Behavioral/Systems/Cognitive

From Salience to Saccades: Multiple-Alternative Gated Stochastic Accumulator Model of Visual Search

Braden A. Purcell, Jeffrey D. Schall, Gordon D. Logan, and Thomas J. Palmeri
Center for Integrative and Cognitive Neuroscience, Vanderbilt Vision Research Center, Deparment of Psychology, Vanderbilt University, Nashville, Tennessee 37240

We describe a stochastic accumulator model demonstrating that visual search performance can be understood as a gated feedforward cascade from a salience map to multiple competing accumulators. The model quantitatively accounts for behavior and predicts neural dynamics of macaque monkeys performing visual search for a target stimulus among different numbers of distractors. The salience accumulated in the model is equated with the spike trains recorded from visually responsive neurons in the frontal eye field. Accumulated variability in the firing rates of these neurons explains choice probabilities and the distributions of correct and error response times with search arrays of different set sizes if the accumulators are mutually inhibitory. The dynamics of the stochastic accumulators quantitatively predict the activity of presaccadic movement neurons that initiate eye movements if gating inhibition prevents accumulation before the representation of stimulus salience emerges. Adjustments in the level of gating inhibition can control trade-offs in speed and accuracy that optimize visual search performance.

Introduction

Many models of visual search and attention assume a map that combines bottom-up and top-down salience (also referred to as “priority”) to guide shifts of attention and eye movements (Treisman and Gormican, 1988; Duncan and Humphreys, 1989; Bundesen et al., 2005; Navalpakkam and Itti, 2007; Wolfe, 2007). Visually responsive neurons in the frontal eye field (FEF), superior colliculus (SC), and posterior parietal cortex modulate their firing rate according to both physical conspicuity (bottom-up salience) and behavioral relevance (top-down salience) (Schall and Hanes, 1993; Gottlieb et al., 1998; Bichot and Schall, 1999; Findlay and Walker, 1999; Constantindis and Steinmetz, 2005; Ipata et al., 2006; Thomas and Paré, 2007; Balan et al., 2008; Cohen et al., 2009; White and Munoz, 2011). However, it is poorly understood how neurons that encode salience are read out to guide saccades. Furthermore, no current model of visual search accounts for full response time (RT) distributions of error and correct responses (Wolfe et al., 2010; Palmer et al., 2011). Predicting full RT distributions can help to distinguish alternative models that may predict mean correct RTs equally well, but make very different predictions about underlying mechanisms (Ratcliff and Smith, 2004).

Stochastic accumulator models of perceptual decision making explain how a single stimulus guides selection of one response among alternative responses (Usher et al., 2002; McMillen and Holmes, 2006; Beck et al., 2008; Churchland et al., 2008; Furman and Wang, 2008; Ditterich, 2010; Leite and Ratcliff, 2010), but the natural world consists of many stimuli. FEF and SC movement neurons project to oculomotor brainstem nuclei and initiate a saccade when a fixed firing threshold is reached (Hanes and Schall, 1996; Hanes et al., 1998; Paré and Hanes, 2003). The firing rate dynamics of movement neurons are well explained by stochastic accumulator models (Boucher et al., 2007; Ratcliff et al., 2007), but the source of inputs driving movement neurons to threshold is not known. Furthermore, neurally plausible implementations of these models have never been generalized to decisions among multiple stimuli, as in visual search.

We present the first neural model of visual search formulated as a multiple-accumulator decision process. We significantly extend a novel modeling technique that defines the source of inputs driving accumulators with the observed discharge rates from FEF visually responsive neurons (Purcell et al., 2010). Our new work assumes that this dynamically modulated representation of stimulus salience drives a network of multiple competing accumulators associated with alternative stimuli in a visual search display. The transformation from salience to saccades is controlled by tonic gating inhibition. Unlike our earlier work developing and validating this modeling technique, our new model incorporates multiple accumulators and explains choice probabilities and full correct and error response time distributions as a function of set size during visual search. We identify the accumulators with FEF movement neurons that select and initiate saccades when threshold is reached. Accumulator dynamics correspond quantitatively...
to movement neuron dynamics. We show that gating and lateral inhibition are necessary by ruling out alternative models that do not include these mechanisms. We also show how the gating inhibition provides a novel mechanism to control trade-offs in speed and accuracy.

Materials and Methods

Behavioral and physiological data were collected from two adult male macaques (Macaca radiata) trained to perform a visual search task in which the number of items in the display (set size) varied randomly across trials (Fig. 1A, inset). Basic behavioral and physiological analyses of these data have been published previously (Woodman et al., 2008; Cohen et al., 2009; Heitz et al., 2010). Briefly, after fixating a central stimulus for ~600 ms, a stimulus array appeared containing a target (T or L) among one, three, or seven distractors (random 90° orientations of the other letter). The eccentricity of the array was always 10°. When the array contained two objects, they were always presented at opposite locations. When the array contained four items, they were separated by an unfilled location. Thus, stimuli were spaced ~7.7°, 14°, 19°, or 20° apart in the visual field. Target identity was varied across sessions. Monkeys were rewarded for making a single saccade to the location of the target within 2000 ms and fixating it for 1000 ms. Search performance was qualitatively similar across monkeys. Therefore, we pooled data across monkeys in addition to fitting data to individual monkeys.

