pieters, Rosbergen & Wedel, 1999). However, rather than using a single Markov process, we assume two such processes, indicative of the specific covert attention state that a subject is in at a particular point in time. We relax the no-memory property of the Markov chain and allow the time of a particular scanpath to depend on the time spent in the covert attention state, which is important substantively (Egeth & Yantis, 1997; Viviani 1990), since it reflects the attention feedback control mechanism exerted by the PFC. We model subjects' switching between the global and local states controlled by the PPC and ITC respectively, where we take the observed pattern of saccades as indicative of the subject's latent attention states. We allow for an arbitrary number and order of switches among the covert attention states for each subject. The model that we develop is estimated with MCMC methodology and calibrated on eye-movement data collected in a study of consumers' attention to print advertisements in magazines. In the next section we derive our model of covert visual attention states, and describe how it is estimated. Next, we offer a test of the model using synthetic data, and proceed with a description of the data collection and findings of the main empirical illustration.

2. A Model of Local and Global Visual Attention

2.1. Likelihood and Priors

In order to enable the identification of patterns of visual attention to stimuli, we extend the recent literature on hidden Markov models (Hodgson & Green 1999; Liechty & Roberts 2001; Robert, Ryden & Titterington 2000) and develop a model of covert global and local spatial visual attention, based on the theory outlined in the previous section. Prior research has employed Markov and semi-Markov models for modeling eye-movements across a stimulus (cf. Engbert & Kliegl 2001; Hacisalihzade, Stark, & Allen 1992; Rimey & Brown 1991), while Salvucci and Anderson (2001) use hidden Markov models for the coding of eye-movement protocols.

Rather than equating characteristics of eye movements (fixations, saccades) directly with visual attention, we consider them as indicators of the patterns of covert visual attention that we intend to uncover. In order to do that, we consider the overt attention process, X, consisting of the sequence of eye fixations across a stimulus, observed for a sample of subjects to be a continuous-time Markov chain, which allows us to obtain insights into the nature of the spatial dependence of eye fixations on a stimulus.

We assume that overt attention is driven by a covert attention process, denoted as D, which attains two states representing the states of visual attention, global or local, that a subject may be in at any point in time. The covert process cannot be directly identified from the data, since

global eye jumps may occur by chance in the local covert attention state and visa versa. The covert process D is a continuous-time Markov or semi-Markov chain.

We begin our description of the model by assuming for simplicity that we have observed a sequence of eye movements observed for a single subject, as s/he views a single stimulus. We assume that the data is summarized in terms of a sequence of fixations on a grid overlaid on top of the stimulus, where the S is a countable collection of locations j on the grid:

$$S \equiv \{j; (x_j, y_j) : j = 1, \ldots, J\}.$$

The pair (x_j, y_j) defines j in terms of the x- and y-coordinates on the lattice. Such a grid facilitates the analysis of the data through aggregation of the fixations, and enables us to define local and global saccades. To that end we define two sets of indices on the lattice, those that define neighboring (local) locations on the grid, respectively those that we call global:

$$L \equiv \{j, l : |x_j - x_l| \leq 1 \wedge |y_j - y_l| \leq 1\},$$

$$G \equiv \{j, l : |x_j - x_l| > 1 \vee |y_j - y_l| > 1\},$$

where $S = L \cap G$, so that fixations are classified as either local or global, but not both. Dealing with a grid rather than with the exact eye-fixation locations introduces measurement error—the chosen grid size affects the definition of local versus global observable saccades—that is accommodated by the stochastic nature of the model specified below. We are primarily interested in inferences on the unobserved attention process and its latent states, for which the observed fixations on the grid elements are used as (imperfect) indicators. In choosing the grid size in the empirical application below, we intend to balance the detail of the information obtained on the indicators of the latent states against data sparseness (e.g., Ellis & Smith 1985).

We consider the observed fixations

$$X_t = \{X_t(j) : j \in S; 0 \leq t \leq T\}$$

to be realizations of a continuous-time Markov chain indexed by t (we omit the dependence of X on the lattice S in the notation in the sequel). The observed chain has waiting times that are exponentially distributed. In line with previous suggestions that saccades depend on the time spent in the attention state (e.g., Wedel & Pieters 2000; Wolfe, 1998), we have the transition matrix of the observed chain X depend on the hidden continuous-time Markov chain, D_t. That hidden chain is characterized in terms of states $(k_0, k_1, \ldots k_T)$ and associated waiting times $(\tau_0, \tau_1, \ldots, \tau_T)$ indexed by $t = 1, \ldots, T$. We consider a two-state $(k = 1, 2)$ hidden Markov chain, representing the local versus global covert attention states. Conditional upon a jump occurring, the probability that D changes states is defined by P_D.

Each state, k, of the hidden Markov chain, D, has an associated transition matrix, $P_k = \{p_{kj\ell}\}$, that characterizes the observed Markov chain describing the saccades. The indices j and ℓ refer to the elements of the lattice S. The diagonal elements of the transition matrices represent re-fixations in the same cell of the grid, and identify the informative or perceptually salient areas of the stimuli for each attention state (Henderson & Hollingworth, 1998, 1999). The off-diagonal elements represent saccades between the grid cells and indicate how, given a latent attention state, the eye shifts between those informative or perceptually salient areas. Since the fixations may be spatially dependent beyond what is accommodated by the transition matrices P_k, spatial dependencies in the fixations are accommodated through the specification of the prior distribution, as will be detailed below.

Explicitly modeling the termination of the observed eye movements across the stimulus is important from a substantive point of view, since it provides insights not only into when subjects terminate attention to a particular stimulus, but also into the covert attention state they are in when terminating. This tells us whether subjects are more likely to end scanning the stimulus after attending to stimulus details deemed sufficiently informative (local), or after exploring the

stimulus for informative areas (global). For the data that we consider, the state space of the observed process, X, is given by the elements of the grid plus an 'exit' state that occurs when a subject stops exposure to the stimulus. This exit state is an absorbing state that naturally accommodates finite length or censored Markov chain data. We represent the probability of jumping to the exit state as q_{kj} and formulate the probability of jumping conditional on not exiting. Thus, $P_k = p_{kj\ell}$, is the probability that the eye jumps from j to ℓ given that D is in state k and X jumps, but not to the absorbing state. We assume that once X enters the absorbing state D no longer changes its state.

In the eye-movement data in the empirical application below we have a collection of observed processes, $\{X_{na}\}$, which for our purposes are indexed by n for a particular subject and by a for a particular stimulus. The dynamics of each observed process X_{na} depends on it own hidden Markov chain D_{na}. We assume a priori that the starting probabilities of the Markov chain, ν_D, and the transition probabilities P_D given the hidden chain D are the same for all subjects and for all stimuli. As motivated extensively earlier, we assume a hidden process with two covert attention states, a global state and a local state, but for purposes of model comparison we will also later estimate a model with a single latent state. Although the times that subjects spend in these states may differ, the switches between the local and global states themselves are assumed timeless, which is in accordance with quantal theories of spatial attention (Sperling & Weichselgartner 1995). The prior for D_{na} is:

$$f(D_{na}|P_D, \lambda_D, \nu_D) = \nu_D(D_{na0}) \left(\prod_{k \neq \ell} p_{Dk\ell}^{N_{na}Dk\ell} \right) \left(\prod_k \lambda_{Dk}^{M_{Dnaj}} \right)$$

$$\exp\left[-\sum_k \left(\int_0^T \lambda_{Dk} I\{D_{nas} = k\} ds \right) \right], \tag{1}$$

where $N_{akDj\ell}$ (unobserved) is the number of times that D jumps from state j to ℓ, where ν_D is the density of the starting values of X, where M_{Dnaj} is the number of times X jumps to state j, where T is the time at which X jumps to the absorbing state, where $I\{\cdot\}$ is the indicator function and where λ_{Dk} is the intensity of the k-th exponential waiting time distribution. We assume that ν_D follows a prior Dirichlet density and that λ_{Dk} follows a prior Gamma density.

Each stimulus has its own starting probabilities, denoted as ν_a. We assume a priori that ν_a follows a Dirichlet density. The dynamics of the observed process are described by the intensities λ_{akj}, the probabilities of exiting q_{akj} and the transition probabilities, P_k. We assume that λ_{akj} follows a prior Gamma density, and that q_{akj} follows a prior Beta density:

$$\lambda_{akj} \sim \text{Gamma}\,(g_k, h_k),$$

$$q_{akj} \sim \text{Beta}\,(d_k, b_k). \tag{2}$$

We assume that the parameters in (2) each follow a prior exponential density.