Neurons were classified using a memory-guided saccade task. A single target (filled gray disk) was presented in isolation for 100 ms. Monkeys were trained to maintain fixation for 400–800 ms after target onset. When the fixation spot disappeared, the monkey was rewarded for making a saccade to the remembered location of the target. Neurons with a consistent visual response (<100 ms following target onset and persistent firing during the delay period) were classified as visually responsive neurons (59 neurons; monkey Q, 40 neurons; monkey S, 19 neurons) and were used to derive input to the stochastic accumulators in the model. This included both visual neurons with no presaccadic discharge and visuomovement neurons with both visual and presaccadic discharge. We included both pure visual and visuomovement neurons as input to the model because previous work has shown that visuomovement neurons, like pure visual neurons, represent the salience of objects in their receptive field (RF) (Thompson et al., 1996; Cohen et al., 2009), but do not have activity sufficient to directly control saccade initiation (Ray et al., 2009). Neurons with a brisk presaccadic discharge, but little or no visual response, were classified as movement neurons (22 neurons; monkey Q, 15 neurons; monkey S, 7 neurons) and were compared qualitatively and quantitatively to model accumulator dynamics.

Simulation methodology

Neurophysiologically derived representation of visual salience. Our model assumes that output from pools of visually responsive neurons encoding stimulus salience converge on movement neurons that initiate saccades. Observed spike trains of visually responsive neurons define the model input to stochastic accumulators identified with movement neurons. Here we describe how those observed spike trains are converted into model input on each simulated trial. In the next section, we describe the architecture of the network of accumulators that integrate that model input over time to generate a saccade to a particular stimulus in the visual search array at a particular time. Predicted proportions of error and correct responses and their associated response time distributions are generated by simulating thousands of trials.

Model input was derived directly from the spike trains recorded from visually responsive neurons following previously described methods (Fig. 2) (Purcell et al., 2010). For each simulated trial, we randomly sampled, with replacement, N spike trains from individual analyses and generated a spike density function for each trial by convolving the spike trains with a function of the form as follows:

\[ y(t) = (1 - e^{-t \tau_{r}}) \times e^{-t \tau_{d}} \]

which mimics the postsynaptic influence of each spike (Thompson et al., 1996). The growth constant, \( \tau_{r} \), is 1 ms, and the decay constant, \( \tau_{d} \), is 20 ms. The pool size, \( N \), was varied to identify the value that optimally predicted behavior (Table 1). Each single-trial spike density function in the sample was normalized to the maximum firing rate of the neuron and summed. The resulting input function is mathematically similar to a Poisson shot noise process (Smith, 2010).

Spike trains from all neurons were sorted according to the set size (2, 4, or 8) presented and response (correct or error) made when they were recorded. Trials in which the animal prematurely broke fixation were excluded (<1%). Spike trains recorded from a particular set size condition were used to generate behavioral and neural predictions for that set.

![Figure 1](https://example.com/fig1.png)
size. The observed choice probabilities for the set size being simulated determined the proportion of simulations that used spike trains sampled from correct or error trials. Thus, the spike trains that drive the model represent the responses of visually responsive neurons recorded under the exact conditions being simulated. Simulations that used input from trials in which the animal erroneously looked to a distractor most often predicted erroneous responses due to the form of sampled input (Heitz et al., 2010), but this was not always the case. Indeed, we present examples of models that fail to adequately predict error probabilities.

Post-saccadic spikes cannot contribute to the saccade choice and must be excluded. However, simply eliminating those spikes would artificially increase noise as the simulation progressed because different trials terminate with different RTs. Instead, we extended each trial with firing rates generated according to a Poisson process with a mean rate determined by the spike rate 10–20 ms before the saccade. This proved to be a practical way to prevent the parameter optimization routine from failing when certain parameter values could not produce any RT or choice. Ultimately, these extrapolated spikes contributed very little to simulations of well fitting models because simulated trials ended within the observed RT range.

The use of trial-by-trial spike trains recorded while animals performed the visual search task distinguishes our modeling approach from other stochastic accumulator models in several important ways. First, simulations began well before the presentation of the stimulus array and proceeded until the response was initiated (Fig. 2B). Afferent delays were determined by visually responsive neuron physiology and efferent delays were short and well established quantities (Scudder et al., 2002), so no “nondecision time” parameter was necessary. Second, this approach

![Figure 2](image-url)
takes into account the heterogeneous nature of neuronal response properties. Whereas many models assume highly simplified, prototypical, or idealized neural response function, our model is subject to the heterogeneity inherent in neuronal populations. Third, most models make specific assumptions about the mathematical form of the input to stochastic accumulators (Churchland et al., 2008; Furman and Wang, 2008; Smith and Ratcliff, 2009). Here, we make no prior assumptions about the form, timing, and variability of inputs within and across trials and conditions. Instead, the form, timing, and variability are given by the pattern of activity of visually responsive neurons representing visual salience. Although this approach does not speak to the mechanisms that generate salience (Tsotsos et al., 1995; Itti and Koch, 2001), it provides strong constraints on the mechanism by which salience is used to generate a saccade. We see this as a crucial first step to constrain new models that explain how the neural representation of salience is generated.

**Stochastic accumulation of visual salience.** The physiologically defined inputs are pooled and integrated by a network of leaky, competing accumulators (Fig. 2A). Trials in which a target was in the RF of a visually responsive neuron were used as input to the accumulator unit encoding the location of the target. Trials in which distractors were in the RF of a visually responsive neuron were used as input to units encoding the locations of distractors. Trials in which the target appeared near the RF edge were excluded.

The network always consisted of eight accumulator units regardless of the specific set size being simulated. For set sizes 2 and 4, input to an accumulator representing a location without a stimulus was defined by visually responsive neuron firing rates on trials in which no stimulus appeared in the cell’s RF. Thus, the model requires no reconfiguration across set size conditions and can generalize naturally to situations with more than eight stimuli. We simulated neural activity from 300 ms before stimulus presentation until the initiation of the saccade. A simulated choice was made to the location represented by the first accumulator unit to reach a constant threshold (Fig. 2B, bottom). Simulated saccade RT was defined as the time when the threshold was crossed plus a fixed ballistic time of 15 ms, which accounts for the time necessary for brainstem processes that shift gaze (Scudder et al., 2002).

Model parameters define for all accumulators the threshold (θ), leakage (k), gating inhibition (g), feedforward inhibition (u<k>), and lateral inhibition (βg). Lateral inhibition varied as a function of distance between locations, but was symmetrical between any two locations. Critically, no parameters varied across set size or with the choice of saccade target or distractor. Therefore, variability in search performance was directly determined by variability in the visually responsive neuron representation of salience and how that representation affects the dynamics of the stochastic accumulator network. Thus, the model assumes that set size effects can be understood as a consequence of changes in the salience representation integrated by stochastic accumulators. It is possible that models assuming architectural or parametric differences across set size could also predict the observed behavior, but such models would provide a less parsimonious account of the data.

Each accumulator was governed by the following stochastic differential equation:

\[
dm_{i}(t) = \frac{dt}{\tau} \left[ \left( v_{i}(t) - \sum_{m \neq i} u_{m}v_{m}(t) - g \right) + \frac{d_{i}}{\tau} \xi \right],
\]

where \( m_{i} \) is the mean activity of the accumulator unit representing location \( i \) and \( v_{i} \) is the input from visually responsive neurons representing the same location. As with other neural stochastic accumulator models (Usher and McClelland, 2001), we rectified \( m_{i} \) to be >0 because firing rates cannot be negative. Interactions between units were determined by symmetrical feedforward, \( u_{m} \), and lateral, \( \beta_{g} \), inhibition weights (Usher and McClelland, 2001), where \( d \) indexes the distance between two stimuli in the array (Fig. 2A, inset; 4 total values corresponding to approximately \( d = 7.7^\circ, 14^\circ, 19^\circ, \) and \( 20^\circ \) of visual angle separating the two stimuli). \( k \) is the leakage constant, \( g \) is the gating constant, and \( + \) denotes rectification above zero. All simulations used an integration time step of \( dt = 5 \) ms. All models included a modest amount of Gaussian noise intrinsic to the accumulator, \( \xi \), with mean, \( \mu = 0 \), and SD, \( \sigma = 0.05 \); this variability is quite small relative to variability in the visual inputs. Initial model exploration showed that models explain search performance with \( \sigma = 0 \), but we include some intrinsic variability to allow for tonic neuronal drive when salience input falls below the level of the gate. We also found that models that assume tonic excitation (\( \mu > 0 \)) in addition to salience inputs can predict elevated baseline firing rates as observed in our movement neuron sample. Note, however, that movement neurons with little or no tonic firing rate are commonly observed (Paré and Hanes, 2003).

**Gating inhibition.** Effective visual search performance requires that saccades be withheld until visually responsive neurons have processed stimulus salience. The model solves this problem with a combination of gating and lateral inhibition. Computationally, \( g \) is a tonic inhibition applied to the salience input, which is rectified to be >0 after the inhibition is applied (Eq. 2). Functionally, gating imposes a firing rate that must be exceeded before the inputs from visually responsive neurons begin to influence the accumulators. All models assumed that gating is independent of set size and stimulus strength. We assumed a tonic inhibition for computational simplicity and because it could be easily implemented by neurons (see Discussion, Neurophysiological implications). The precise form of nonlinearity mediating the transformation from visually responsive neurons to accumulator units is probably not critical, so long as accumulator units have little or no response until the inputs exceed some critical level. Alternative nonlinearities, such as sigmoid or power functions, may also work if the slope or acceleration is sufficiently steep (Simen et al., 2006; Simen and Cohen, 2009).