We also assume a priori that each row of the jump chain transition matrices follow a constrained Dirichlet density, so that the density for the j-th row of P_k is given by

$$f(p_{kj}) = \frac{\Gamma\left(\sum_\ell \alpha_{kj\ell}\right)}{\prod_\ell \Gamma(\alpha_{kj\ell})} \prod_\ell p_{kj\ell}^{\alpha_{kj\ell}-1}, \tag{3}$$

where the restriction $\sum_\ell p_{kj\ell} = 1$ holds. We specify the parameters of this prior to depend on whether the cells j and ℓ are adjacent to each other. Thus we use the spatial adjacency pattern of the cells as prior information. We now parameterize α_{akj} as:

$$\alpha_{akj\ell} = \begin{cases} \beta_{kL}, & \text{if } j, \ell \text{ is a local move} \\ \beta_{kG}, & \text{if } j, \ell \text{ is a global move}. \end{cases} \tag{4}$$

In order to identify the states of the latent chain D, we place restrictions on the transition matrix. Because we are interested in distinguishing between periods in which local moves (moves to neighboring cells in the grid) are preferred to global moves (nonlocal moves) and periods in which global moves are preferred to local moves, we consider a two-state ($k = 1, 2$) hidden Markov chain where the two states are identified by assuming that in State 1 the probability of a local move is larger than that of a global move, while in State 2: the reverse holds. That restriction to identify the parameters of the latent chains is imposed as

$$\beta_{1L} > \frac{n_G}{n_L}\beta_{1G} \quad \text{and} \quad \beta_{2L} < \frac{n_G}{n_L}\beta_{2G}, \tag{5}$$

where n_L is the number of adjacent (local) cells and n_G is the number of nonadjacent (global) cells for a cell that is not on the edge of the grid. This restriction is implemented, because prior eye movement research has shown that fixations on a cell at the edges of the stimulus as a whole are more rare (Pieters, Rosbergen & Wedel, 1999). To complete the model specification we assume that each β in (5) follows a prior exponential density.

In our model the waiting times of X are not necessarily independent of D. If they are, we have a continuous time Markov model that arises as a special case. In describing the eye-movement data, the time spent in one of the cells of the lattice, before jumping may depend on the realization of the hidden Markov chain. Hence, it is reasonable to have the waiting times of X conditional on D, denoted as w, come from a piecewise-continuous exponential density, with constant intensities within time intervals where D is constant. If the waiting times of X come from such a density, then the observed process is no longer memory-less (i.e., the hazard function is no longer constant) and X can be regarded as a semi-Markov process; see Ross (1983). The intuition behind this specification of the semi-Markov process is that to predict an observed saccade given a covert attention state, the entire time spent in that particular covert state is relevant, and determined by attention feedback from the PPC and ITC to the PFC.

The likelihood of the model can now be formulated as

$$f(X|-) = \prod_{na}\prod_{k}(\nu_{ak}(X_0))^{I\{D_{na0}=k\}}\left(\prod_{kj\ell}(p_{akj\ell}(1-q_{akj}))^{N_{nakj\ell}}\right)$$

$$\times \sum_{kj}q_{akj}I\{D_{naT}=k \wedge X_{naT}=j\}\left(\prod_{\ell}f(w_{na\ell}|D_{na},\lambda_a)\right), \tag{6}$$

where $f(w_\ell|D,\lambda)$ is the density of the ℓ-th waiting time of X (now suppressing subscripts a and n for convenience), which is given by

$$f(w_\ell|D,\lambda) = \zeta_\ell\sum_{m=0}^{\infty}\phi_{\ell m}\exp\{-\lambda_{X_{t_\ell}D_{w_\ell}}(w_\ell-\tau_{\ell+m})\}I\{\tau_{\ell+m}<w_\ell\leq\tau_{\ell+m+1}\}. \tag{7}$$

Here $\lambda_{X_{t_\ell}D_{w_\ell}}$ is the intensity associated with the state of X when the ℓ-th jump time of X began and with the state of D at time w_ℓ, and $\tau_{\ell+m}$ is the m-th time after t_ℓ—the ℓ-th jump time of X—that D changed states, with $\tau_{\ell+0}=0$. In addition, $\phi_{\ell 0}=1$ and $\phi_{\ell m}$ is defined recursively by

$$\phi_{\ell m+1} = \phi_{\ell m}\exp\left\{-\lambda_{X_{t_\ell}D_{\tau_m}}(\tau_{m+1}-\tau_m)\right\}. \tag{8}$$

The normalizing constant of this density complete the model description, and it is given by,

$$\zeta_\ell^{-1} = \sum_m\left(\frac{\phi_{\ell m}}{\lambda_{X_{t_\ell}D_{\tau_m}}}(1-\exp\{-\lambda_{X_{t_\ell}D_{\tau_m}}(\tau_{\ell+m+1}-\tau_m)\})\right). \tag{9}$$

To summarize, the important features of the model are that it describes overt visual attention, that is, eye fixation patterns on a grid overlaid on a stimulus and their termination, as an indicator of the covert attention process of primary interest, comprising of a global and a local state, where the time spent in one of the cells of the grid may depend on the overt realization of the attention process. The proposed model provides key measures of covert visual attention: the probability of starting and of terminating exposure to the stimulus in a particular attention state, the probability of jumping to a certain state given the current attention state that people are in, the mean duration in each attention state, and the mean fixation duration per attention state.

2.2. Markov Chain Monte Carlo Algorithm

We use a standard Markov Chain Monte Carlo (MCMC) algorithm, with Hasting Metropolis (HM) subalgorithms, to draw from the posterior distributions generated by our model specification; see Gilks, Richardson and Spiegelhalter (1996) for a general discussion of MCMC methods and the HM algorithm. The full conditional posterior densities for the Markov chain and semi-Markov version of these models are identical except for the full conditional density of the hidden Markov chain D and the density of the intensities of the waiting times of X. The full conditional posterior densities for rows of the transition matrices, P_{ak}, and for v_D, the initial distributions of the hidden and observed Markov Chains are Dirichlet densities, those of β_{kL} and β_{kG} are proportional to products of Dirichlet densities. The full conditional density of q_{akj} is a Beta density. The full conditional densities of the intensities of D, λ_{Dk}, and of X, λ_{akj} are Gamma densities. The full conditional distribution of g_k is a Gamma density, but that of h_k is nonstandard. Those of d_k and b_k are also nonstandard. The full conditional densities of the intensities of X, when X is a semi-Markov process, and of D, regardless of whether X is semi-Markov or Markov, are nonstandard. The details of these full conditional posterior distributions are not provided here to save space.

We use the reversible jump HM algorithms proposed by Liechty and Roberts (2001) to generate samples of the hidden Markov chain D. Only a brief summary of these algorithms is given here as they are fully described in Liechty and Roberts (2001). The difference between their algorithm and ours is based on the different likelihood functions. Samples for the parameters that have conjugate full conditional densities were generated using the Gibbs sampler, and samples for the remaining parameters were generated using a random walk HM algorithm. The algorithms used to update D are reversible jump algorithms, where each hidden Markov chain is parameterized using a variable number of parameters. The number of parameters depends on the number of times that the Markov chain changes state. One advantage of the HM algorithm, as pointed out in Green (1995), is that it can jump between models with different dimensions; for example, between models where the hidden Markov chains have a different number of jump times.

In order for the HM algorithm to generate a Markov chain that converges to the target density, it must propose a reversible change to the parameter that is being updated—that is, the probability of proposing the current parameter value given the proposed parameter value must be greater than 0—hence the name reversible jump. While this is a general requirement with respect to constructing HM algorithms, the name reversible jump has become associated with HM algorithms that are designed for jumping between model spaces that have different dimensions. A second issue that needs to be considered when constructing an appropriate reversible jump HM algorithm has to do with what Green (1995) referred to as dimension matching. While dimension matching is needed to ensure that the resulting acceptance probability calculated in the HM algorithm is actually a probability, dimension matching can be taken care of in a natural way if the full conditional densities of the "variable dimension" parameters are defined with respect to a common dominating measure. The likelihood of the models that we propose is defined with respect to a common dominating reference measure. Therefore, the resulting full conditional

densities for the hidden Markov chains are densities with respect to this common dominating measure. This construction takes care of the dimension matching issues.