Figure 2B illustrates how gating inhibition and lateral inhibition mediate the transformation of salience into response accumulator dynamics. Drive from visually responsive neurons exists even before the search array appears, but no accumulation takes place because baseline input firing rates are below the gate. By 100 ms after array presentation, visually responsive neuron firing rates exceed the gate, but top-down salience has not yet emerged. Salience for targets and distractors is equal and balanced; lateral competition holds the system at a temporary equilibrium. After 150 ms, salience for the target exceeds salience for distractors, which triggers a rise in the accumulator unit encoding the target location. Thus, gating and lateral inhibition explain how the selection of salient saccade targets by visually responsive neurons is transformed into a competition for saccade execution in movement neurons.

**Alternative model architectures.** We evaluated several different model architectures by fixing various subsets of the parameters in Equation 2 to zero and varying others to fit search performance. Different model architectures instantiate competing hypotheses about the processes that lead to the observed response. We compared models on their quality of fit to search performance (see below, Fits to visual search performance) and their ability to predict the time course of movement neuron activity (see below, Measures of movement neuron and accumulator dynamics). Models that failed to fit the observed pattern of behavior, that fit significantly better or worse than alternative model architectures, or that failed to predict movement neuron dynamics were rejected. This application of strong inference (Platt, 1964) enabled us to systematically eliminate several plausible neural architectures for using salience to make saccades.

The best-performing model architecture across all of our criteria was the gated competitive model (see Figs. 4–7, 12). This model was defined by setting \( u_{g} = 0 \) (no feedforward inhibition) and included gating inhibition, leakage, and lateral inhibitory connections between units (8 free parameters). The term “gated accumulator model” refers to this particular architecture unless otherwise specified.

We evaluated two alternative model architectures that did not assume competitive interactions. The gated race model (see Fig. 9A) was defined by setting \( u_{g} = 0 \) and \( \beta_{g} = 0 \); this model included gating inhibition and leakage, but no competition (4 total free parameters). The gated diffusion-like model (see Fig. 9B) was defined by setting \( u_{g} = 0 \) and \( \beta_{g} = 0 \); this model included gating inhibition, leakage, and feedforward, but not lateral, inhibitory connections (8 total free parameters). We also evaluated a model with both diffusion-like and competitive interactions (12 total...
free parameters), but the fit was not significantly improved relative to the gated competitive model.

We evaluated two other model architectures without gating inhibition. The non gated, leaky model (see Fig. 9C) was defined by setting $u_d = 0, k = 0, and g = 0; this model included lateral inhibition, but integration of the salience occurred without loss (6 total free parameters). The non gated, leaky model (see Figs. 9D, 10) was defined by setting $u_d = 0$ and $g = 0$, and included leaky integration and lateral inhibition, but no gating inhibition (7 total free parameters).

Fits to visual search performance. We identified parameter values that best fit search performance. We simulated 5000 trials to generate RT distributions and response probabilities for each set size. All models were fit to behavioral data using a custom combination of a genetic algorithm and response probabilities for each set size. All models were estimated (Efron and Tibshirani, 1993; Wichmann and Hill, 2001). We accounted for variation in RT within each set size, we computed the slope of the average measure as a function of set size, as opposed to specific details likely to be idiosyncratic to particular neurons. Thus, the simulated model dynamics are true predictions and not optimized fits to physiology. As predictions, we focus on the general changes in response dynamics across and within conditions, as opposed to specific details likely to be idiosyncratic to particular neurons.

Receiver operating characteristic analysis. We assessed how reliably FEF neurons signaled the target location by computing the area under the receiver operating characteristic (ROC) curve from the distribution of trials in which the target appeared inside the receptive field of the neuron and trials in which the target appeared outside the receptive field (Thompson et al. 1996). The ROC was computed by incrementing a criterion from 0 spikes/s to the maximum firing rate observed across all trials in steps of 1 spike/s. For this analysis we measured the average area under the ROC curve from 100 to 200 ms after array presentation.

Speed–accuracy trade-off simulations. We explored how the gated competitive accumulator model could accomplish speed–accuracy trade-offs by systematically varying two key parameters of the model. We computed the mean RT, percentage correct, and expected reward rate while varying the gate or threshold parameters and fixing all other parameters at the values fitted to pooled performance data. Each simulation consisted of 700 trials of each set size corresponding approximately to the
average number of trials in an actual recording session. Reward rate is defined by $r = P/T$, where $P$ is the average proportion of correct choices across all set size conditions and $T$ is the average trial duration (Gold and Shadlen, 2002; Lo and Wang, 2006). $T$ was defined by the following equation:

$$\sum \sum P_x (RT_{x} + ITI),$$

where $p$ is the response proportion and $RT$ is the average RT for each set size, $s$ (2, 4, or 8), and response, $r$ (correct or error). The inter-trial interval, ITI, was 1 s as in the experiment protocol. Confidence intervals were computed from the distribution of reward rates obtained by simulating expected reward rate 1000 times for each set of parameters with randomly sampled inputs.

**Results**

**Visual search behavior and neurophysiology**

We modeled behavior and neurophysiology observed from macaque monkeys trained to perform an attentionally demanding visual search task (Fig. 1A), details of which have been previously published (Cohen et al., 2009; Heitz et al., 2010). Monkey behavior exhibited hallmarks of inefficient search as observed in humans. The percentage correct declined (Fig. 1A) and mean RTs increased (Fig. 1B, C) with increasing set size (Cohen et al., 2009). Monkeys also exhibited substantial variability in RTs within each set size (Fig. 1B, C, insets). The variance of the distribution increased with set size and there was a systematic lengthening of the upper tail of the distribution, which is also observed in human search performance (Ward and McClelland, 1989; Palmer et al., 2011). Current model of search explains full response time distributions (Wolfe et al., 2010) or the neural source of this variability. Our new model explains both the systematic and random variability in RTs using single-unit physiology recorded during visual search.

The first observation guiding our modeling is that visually responsive neurons in FEF, SC, and lateral intraparietal area (LIP) appear to encode the salience of potential saccade targets. All single-unit data in the present report were collected in FEF, but we note that neurons with very similar response properties are observed in other sensorimotor areas. Thus, although we used the discharge rate of visually responsive FEF neurons as the model input, we believe that similar results would be observed using discharge rates from visually responsive SC or LIP neurons. During the visual search task, the firing rates of visually responsive neurons increase following the onset of the search array and are maintained at a higher level when the target is within the cell’s RF (Fig. 3A). The average time of target selection increases with set size and the average firing rate decreases with set size (Cohen et al., 2009). Decreases in the average firing rate with set size have also been reported in LIP (Balan et al., 2008). In addition, we found that visually responsive neurons signaled the location of the target less reliably as set size increased, as indicated by a decline in the average ($\pm$SE) area under the ROC curve (set size 2, 0.63 ± 0.05; set size 4, 0.55 ± 0.03; set size 8, 0.54 ± 0.03). All differences were significant (paired $t$ test, all $p < 0.05$). These changes may reflect the limited capacity of the visual system to simultaneously analyze all stimuli in the visual field, although we cannot make strong claims about the specific mechanisms responsible for this limitation. Importantly, the decrease in firing rate with set size is consistent with the identification of these neurons as encoding visual salience. A model that explains how salience is translated into a saccade command must explain how these firing rate dynamics are read out to initiate a saccade.

The second observation guiding our modeling is that movement neurons in FEF and SC initiate saccades at a fixed firing rate threshold. Note that LIP has very few movement neurons without visual responses (Gottlieb and Goldberg, 1999) and LIP neurons do not project directly to oculomotor brainstem nuclei. Therefore, we do not identify our model accumulator with LIP neurons. Figure 3B illustrates the population response of FEF movement neurons during the T/L search task. In contrast to the firing rates in the visual neuron population, the firing rates of these neurons converge at a fixed firing rate threshold immediately before the saccade regardless of the set size and RT. This adds to converging evidence that presaccadic movement neurons initiate a saccade when a fixed firing rate threshold is reached (Hanes and Schall, 1996; Hanes et al., 1998; Paré and Hanes, 2003). A model that explains how stimulus salience is translated into a saccade command must also explain how presaccadic movement neurons are driven to threshold.

**Gated competitive accumulation explains search performance**

The gated accumulator model quantitatively accounts for all aspects of visual search performance. We optimized parameters to simultaneously fit the observed choice probabilities and RT distributions for both correct saccades to the target and error saccades to distractors (Table 1). Figure 4 compares observed search performance with simulated performance of the network. The
model accounts for the increase in mean RT with set size (Fig. 4A) because the average firing rate of visually responsive neurons declines with set size (Fig. 3A) and because neurons encoding competing locations are increasingly active due to the presence of distractors in their RF. In other words, as set size increases, more locations become activated on the salience map. The model accounts for the decrease in percent correct with set size because increased activation at competing locations decreases the reliability with which the target can be discriminated and because visually responsive neurons select distractors when the monkey makes an error (Thompson et al., 2005a; Heitz et al., 2011). Predicted errors are slow relative to correct responses because firing rates are lower, on average, on error trials. Thus, the properties of visually responsive neurons coupled to a network of stochastic accumulators have exactly the properties necessary to explain the basic properties of search performance. Although we only manipulated set size here, we note past results showing that visually responsive neurons encode a stimulus similarity as expected of a salience representation (Bichot and Schall, 1999; Ipata et al., 2006; White and Munoz, 2011). This is exactly the pattern of modulation that would be required for our model to predict the observed decreases in performance with increased target–distractor similarity, as we have demonstrated previously (Purcell et al., 2010).