We used three different algorithms for updating D. The first algorithm is an *independence algorithm*, which ignores the current realization of D and proposes realizations of D by drawing from the prior density of D. This results in proposed realizations that are considerably different, in terms of the posterior density, and as a consequence this algorithm tends to result in large but infrequent moves. The other two algorithms create proposed realizations of D by making small modifications to the current realization of D. The second algorithm is a *refinement algorithm* where the proposed realization of D is created by modifying one of the jump times of the current realization of D. The third algorithm is a *birth-death algorithm* in which the proposed realization of D is created by either inserting a new interval into the current realization of D—a birth—or removing an interval from the current realization of D—a death. The independence algorithm has obvious advantages when the posterior distribution is multi-modal or when a poor initial value of D has been chosen, where the refinement algorithm and the birth-death algorithm have the advantage of more efficiently exploring the modes of the posterior distribution. In order to take advantage of the properties of these three algorithms, one of these three algorithms is randomly chosen at every sweep of the MCMC algorithm, to update each hidden Markov chain. We apply the algorithms proposed by Liechty and Roberts (2001; see their detailed descriptions) to estimate our model of visual attention states based on eye-movement data.

3. Analysis of Simulated Data

To illustrate the performance of our model and MCMC estimation algorithm and compare this with alternative models and algorithms, we generated synthetic data according to our model, but with a single semi-Markov process, with 16 observed states (i.e., a 4 by 4 grid) and two latent states. We analyze this data using a collection of models in order to investigate the effect of incorporating various model components. The different models that were considered are as follows:

1. A semi-Markov model with waiting times that depend on the latent states. This is our proposed full model. For this model we used reversible jump HM algorithms that identifies the number of jumps between the latent states.
2. A Markov model with constant waiting time hazard. This is the restricted Markov version of our proposed full model. Again we used the same reversible jump HM algorithm that identifies the number of jumps between the latent states.
3. A semi-Markov model with waiting times that depend on the latent states, but that has a fixed number of jumps between the latent states. For these models we used a stripped version of the reversible jump MH algorithm that assumes the number of jumps between the latent states to be fixed. We apply this model with an assumed value of 1 to 10 jumps.

We present the results of the analysis in graphs in Figure 2. Figure 2 shows the realization of the true hidden Markov Chain in the top panel. The second panel shows the results of the semi-Markov and Markov model estimates (Models 1 and 2), and the third panel shows several of the fixed switch point models (Model 3). In each panel the vertical axis shows the probability of being in latent State 1, the horizontal axis shows the simulated time scale. The top panel of the Figure displays the realized states of the hidden chain. It should be noted that the particular realization generated has a number of short waiting times, near the beginning of the time axis and near the end of the time scale. It seems obvious that such short visits to latent states are difficult to identify from the data (plotted in the graph as jittered points), which is desirable for the investigation of the performance of the various models.

The nonconstant waiting time model does a better job of estimating the hidden Markov chain than the constant model, as can be seen in the second panel in the Figure. The "long"
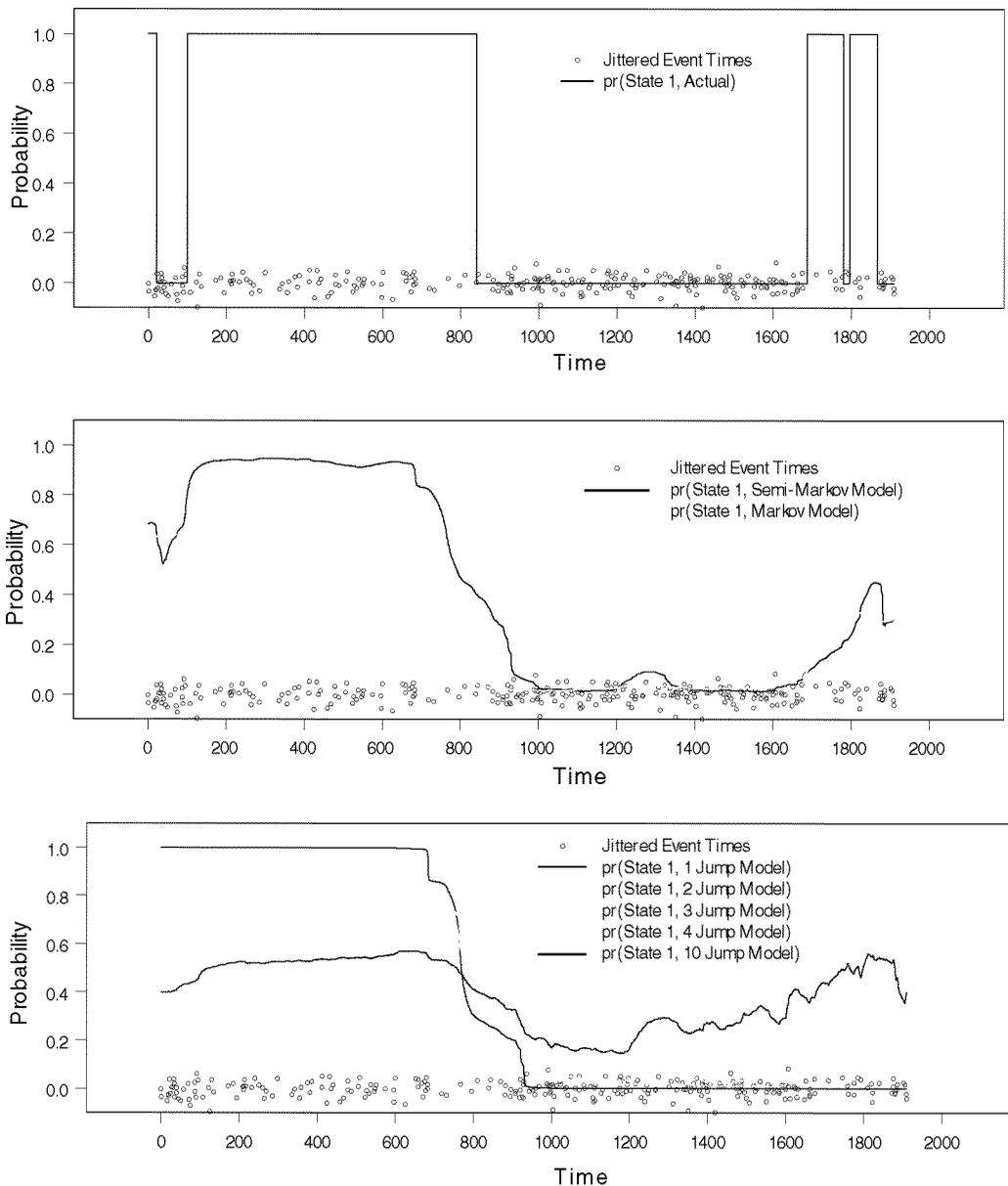
FIGURE 2.
Dots are jittered observed event times. Lines are the probability of being in the first latent state, for the true data (top panel), the Markov and semi-Markov models with variable numbers of jumps (middle panel) and the semi-Markov models with the numbers of jumps fixed (bottom panel).

period that the system spends in State 1 from about $t = 200$ to $t = 800$, is identified with high probability (around 0.9), and the same holds for the time spend in State 2, from about time 800 to 1600. Even the very short transitions between States 1 and 2 and visa versa at both ends of the time scale are visible in the graph. The model that assumes a constant waiting time does worse in that respect. The long State 1 visit in the beginning is identified with a relatively low probability of 0.6, but for a substantial portion of the relevant time scale the probability is 0.5 or lower. The short jumps between the latent states at both ends of the time scale are not as clearly identified when compared to the semi-Markov model. We conjecture that the difference in performance

between these two models arises because the semi-Markov model capitalizes on the information about the difference in the waiting times for each state to distinguish between the different states.

The analyses where the number of jumps between states is fixed reveal that the results are very dependent upon that assumption. When one change point is assumed, long visits to State 1 and State 2 are identified with high probability, but the short visits are not identified at all. That situation improves when the number of change-points set in the model is increased. Models with 2 to 4 change-points (shown in the figure) come close in performance to the semi-Markov model, but the profile of transition probabilities is less sharply defined. The semi-Markov model resulted in an a posteriori mean number of jumps of 3.74 (the Markov model resulted in a lower posterior mean number of jumps: 2.81). For models with more than four change points, performance starts to decline. The graphs of the estimated jump chain probabilities become flatter and the individual states are less well identified, as exemplified by the graph for the model in which 10 change points are assumed: the probability of State 1 hardly ever rises above 0.5.

This analysis of simulated data demonstrates that even for a single process (i.e., one subject and one stimulus) with limited information, and a structure that is difficult to identify due to short stays in the latent states, our model and MCMC estimation algorithm do a good job in identifying the transitions between those latent states from the limited information in the data, and a better job than several competing models and algorithms, as is apparent from Figure 2. With that support, we set out to analyze real eye-movement data.

4. Application to Visual Attention to Advertisements in Magazines

4.1. Subjects and Stimulus Material

A random sample of sixty-nine female consumers, between 19 and 52 years, were invited to the market research agency (Verify International in Rotterdam, The Netherlands) that collected the data. The stimulus material was comprised of seventeen full-page advertisements with their surrounding editorial material from consumer magazines. The advertisements appeared in a larger stimulus set exposed to subjects on 21-inch (3:4) high-resolution NEC monitors. The target advertisements promote different services (airline, investment, computer shop), food products (tortilla chips, cookies, sausage), detergents, tobacco, alcoholic beverages and durable products. The target advertisements all contained text and pictorial (scene) information. There has been little research into visual attention for such mixed information mode stimuli (see, e.g., Rayner, 1998; Rayner et al., 2001). We expect local and global covert attention to be both more prevalent for such complex stimuli than for example in reading or scene perception tasks as used frequently in previous eye-tracking studies. This makes this data set very well suited for our purposes.

4.2. Eye-Movement Recording

Subjects were seated in a comfortable chair in front of the monitor, with their head against a forehead rest to ensure reliable eye tracking. Instructions appeared on the monitor. First, a calibration task was performed. Next, subjects engaged in a typical visual exploration task as they would at home or in a waiting room (Wedel & Pieters, 2000). They proceeded through the stimulus set at their own pace by pushing a switch in front of them. Stimuli were randomized across subjects to control for order effects.

While paging at their own pace, subjects' eye-movements were recorded with infrared corneal reflection eye-tracking methodology (e.g., Ober 1994). The specific eye-tracking equipment we used has been developed by Verify International. It measures the position of the fovea at 50 Hz and records the exact locations of eye fixations continuously. We retain the data at the recording level of 20ms before processing. The systems' precision is 0.5° (for details of

the eye-movement recording, see, for instance, Wedel & Pieters, 2000). The viewing distance (eye-screen) was 30 inches.

4.3. Grid Overlay

In basic reading research, fixations are typically analyzed at their exact position. However, there are advantages to assigning fixations to larger areas in visual exploration research. First at each eye fixation the perceptual span covers more than just the exact fixation point due to parafoveal and peripheral vision (Anstis 1974; Findlay & Gilchrist, 1998; McConkie, 1983), particularly in scene perception. Since a larger part of the stimulus is covered in each fixation than is suggested by exact fixation positions, predefined areas are more appropriate as the unit of analysis than exact fixation points. Second, given the number of fixations and saccades that occur during a single exposure to an ad, the dimensionality of the data become prohibitively large when based on exact fixation positions instead of on a limited number of areas. Aggregation of raw eye-tracking data is thus desirable in visual exploration research, but the level of aggregation needs to be determined.

Previous research has used coarse grids of up to 16 elements superimposed on the stimuli (e.g., Pieters, Rosbergen & Wedel, 1999; Viviani 1990). To examine visual attention to magazine advertisements in detail, we superimposed a grid of six by eight (48 elements) on the stimuli. In choosing the grid size, one balances level of detail against data sparseness, where in this study we need to accommodate different stimuli (ads) and different modes (text/pictorial) within the stimuli. The advertisements were resized to fit the high-resolution NEC monitor, with a diagonal of 21 inch (12.6 by 16.8 inch), the 48 grid elements being 2.1 by 2.1 inch. This grid sizes optimally balances the amount of information retained versus sparseness of the data. All saccades between and fixation frequencies within elements of the grid for all stimuli and subjects are retained, and individual fixations and their duration are preserved as well. Below, we provide a check on the validity of our choice of the grid size by investigating the distribution of the observed saccades versus that predicted from our model.

4.4. Data Description

Figure 3 provides a description of the data. The upper left-hand panel shows the distribution of the numbers of fixations per stimulus. The distribution is quite skewed, with a maximum number of 105. The upper right-hand panel provides the distribution of the saccadic lengths across the grid elements, in numbers of grid points. One may observe that a majority of saccades occurs within one or two elements, but long saccades of up to seven do occur. The lower left-hand panel shows that across all stimuli and subjects initial fixations tend to be located near the center of the stimulus. However, the lower right hand panel shows that three locations of high density of fixations tend to exist, possibly caused by the typical spatial layout of the stimuli.

5. Results of Model Estimations

5.1. Convergence and Prior Sensitivity

We assess convergence of the MCMC algorithm for the various models estimated by monitoring the behavior of the scalar parameters, which are at the "top" of the model hierarchy using the Gelman and Rubin (GR; 1992) statistic. In addition we monitor the convergence of six, randomly selected, individual hidden Markov chains, by calculating the L^1 statistics discussed in Liechty and Roberts (2001). Both the GR statistic and the L^1 statistic assess convergence by monitoring the behavior of parameters from multiple MCMC analysis that are run in parallel. In our analysis we run five chains in parallel, the GR statistics and the L^1 statistics show that the
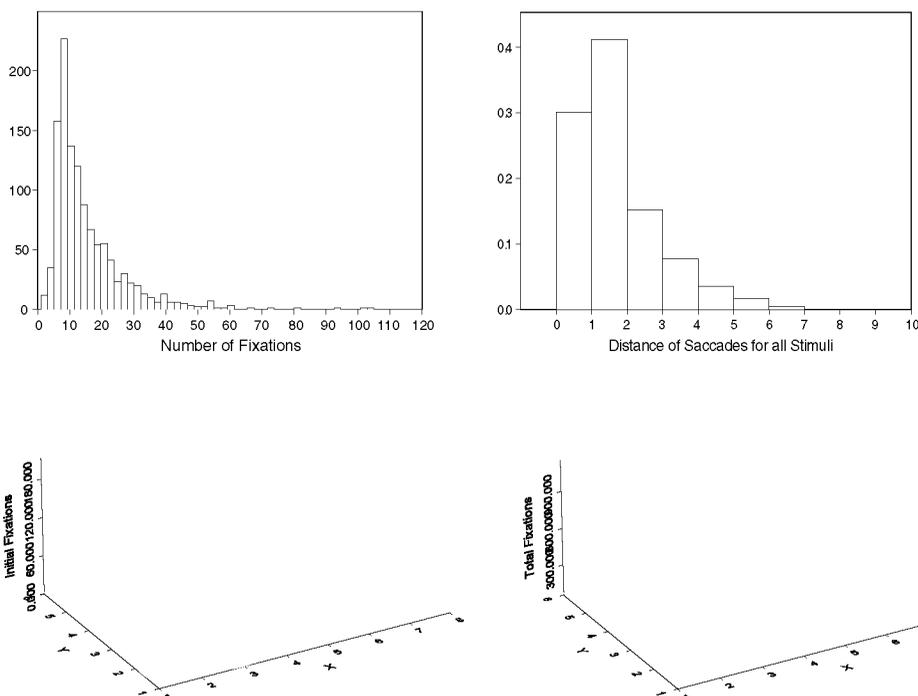
FIGURE 3.

Summary of the eye-movement date: Histogram of fixation frequency (top-left), saccadic length (top-right), distribution of initial fixation locations (bottom-left), total fixation locations (right) across subjects and stimuli.

MCMC chain has converged after 7,000 iterations. We investigate prior sensitivity of our model by varying the parameters of the prior distributions. In all cases we found that the posterior distributions of the parameters were similar for a range of prior specifications, which shows that the prior is dominated by the likelihood due to the large amount of information in the data.

5.2. Local and Global Covert Attention States

We first investigate the support that exists for our claim that there are local and global covert attention states (e.g., what evidence is there for a two-state model, rather than for a one-state model that does not distinguish between global and local attention states), and for a semi-Markov model over a Markov model. We assume a priori, that each of the models being considered is equally likely and calculate the posterior odds ratio using the fourth sampling based estimator proposed by Newton and Raftery (1994), which in this case is the Bayes Factor (O'Hagan, 1994).

There is very strong support for the two-state model when compared with either version (Markov or semi-Markov) of the one-state model: the Bayes Factor is 237.5. Thus, importantly there is strong statistical evidence that indeed global and local covert attention states underlie subjects' eye-movement patterns in this study. Further, there is strong support for the two state semi-Markov model, compared to the two state Markov model: the Bayes Factor is 468.7, which implies that global and local attention states have different associated fixations durations. To examine the fit of our model and provide face validity to the choice of our grid-size, we predict the mean and standard deviation of the length of the saccades. Note that saccade lengths themselves are not included as a part of our model. The saccade data have Mean $= 0.216$, SD $= 0.134$, while those predicted from our model have Mean $= 0.199$, SD $= 0.202$, which is fairly close.