The model explains the basic changes in mean RT and accuracy during inefficient visual search, which must be true for any viable model of search, but a more rigorous evaluation of model performance would determine whether the model can predict full RT distributions. Figure 4 plots the observed and predicted RT distributions for correct (Fig. 4C) and error (Fig. 4D) trials for each set size condition. The model captures the increasing spread and the systematic lengthening of the upper tail of the distribution, which is characteristic of inefficient visual search in humans and our monkeys (Wolfe et al., 2010; Palmer et al., 2011). We verified that this was also true for individual monkeys (Fig. 5). These results demonstrate that nearly all of the variability in search performance can be explained by a gated accumulation of the neural representation of stimulus salience.

Gated competitive accumulation predicts dynamics of presaccadic movement activity

We identify model accumulator units with movement neurons in FEF and SC. If this mapping is correct, then accumulator dynamics should correspond to movement neuron dynamics observed during visual search. Figure 6A–C plots observed movement neuron spiking dynamics for slow, intermediate, and fast responses within each set size condition. The onset of neuronal modulation is later, on average, when the monkey responds slower whether within or across set size conditions (Woodman et al., 2008), but there is little change in the rate of growth across set size when aligned on saccade initiation. The apparent differences in rate of growth when aligned on array onset are merely a consequence of averaging across trials with different RTs. Consistent with a fixed response threshold, the presaccadic firing rate remains constant regardless of task difficulty or response speed. The onset of activation also increases within each set size condition and there is a slight decline in the rate of growth to threshold.

Figure 6D–F plots the average simulated trajectories sorted according to the same procedure as the movement neuron data. We simulated each set size condition 5000 times and generated spike trains according to a Poisson process with a nonhomogeneous rate given by the simulated model trajectories (see Materials and Methods). The model qualitatively captures key physiological observations in the movement neurons. The time when the accumulated activation first modulates above baseline increases across and within set size. When aligned on saccade initiation, there is little or no change in the rate of growth across set size and only marginal change in the rate of growth within set size. Finally, the model reaches a fixed threshold before the response by design.

Following Woodman et al. (2008), we quantified four measures of neural dynamics to verify the correspondence between accumulators and movement neurons: delays in the onset of the presaccadic rise (onset), the average rate of rise (growth rate), the starting point (baseline), and the firing level at the time of saccade (threshold; see Materials and Methods). All simulations used pa-
parameters fitted to search performance (Table 1). Critically, no parameters were adjusted to generate predicted model dynamics.

First, for the across set size assessment, we computed the average of each measure across RT groups and computed the least-squares regression slope of that measure as a function of set size. Figure 7A–D shows the resulting slopes for a representative movement neuron. Only the onset of the presaccadic burst correlates significantly with set size. Figure 7E–H shows the resulting slopes for the population of all movement neurons (bar histogram) and simulations of the gated competitive accumulator model (line histogram). Across the movement neuron population, the median slope of onset as a function of set size is significantly >0 (14.4 ± 2.5 ms/item; Wilcoxon rank sum test, p < 0.001), but does not differ significantly from the median slope of model predictions (14.9 ms/item, p = 0.47). No other slope distribution was statistically different from zero or from model predictions (all p ≥ 0.05). Thus, the model predicts changes in neural dynamics with changes in set size comparable to observed movement neuron dynamics.

Next, for the within set size assessment, we computed the correlation between each measure of neural dynamics and RT within each set size independently. Figure 8A–D illustrates the correlation between each measure of neural dynamics and RT bins from a representative neuron recorded during set size 4 trials. Only the onset of the presaccadic burst correlates significantly with response time (Fig. 8A), although there is a small, but non-significant, decrease in the rate of growth (Fig. 8B). Figure 8E–H plots the median observed correlation between each measure of neural dynamics and RT across the population of movement neurons for all set sizes. The median correlation is significantly >0 for the onset of the presaccadic burst (Fig. 8E; Wilcoxon rank sum test, set size 2, 0.67 ± 0.12; set size 4: 0.81 ± 0.05; set size 8, 0.93 ± 0.02; all p < 0.01). We also observed a weak correlation between baseline and RT in set size 2 that was marginally significant (Fig. 8G; set size 2, −0.24 ± 0.07, p = 0.04), but no other measures of neural dynamics correlated significantly with RT (all p > 0.05). Importantly, the strength of correlation with RT was statistically indistinguishable between movement neurons and model predictions for most measures of neural dynamics (all p > 0.05). The only exceptions were the median correlation between baseline and RT for set size 4 (data, 0.03; model, −0.29; p = 0.02) and between threshold and RT for set size 4 (data, 0.12; model, −0.17; p = 0.03), which differed slightly, but significantly, from the data. However, the lack of a systematic relationship with set size suggests that this deviation can be attributed to noise. Thus, the model predicts the essential quantities seen in the neural dynamics using the same parameters that were optimized to fit search performance.