TABLE 1.
Posteior means and standard deviation (in parentheses) of key parameters

Parameter (-function)	Interpretation	$k = 1$ Local Covert Attention	$k = 2$ Global Covert Attention
ν_k	Probability of starting	0.992 (0.003)	0.008 (0.003)
d_k	Location parameter of exit probability distribution	1.01 (0.004)	1.03 (0.027)
b_k	Scale parameter of exit probability distribution	9.88 (0.525)	3.39 (0.013)
$\dfrac{d_k}{d_k + b_k}$	Probability of terminating	0.093 (0.005)	0.234 (0.013)
β_{kL}	Location parameter of local jump distribution	217.34 (4.697)	31.39 (0.489)
β_{kG}	Location parameter of global jump distribution	71.84 (1.852)	51.91 (1.166)
$\dfrac{\beta_{kL}}{\beta_{kL} + \beta_{kG}}$	Probability of a local jump, given current state	0.758 (0.000)	0.385 (0.003)
λ_k	Location parameter of waiting time distribution	0.883 (0.021)	4.58 (0.123)
$\dfrac{1}{\lambda_k}$	Mean duration, in seconds	1.130 (0.027)	0.219 (0.006)
g_k	Location parameter of fixation intensity distribution	21.39 (1.145)	16.47 (0.243)
h_k	Scale parameter of fixation intensity distribution	4.32 (0.243)	2.42 (0.257)
$\dfrac{h_k}{g_k}$	Mean fixation duration, in seconds	0.202 (0.003)	0.147 (0.004)

Based on these model tests, we focus on the aggregate results for the two-state semi-Markov model and we illustrate subject and stimulus level inference on the inferred attention patterns. Our inferences based on the posterior distributions of the model parameters are confined to the covert attention states, since identifying these is the primary goal of this research. We report posterior Means and Standard Deviations (SD) for key parameters in Table 1 and in the text below.

The results of the application of the multi-process, semi-Markov model to the observed fix-ation sequences across subjects and stimuli in the main study reveal the following. The expected number of fixations per stimulus is 15.300 (SD = 11.136), while the expected total viewing duration per stimulus is 3.339 sec (SD = 0.552). As expected, the probabilities of starting the fixation sequence are substantially larger (Mean = 0.101, SD = 0.037) for the center cells of the lattice ($x = 3$ to 4, $y = 3$ to 5) than for the exterior cells (Mean = 0.009, SD = 0.007).

Next, we examine the covert attention states, local or global, in which people generally start and terminate exposure to the advertisements. Then, we examine the patterns of switching between the covert attention states, during exposure to the stimulus, as well as the time they spend in the two states, and the average fixation duration in each state. After that we integrate these findings, and present subject and stimulus inferences based on the model estimations.

5.3. Start and Termination States of Visual Attention

Subjects nearly always start in the local attention state ($k = 1$), with an estimated prior probability of 0.992 (Table 1). This is partly caused by the observed duration of the first fixation, which is significantly larger on average (Mean = 0.268 sec, SD = 0.146) than the subsequent ones (Mean = 0.213 sec, SD = 0.132). Fixations in the local attention state tend to be substantially longer, Mean = 0.202 sec, (SD = 0.003), than those in the global attention state, Mean = 0.147 sec, (SD = 0.004), see Table 1. The probability of terminating exposure to the advertisements is much larger when in the global attention state (Mean = 0.234, SD = 0.013) than when in the local attention state (Mean = 0.093, SD = 0.005), see Table 1.

5.4. Switching Between and Time in Attention States

During exposure to an advertisement, subjects jump between the local and global states 2.6 times on average (SD = 0.049), and tend to spend longer in a local (Mean = 1.13 sec, SD = 0.027) than in the global state (Mean = 0.219, SD = 0.006), as can be seen from the parameter estimates in Table 1. Also, fixations in the local attention state tend to be substantially longer (Mean = 0.202 sec, SD = 0.003), than those in the global attention state (Mean = 0.147 sec, SD = 0.004), see Table 1. Note that the expected time spend in a global attention state is only around 1.5 times as long as the expected duration of a fixation in that state. Subjects spend 84.7% of the time in the local attention state.

When people are in the local attention state they are a little more than three times as likely to make an observable local saccade on the grid: $\Pr(\text{local jump} \mid \text{local state}) = 0.758$ (SD = 0.000), and when in the global attention state they are almost two times as likely to make an observable global saccade: $\Pr(\text{global jump} \mid \text{global state}) = 0.615$ (SD = 0.003), see Table 1. This in itself provides a validation of our choice of the grid size. Additional validation comes from the distribution of the length of a saccade in the local state and global state when compared with the distribution of the length of all of the saccades. The difference in these distributions is shown in Figure 4, which reveals that indeed the saccade length in the global state tends to be longer.

Taken together, these results show that the local attention state is dominant, with a high probability of starting in that state, where there are occasional short jumps to the global attention state with a high probability of ending attention to the stimulus in that state.

5.5. Integrating Model Findings

We believe that the interpretation of these results is as follows. The local covert attention state serves to extract detailed information from specific regions of the stimulus to build or strengthen memory traces, for purposes of object identification or recognition, which is indicated by the longer duration of the fixations in this state (Viviani, 1990). Also, subjects tend to spend most of the exposure time in this local state where information is extracted and the cortical representation of the objects in the scene is strengthened. Subjects appear to start visual attention in the local state, with a quite long fixation, much more likely at the center of the stimulus. This however, may not be strictly a "local fixation", in that it serves a different purpose. We conjecture that around the initial fixation, pre-attentive computation of perceptual salience occurs, based on features such as color, orientation, shape and size, relative to their background (cf., Itti & Koch, 2001). Here, salient regions in the stimulus are identified.

The global covert visual attention state, controlled by the posterior parietal cortex, then tends to consist of short bursts of longer saccades and fixations of short duration. The global state serves to redirect attention to the next salient region of the stimulus, through winner-takes-all and inhibition of return neural mechanisms (e.g., Klein, 1988; LaBerge, 1998). In that global attention state, as suggested by the short duration of fixations, limited or no detailed information
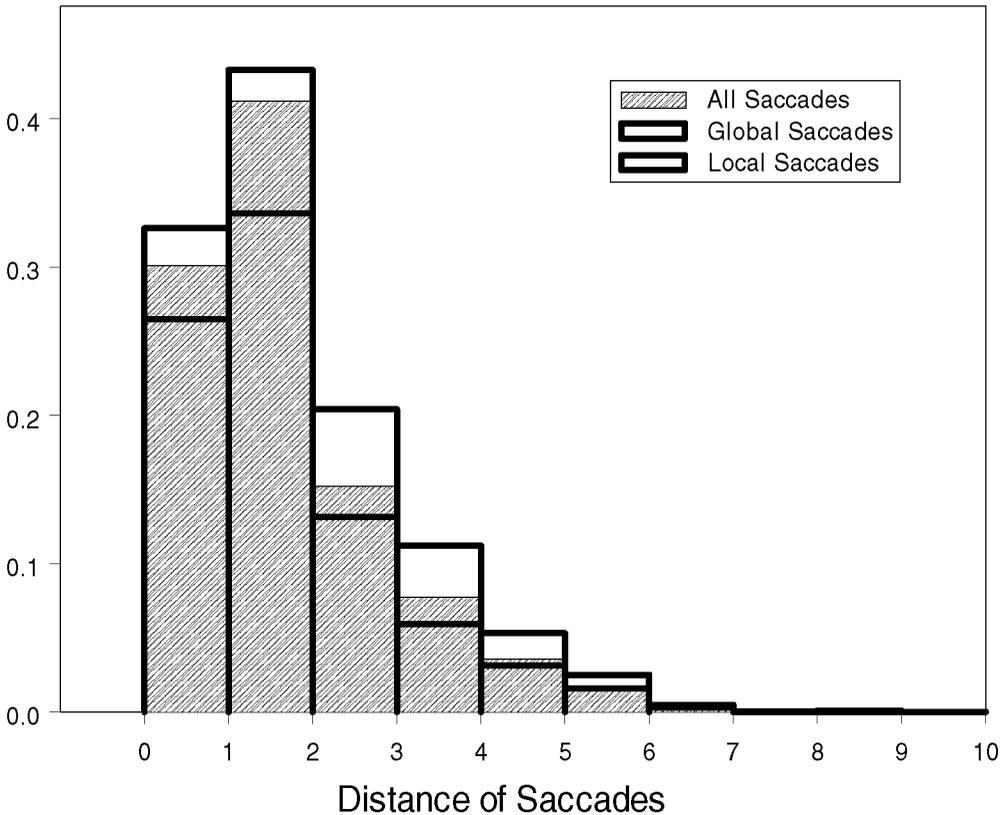
FIGURE 4.
Distribution of saccade lengths for local and global states: A histogram summary of the posterior distribution of saccade lengths for both the local and global states, compared to the distribution of all of the saccade lengths.

is extracted from the stimulus for further higher-order processing or memory storage (cf., Mc-Conkie, Reddix, & Zola, 1992). After such a short sequence of longer ballistic eye movements in the global state, the probability of local information processing increases again and the prefrontal cortex control mechanism switches to the local attention state to enable information extraction. The time spent in the global state increases the probability of switching to the local state, and visa versa.