Gating inhibition and lateral competition are necessary

Using the logic of nested model testing, we determined which model elements are necessary to account for visual search performance and replicate movement neuron physiology. First, we explored a gated race model architecture without competitive interactions (all $\beta_d = 0$). This architecture qualitatively failed to predict the ordering of mean error RTs (Fig. 9A). This architecture failed because it has no mechanism to suppress the initial visual responses that drive saccades to distractors. Quantitatively, this produced significantly elevated $\chi^2$ relative to the gated competitive model for both individual and pooled monkey data ($p < 0.05$, nonparametric bootstrap, 5000 simulations; Fig. 10). The difference was not statistically significant using AIC (all $p > 0.05$), but there are several reasons to believe that this additional complexity is warranted. First, the gated race model cannot predict the correct ordering of error RTs and therefore qualitatively fails to predict a key component of visual search performance. Second, the need for competition is supported by studies demonstrating that microstimulation of FEF can suppress firing rates in the opposite hemisphere (Schlag et al., 1998) and that the presentation of a visual distractor in the ipsilateral visual field reduces the firing rate of SC movement neurons (Dorris et al., 2007). Last, existing theoretical work has demonstrated some form of competitive interaction is necessary to optimize the rate of reward for a set amount of sensory evidence (Bogacz et al., 2006). Thus, some form of competitive interaction appears to be necessary.

We next explored a diffusion-like gated accumulator model with feedforward inhibition between model inputs but no competition between model units (all $\lambda_w$ were allowed to vary, but all $\beta_d = 0$). This architecture failed to predict both the ordering of error RTs and the observed percentages of correct responses (Fig. 9B). We observed a significantly worse fit for this model than the gated competitive model for all datasets whether or not model complexity is taken into account (all $p < 0.05$; Fig. 10). This model fails because some mechanism must suppress the initial visual responses to distractor locations, but feedforward inhibi-
tion must necessarily be weak when the number of interacting accumulators is >2. If feedforward inhibition is strong, then the inhibition from distractors will overwhelm excitation to the target location, resulting in excessively long RTs (Usher and McClelland, 2001). These results appear to challenge the generality of a large class of drift diffusion models that require feedforward inhibition for their neurophysiological implementation (Gold and Shadlen, 2001; Bogacz et al., 2006; Ratcliff et al., 2007; Churchland et al., 2008; Ditterich, 2010; Purcell et al., 2010).

To investigate the necessity of gating inhibition and leakage, we explored a nongated, nonleaky model architecture with lateral inhibition but no gate or leak. This architecture failed to quantitatively predict both correct and error RTs as well as the percentage of correct responses (Fig. 9C). The fit was significantly worse than that of the gated competitive model even when taking into account differences in model complexity (Fig. 10). It fails because tonic activity of visually responsive neurons propagates directly to accumulators that initiate saccades before salience can be computed. Thus, some mechanism is needed to limit the rate of integration until an informative salience signal can be generated. Leakage, gating, or both leakage and gating are each viable mechanisms for limiting the rate of integration.

To investigate the necessity of gating inhibition, we explored a nongated, leaky architecture with lateral inhibition and leakage, but no gating inhibition. This architecture accounted qualitatively for the basic pattern of search performance (Fig. 9D). Although the fit appears qualitatively worse than that of the gated competitive model architecture, this difference was not statistically significant for both fit statistics and all datasets (Fig. 10). We previously found the same result using a simplified two-
accumulator model framework (Purcell et al., 2010). Thus, leaky and gated models cannot be discriminated by behavioral data alone even in this more complex network that explains richer behavioral data. However, gated accumulation is necessary for the accumulators to replicate movement neuron dynamics. Models that do not include gating predict a strong negative correlation between the baseline and RT that is not observed in the data. Figure 11 shows that this result replicates in a multiple-accumulator framework. Thus, gating inhibition is necessary to explain movement neuron physiology during visual search.

Control of speed–accuracy trade-off

Most stochastic accumulator models assume that adjustments in speed and accuracy are controlled via adjustments in the height of the response threshold (Gold and Shadlen, 2002; Bogacz et al., 2006; Lo and Wang, 2006; Nakahara et al., 2006; Simen et al., 2006; Ratcliff and McKoon, 2008). However, no current study has reported changes in threshold firing rate at the level of individual neurons. In fact, recent neurophysiological evidence indicates that systematic adjustments of RT are accomplished through changes in the onset of accumulation (Pouget et al., 2011). We found that the gated accumulator model produces a speed and accuracy trade-off when either threshold or gating inhibition is varied. Rather than fit parameters to data, as in the preceding section, we manipulated the level of the threshold or the gate while fixing all other parameters at values fitted to pooled search performance. When the threshold increases, the system is more likely to correctly locate and respond to the target, but requires more time to accumulate additional salience (Fig. 12A, inset). When the gate increases, visual salience is processed to a higher resolution before it begins to drive saccade initiation (Fig. 12B, inset). Thus, gating inhibition provides an alternative account for trade-offs between speed and accuracy during visual search.