Thus, by switching back and forth between local and global attention, the problem of interpreting the complex visual scene is broken down into a sequence of simpler localized interpretations of the most salient regions (Itti & Koch, 2001). The fixation sequence is terminated in the global attention state when no regions are identified that are sufficiently salient to warrant further detailed local examination.

5.6. Subject and Stimulus Inferences

In order to illustrate the subject- and stimulus-specific posterior inferences that can be obtained from our model, we present results for selected stimuli and subjects in Figure 5. The figure displays the inferred attention pattern of selected subjects and stimuli, defined in terms of durations of the two covert attention states and switching patterns among them. The figure reveals substantial variety in the probability of local (displayed as a black line) versus global covert attention, as derived from the observed fixations (displayed as dots at the bottom of the graphs) across subjects and ads.
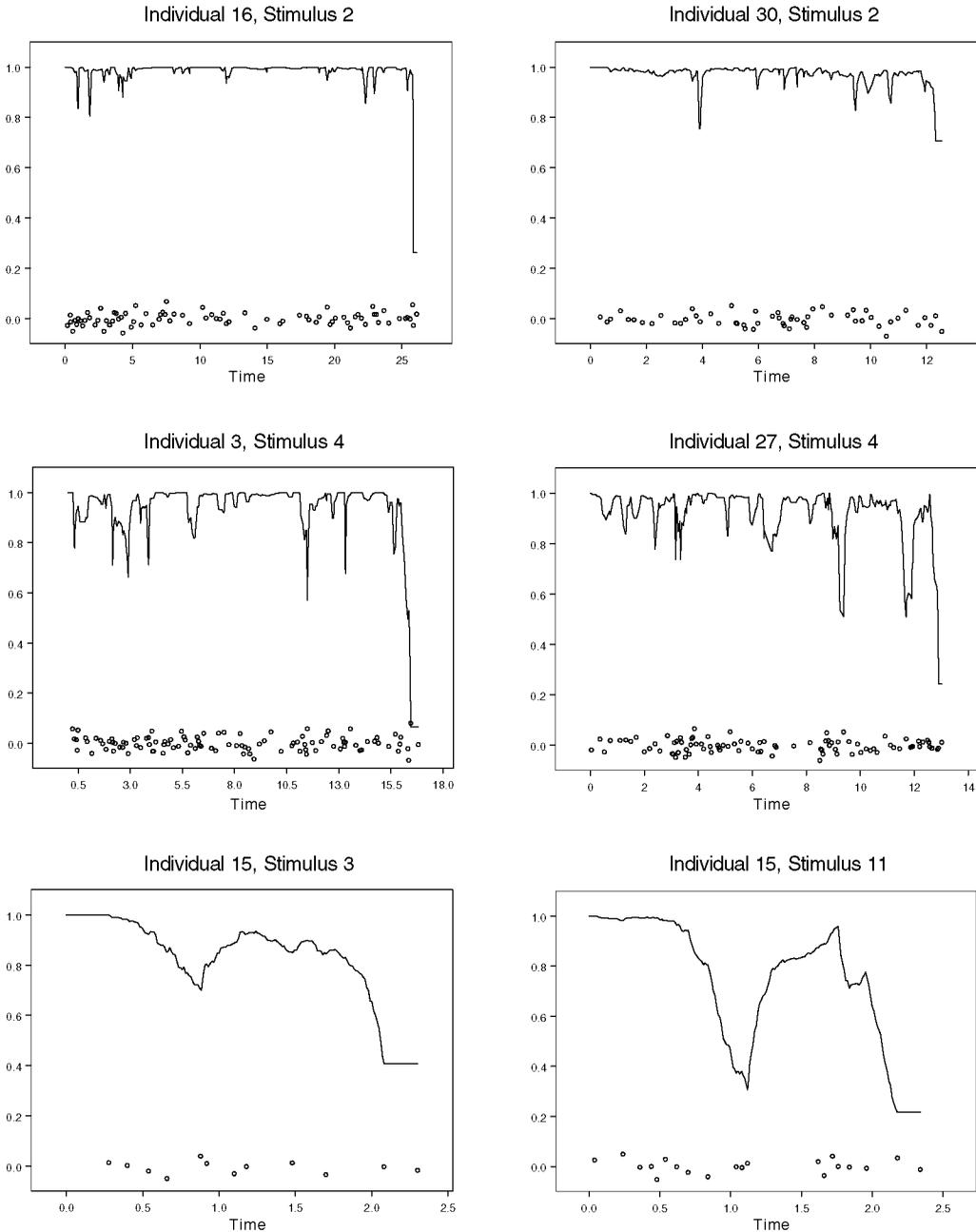
FIGURE 5.
Probability of being in local attention state: Lines indicate the probability of being in the local attention state across duration of the exposure. Dots indicate individual fixations. Plots are presented for selected subjects and stimuli only.

Our review of the graphs suggests that there is a large variation in the estimated probabilities likely to be caused by effects specific to the stimuli. Take for example Stimuli 2 and 4 in the top two rows of Figure 5. These stimuli attract a relatively large number of fixations from several individuals—e.g., Subjects 16 and 30 for Stimulus 2, and Subjects 3 and 27 for Stimulus 4, as shown. Given the large number of fixations, and the large proportion of time spend in the

local covert attention state, it seems that these stimuli contain much relevant information for the subjects in question, who therefore peruse them in detail, without much need to jump between distant salient regions. Both stimuli are effective in eliciting local attention for all four subjects, but the pattern of probabilities of jumping between the two attention states is quite different for Stimulus 2 compared to 4. When Subjects 16 and 30 explore Stimulus 2, across the whole exposure duration they are in the local attention state with high probability, where bursts of global attention occur infrequently and with fairly low probability.[1] On the other hand, Subjects 3 and 27 exhibit a pattern of long spells of local attention, interchanged with short bursts (of high probability) of global attention, when exploring Stimulus 4. For both subjects the global attention spells tend to increase in intensity and frequency over the time of exposure, as if the interruption of global jumps by sufficiently salient regions eliciting local attention diminishes over time. Stimulus 4 likely contains relevant information, that is however, more localized than Stimulus 2. In all cases, subjects terminate exposure to the stimuli in the global state. We found similar patterns for other subjects viewing each of these stimuli.

For scanpaths of moderate length, some subjects exhibit rather consistent attention patterns for several of the stimuli viewed. For example, Subject 15 exhibits similar patterns for Stimuli 3 and 11 as shown in the bottom row of Figure 5. Notably, these scanpaths consist of far fewer fixations as compared to those discussed above. Exposure starts with a high probability of the local attention state that decreases, and then increases somewhat and finally decreases towards zero, since attention is terminated in the global state again. The graphs seem to reflect the subject scanning the stimulus for relevant information, exploring a salient area in more detail and terminating exposure after some more global search. It thus seems that these two stimuli contain little relevant information for this subject.

The first four graphs in Figure 5 suggest a dominant stimulus effect on the attention state, but the final two graphs reveal a similar pattern for Subject 15 when attending to Stimuli 3 and 11 that may be caused by these two stimuli sharing low relevant information content for the subject in question.

The strong communalities in attentional patterns between subjects for particular stimuli are suggestive of bottom-up control of attention by salient locations in these scenes. Saliency might be computed pre-attentively, possibly largely around the first fixation, that is of a particularly long duration and located at the center of the stimulus. In that long, centrally located fixation saliency across the entire scene may derived using "center-surround mechanisms", involving neurons that respond to image contrast differences between the center region fixated at, and broader concentric surround regions (e.g., Palmer 1999). If the scene is composed of sufficiently salient regions that are conspicuous or pop-out in their context due to features such as light intensity, color or orientation contrast, those regions involuntarily attract attention, and scene perception is broken down into a sequence of localized perception tasks, with controlled switching between local and global covert attention states, where the global state may become more dominant over time as the most salient regions have been explored locally. In those cases, the stimulus will elicit a comparable pattern (in a stochastic sense) of covert attention from different subjects, as evidenced by their eye-movements.

The different pattern of covert visual attention observed, however, for some stimuli (e.g., Stimulus 4 above), suggests that here already in the pre-attentive phase low overall saliency is identified, which substantially reduces the probability of local attention throughout, reducing the number of fixations and exposure time. This may point to some higher level processing of extracted information in the first stages of stimulus exposure, perhaps even pre-attentively.

The results illustrated in Figure 5 illustrate that top-down volitional control of attention may also play a role, as evidenced by the significant differences in attentional patterns between subjects for the same stimulus. This deliberate form of attention is probably controlled from higher

---

[1] We define "changing state" as the point where the transition probability of switching the latent states crosses the 50% probability line.

brain area's, such as the frontal lobes and is slower than bottom-up, salience driven attention (Itti & Koch, 2001). Such top-down attention may be driven by prior expectations on the layout of the scene, since subjects in general have certain expectations about the relative position of informative regions in print advertisements, such as the headline, pictorial, body text and packshot (Rosbergen, Pieters, & Wedel, 1997). Subject specific factors such as involvement and familiarity with the stimulus may also play a role (Pieters, Warlop, & Wedel, 2002). It is likely that top-down and bottom-up control operate in parallel and interact.

6. Conclusion

Whereas the local and global covert visual attention states of subjects are unobservable, the Hidden Markov Model proves to be a powerful tool to identify such states from eye-movement data. The model that we have developed was derived from prior theory on visual attention and has several features that make it attractive as a tool for the analysis of empirical data in that field. The empirical illustration on visual attention to magazine advertisements showed the performance of the model, and the new insights it can provide into issues that have been debated extensively in the literature. Print advertisements are heterogeneous stimuli with complex configurations of text and pictures (scenes), which are more likely to activate both covert attention states and switching between them, than do the typical stimuli used in, for instance, reading tasks. This makes magazine advertisements well suited to test the presence of those qualitatively different visual attention states and how they evolve over the time course of stimulus exposure.

Typically, experimentally observed fixation durations in reading show a mean of around 200 ms, but have a substantial proportion of durations of 150 ms or less. In a recent study of attention to print advertisements, Rayner et al. (2001) found average fixation durations of around 220 ms. Our mean fixation lengths are in accordance, but somewhat shorter on average: 202 ms in the local attention state, which dominated, and 147 ms in global attention. Our shorter fixation durations may be due to the fact that we used multiple stimuli (17 advertisements) in a natural context as compared to the few stimuli and increased awareness conditions regularly present in controlled eye-movement experiments. In accordance with this, free viewing of scenes in their natural context has been reported to produce much more sparse cortical neural activity then exposure to small numbers of experimental stimuli in laboratory settings. This is caused by inhibitory neural activity that arises when stimuli extend beyond the classic receptive field of the neurons in question, which seems required for perceptual grouping (Itti & Koch, 2001). Since the subjects in our study controlled exposure duration to the ads themselves, the total looking time was typically short and fixations often few in number. Subjects in the present study spend on average only about 3.4 seconds on each stimulus, fixating about 15 times. Here, inference based on MCMC pays off, since it enables the identification of the hidden states from limited data, using Bayesian shrinkage that pools subsets of the parameters across subjects to allow for more powerful inference.

6.1. Implications for Theories of Visual Attention

Our empirical findings are interesting from a substantive point of view, and extend current knowledge about attention processes and their relationships to eye-movements. Instead of following an orderly sequence of global-local attention states during stimulus exposure, as has been suggested previously (e.g., Antes, 1974), our subjects used visual attention states more dynamically, switching back and forth between them. A particular advantage of the application of the reversible jump MCMC algorithm is that it enables us to estimate a hidden Markov model that allows subjects to switch between latent states an arbitrary number of times. Subjects generally started in local attention state with a first long fixation, argued to be directed at determining salient regions, sometimes switched between states, where the intensity of the global state often increased over time, and ended in a global attention state. Thus the overall scanpath serves to

break down the complex problem of advertisement perception into a smaller number of simpler subproblems of perception of local regions in the ad, where the global attention state serves to redirect attention between them. The results suggest that higher-order (lexical, conceptual) information is more likely to be extracted while subjects are in a local attention state, while the global covert attention state serves to redirect attention to perceptually salient or potentially informative regions of the stimulus. The shorter fixations durations in the global state reflect this process of exploration rather than of intense, higher level information pick-up, which is in line with reading and scene perception research. For instance, McConkie et al. (1992) found that, in reading tasks, little lexical information is extracted through fixation durations of about 140 ms or less.

We speculate that a "Start Local—End Global" attention pattern as we observed is more typical for exploration tasks with heterogeneous stimuli like magazine advertisements. In such tasks, subjects apparently start constructing a rough mental map of the ad based on saliency of regions, and then explore the stimulus for other informative or perceptually salient areas, globally redirecting attention. They switch between local and global states, increase the intensity of the global state as less remaining salient information seems available, exploring in a last "sweep" whether there are areas left that appear perceptually or semantically sufficiently attractive, and terminate attending the stimulus in the global attention state.

The visual attention process observed here, characterized by fairly extensive switching between local and global covert attention states, might be specific for the task and stimulus. In other tasks like target search from multi-element displays (Findlay & Gilchrist, 1998; Monk, 1984) different patterns may be observed. In target search tasks, subjects may quickly scan the stimulus to determine a priority sequence of areas with a decreasing likelihood of containing the target or to maximize the likelihood of a perceptual pop-out (Pomplun, 1998; Wolfe, 1998), which would reflect dominance of the global attention state. In tasks for much more homogeneous stimuli, for example reading, presence of both attention states and extensive switching between them is less likely to occur. Here the local attention state may dominate (McConkie, 1983; Rayner, 1998). In other words, we speculate that the prevalence and the order in which the two covert attention states occur may depend on bottom-up factors such as the type of stimuli used (e.g., text, scenes, or a mix of the two) and on top-down factors such as the specific tasks that subjects engage in (e.g., exploration, target search, reading). Future work that tests these speculations may offer improved understanding of the role that stimulus and task characteristics play in visual attention, which has been identified as a key white spot (Henderson & Hollingworth, 1998, 1999).

For future research it would be of interest to design more detailed experiments that allow for a deeper substantive interpretation of the findings derived from our model. For example, it would be of interest to investigate to what extent the differences in the local and global states are caused by specific stimulus features, in particular differences in textual and pictorial information. It may be that the local state occurs more often in text reading and the global state more often in scanning the pictorial. But perhaps, the global state occurs when people switch back and forth between text and pictorial. Further and more detailed experiments, where our model is applied to more precisely controlled stimuli that consist of text only, or pictorial only, could shed further light on this issue. For example, in text reading, it would be of interest to see whether our model picks up the skipping of familiar words, and the back-sweep from the end of one line to the beginning of the subsequent line as a jump in the global rather than the local attention state (Reichle, Pollatsek, Fisher, & Rayner, 1998).

Further, experiments where the grid is based upon a substantive partitioning rather than a spatial partitioning of the stimulus, as was done in Rosbergen, Pieters and Wedel (1997), may enable deeper psychological interpretation of the model results. For example, it may resolve questions whether repeated local fixations represent attention to and within specific *objects* in the stimulus and global jumps represent shifts in attention between those *objects*. This may shed new light on the issue whether and when covert attention selects objects or locations (e.g., Vecera & Farah, 1994). We do believe that our model is a powerful tool to explore these issues in future research.

The model may also inspire more work on the role of top-down (subject) and bottom-up (stimulus) factors in visual attention, which has long been debated (e.g., Egeth & Yantis, 1997). Previous research in a marketing context has observed significant individual differences in the extent to which perceptual features of advertisements, such as size, contrast and color, influence the overall duration of visual exploration (Rosbergen, Pieters & Wedel, 1997). The stimulus level attention process estimated in the current research corroborates this finding. Follow-up research with the proposed model may provide new insights into the contribution of top-down and bottom-up factors by examining their impact on local and global visual attention states. Such research may identify specific characteristics of subjects and stimuli that affect the latent transition probabilities, event durations and exit probabilities. The model's capability to compare stimuli across subjects and subjects across stimuli in intuitive ways facilitates this.

Obviously, this study is not without caveats. For the data analysis, the exact eye fixation locations were aggregated into cells on a lattice overlaid on top of the ads. The fine grid that we employed balanced the level of detail in fixation locations with size of the modeling and estimation task. The size of the elements of the lattice may affect our model estimates. But, we would like to note that we are primarily interested in identifying the latent attention process, using the observed saccades on the elements of the lattice as indicators. The empirical finding that in the local latent state "local" saccades are more than three times as likely, while in the global state "global" saccades are over two times as likely, as well as the comparison of the observed and predicted distribution of saccades, supports our choice of the lattice in the application. Still, the use of a lattice to classify the saccades introduces measurement error in the data to be analyzed. Such error is accommodated in our model to a certain extent, since it is formulated as a stochastic process describing the transitions on the lattice. The aggregation offers the advantages of allowing for an appealing definition of global and local jumps, as jumps on (non) neighboring cells and of reducing data sparseness. The grid size was chosen to balance information and sparseness of the data, and it is larger than general. However, the issue of the sensitivity of our results to aggregation remains, and it would be of interest in future research to derive optimal lattice sizes.

Another caveat pertains to the first fixation. Whereas due to its long duration our model classified it as arising from the local covert attention state, this may not be appropriate. There is some evidence that the first fixation may serve different functions than follow-up fixations. In scene perception and target search, the perceptual saliency of the stimulus might be determined pre-attentively around this first fixation, which is a quite different process than that of information extraction and recognition in the local attention state. In reading research, the first fixation duration is sometimes distinguished from follow-up fixations, because it is expected to be more responsive to lexical, orthographic and other properties of the text (Inhoff & Radach, 1998). Thus, our model may be extended to deal with the specific nature of the first fixation and pre-attentive identification of stimulus salience. Here, it may be that some higher-level cognitive effort, rather than only that based on basic stimulus features are involved.

Much of the detail in the formulation and estimation of our mathematical model was enabled through substantially improved methodology for recording eye-movements. The current systems allow for automated precise recordings of eye-movements in more natural settings than has been the case heretofore. This in turn facilitates the collection of data on relatively large numbers of subjects and stimuli, enabling us to use finer grids and to specify and calibrate models with much behavioral detail. We conjecture that the visual attention model presented in the current study may serve as a useful starting point for such further work.

References

Anstis, S.M. (1974). A chart demonstrating variation in acuity with retinal position. *Vision Research, 14,* 589–592.

Antes, J.R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103,* 62–70.

Egeth, H.E., & S. Yantis (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269–297.

Ellis, S.R., & Smith, J.D. (1985). Patterns of statistical dependency in visual scanning. In R. Groner, G. W. McConkie, & C. Menz (Eds.), *Eye movements and human information processing* (pp. 221–238). Amsterdam: Elsevier Science Publishers.

Engbert, R., & Kliegl, R. (2001). Mathematical models of eye-movements in reading: A possible role for autonomous saccades. *Biological Cybernetics, 85*, 77–87.

Findlay, J.M., & Gilchrist, I.D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295–312). Amsterdam: Elsevier.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–511.

Gilks, W.R., Richardson, S., & Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in practice*. Boca Raton, FL: Chapman and Hall/CRC.

Green, P.J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika, 82*, 711–32.

Groner, R. (1988). Eye movements, attention and visual information processing: Some empirical results and methodological considerations. In G. Lüer, U. Lass, & J. Shallo-Hoffman (Eds.), *Eye movement research: Physiological and psychological aspects* (pp. 295–319). Toronto, Canada: Hogrefe.

Hacisalihzade, S.S., Stark, L.W., & Allen, J.S. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(3), 474–481.

Henderson, J.M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Amsterdam: Elsevier.

Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271.

Hodgson, M.E.A., & Green, P.J. (1999). Investigating Markov model discrimination for ion channels. *Journal of the Royal Statistical Society, Series A, 455*, 3425–3448.

Inhoff, A.W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Amsterdam: Elsevier.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Neuroscience, 2*, 1–11.

Itti, L., Koch, C. & Niebur, E. (1998). A model of saliency based visual attention for rapid scene analysis. *IEEE Transactions of Pattern and Analytical Machine Intelligence, 20*, 1254–1259.

Just, A.M., & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Boston, MA: Allyn and Bacon.

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*, 430–431.

LaBerge, D. (1998). Attentional emphasis in visual orienting and resolving. In R.D. Wright (Ed.), Visual attention (pp. 417–454). New York, NY: Oxford University Press.

Lévy-Schoen, A. (1981). Flexible and/or rigid control of oculomotor scanning behavior. In D.F. Fischer, R.A. Mony, & J.W. Senders (Eds.), *Eye-movements: Cognition and visual perception* (pp. 299–314). Hillsdale, NJ: Lawrence Erlbaum.

Liechty, J.C., & Roberts, G.O. (2001). Markov Chain Monte Carlo methods for switching diffusion models. *Biometrika, 88*(2), 299–315.

McConkie, G.W. (1983). Eye movements and perception during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 65–96). New York, NY: Academic Press.

McConkie, G.W., Reddix, M.R., & Zola, D. (1992). Perception and cognition in reading: Where is the meeting point. In K. Rayner (Ed.), *Eye movements and vision cognition: Scene perception and reading* (pp. 293–303). New York, NY: Springer Verlag.

Monk, T.H. (1984). Search. In J.S. Warm (Ed.), *Sustained attention in human performance* (pp. 293–321). New York, NY: John Wiley & Sons.

Newton, M.A., & Raftery, A.E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion). *Journal of the Royal Statistical Society, Series B, 56*, 3–28.

Ober, J. (1994). Infra-red reflection techniques. In J. Ygge & G. Lennerstrand (Eds.), *Eye movements in reading* (pp. 9–19). Tarrytown, NJ: Elsevier Science.

O'Hagan, A. (1994). *Kendall's advanced theory of statistics: Volume 2B Bayesian inference*. New York, NY: John Wiley & Sons.

Palmer, S.E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: The MIT Press.

Pieters, R., Rosbergen, E., & Wedel, M. (1999). Visual attention to repeated print Advertising: A test of scanpath theory. *Journal of Marketing Research, 36* (November), 424–438.

Pieters, R., Warlop, L., & Wedel, M. (2002). Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. *Management Science, 48*(6), 765–781.

Pomplun, M. (1998). *Analysis and models of eye movements in comparative visual search*. Gottingen, Germany: Cuvillier.

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

Posner, M.I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). Hove, U.K.: Lawrence Erlbaum Associate.

Rayner, K. (1998). Eye movement in reading and information processing: 20 Years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Rotello, C.M, Stewart, A.J., Keir, J., & Duffy, S.A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*(1), 219–226.

Reichle, E.D., Pollatsek, A., Fischer, D.L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157.

Rimey, R.D., & Brown, C.M. (1991). Controlling eye movements with hidden Markov models. *International Journal of Computer Vision, 7*(1), 47–65.

Robert, C.P., Ryden, T., & Titterington, D.M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov Chain Monte Carlo Method. *Journal of the Royal Statistical Society, Series B, 62,* 57–76.

Rosbergen, E., Pieters, R., & Wedel, M. (1997). Visual attention to advertising: A segment-level analysis. *Journal of Consumer Research, 24,* 305–314.

Ross, S.M. (1983). *Stochastic processes.* New York, NY: John Wiley & Sons.

Salvucci, D.D., & Anderson, J.R. (2001). Automated eye movement protocol analysis. *Human Computer Interaction, 16,* 39–86.

Sperling, G. & E. Weichselgartner (1995). Episodic theory of the dyanmics of spatial attention. *Psychological Review, 102*(3), 503–532.

Stolk, H., Boon, K., & Smulders, M. (1993). Visual information processing in a study task using text and pictures. In G. d'Ydewalle & J. van Rensbergen (Eds.), *Perception and cognition: Advances in eye movement research* (pp. 285–296). Amsterdam: North-Holland.

Underwood, G., & Radach, R. (1998). Eye guidance and human information processing: Reading, visual search, picture perception, and driving. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 1–28). Oxford, England: Elsevier.

Vecera, S.P., & Farah, M.J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General, 123*(2), 146–160.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 353–393) Amsterdam: Elsevier Science Publishers.

Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science, 19,* 297–312.

Wolfe, J.M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). Hove, U.K.: Psychology Press.

Wright, R.D., & Ward, L.M. (1998). The control of attention. In R.D. Wright (Ed.), *Visual attention* (pp. 132–186). New York, NY: Oxford University Press.

Yarbus, A.L. (1967). *Eye movements and vision* (L.A. Riggs, Trans.). New York, NY: Plenum Press.

Zangemeister, W.H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic pictures. *Neuropsychologia, 33,* 1009–1025.