Mechanisms that trade speed for accuracy should optimize behavior. As one measure of optimization, we computed the expected reward rate as a function of threshold and gate. Consistent with previous studies, we found that varying threshold produced a nonmonotonic change in the expected reward rate such that a range of threshold values maximized reward (Fig. 12A). However, we also found that systematically varying the level of the gate produced a nonmonotonic change in reward rate (Fig. 12B). Furthermore, the value of threshold and gate that best fit the monkey’s performance fell within the range of parameter values that maximized reward rate. This suggests that after extensive practice, the circuitry of the network responsible for visual search performance adapted precisely to optimize performance.

Monkeys could strategically adjust their performance via adjustments in threshold or gating, but the alternative mechanisms are indistinguishable in terms of both maximal reward rate and robustness to parameter variation. However, the alternative
A neurophysiologically constrained account of visual search

Visually responsive neurons in FEF, LIP, and SC have been identified as a neurophysiological salience map (Findlay and Walker, 1999; Thompson and Bichot, 2005; Fecteu and Munoz, 2006; Gottlieb, 2007; Bisley and Goldberg, 2010) because these neurons modulate their firing rate according to both physical properties of the stimuli in the display (Bichot and Schall, 1999; Ipata et al., 2006; White and Munoz, 2011), reflecting bottom-up salience, as well as variables that determine the significance of a stimulus, such as its expected reward (Ding and Hikosaka, 2006) or stimulus history (Bichot et al., 1996; Bichot and Schall, 1999, 2002; Sato et al., 2003), reflecting top-down salience. This salience representation does not depend on response or effector (Thompson et al., 1997, 2005b; Bisley and Goldberg, 2003; Sato and Schall, 2003; Gottlieb et al., 2005; Murthy et al., 2009; White and Munoz, 2011).

We used the salience representation provided by visually responsive neurons in FEF to investigate the mechanisms that translate salience into saccades. This is the first model to explicitly link the neural representation of salience to a multiple-alternative stochastic accumulator network. Some previous search models include stochastic accumulation, but these have only been formulated conceptually (Wolfe, 2007) or are not grounded in physiology (Ward and McClelland, 1989; Thornton and Gilden, 2007). We previously modeled correct search RTs using a simple two-unit accumulator network (Purcell et al., 2010). This preliminary model accounted for systematic variation of RT with search efficiency, but the dataset was insufficient to investigate error rates or set size effects; therefore, it was not a viable model of visual search. The behavioral and neurophysiological data modeled here impose substantially greater model constraint.

We found that gating inhibition is necessary to limit the accumulator input until explicit visual salience is present. Many models assume a static salience representation (Palmer et al., 2000; Navalpakkam and Itti, 2007), but the neural representation of salience evolves dynamically over time (Fig. 3A). In our model, saccade preparation cannot begin until the visual response exceeds the gate and an informative salience signal evolves (Fig. 2B). This mechanism contrasts with all previous stochastic accumulator models that require the ad hoc assumption that accumulation is simply delayed by an arbitrary nondetection time parameter (Lo and Wang, 2006; Ratcliff et al., 2007; Churchland et al., 2008; Stanford et al., 2010). Other work has demonstrated that balanced excitation and inhibition can delay the onset of accumulation (Furman and Wang, 2008), but such a system can be susceptible to perturbations that can disrupt stability, leading to premature threshold crossings.

Extending perceptual decision tasks to multiple stimuli

Several frameworks have been proposed to explain perceptual decisions about individual stimuli. Many such frameworks include discrete or continuous networks of accumulators (Usher et al., 2002; Beck et al., 2008; Churchland et al., 2008; Furman and Wang, 2008; Ditterich, 2010; Leite and Ratcliff, 2010). Some of these models refer to how visually responsive LIP and FEF neurons integrate the firing rates of neurons encoding visual features to categorize information from a single foveal stimulus (Roitman and Shadlen, 2002; Ding and Gold, 2011). However, a general theory of perceptual decision making must explain more complex conditions involving multiple stimuli, not just multiple responses. These models identify the same visually responsive neurons that encode salience with accumulators that signal the end of stimulus categorization when discharge rates reach a fixed threshold. However, the dynamics of these neurons are markedly different during search tasks where one or more stimuli appear abruptly in the RF of the neuron and must be categorized and selected to guide response selection. The initial visual response often exceeds the presaccadic firing rate, and these neurons do not reach a fixed threshold before saccade initiation (Fig. 3A) (Hanes et al., 1998; Brown et al., 2008). Therefore, we propose a
new framework to account for neural dynamics and search performance.

According to this framework, visual neurons in FEF, SC, and LIP have already integrated and reweighted sensory inputs according to task demands before being input to the accumulator network. This interpretation is consistent with cognitive models of perceptual decision making in which perceptual processing precedes the start of the accumulation process to determine the degree of support for a particular response (Nosofsky and Palmeri, 1997; Logan, 2002). This interpretation also highlights the distinct role of different populations of neurons in decision making. While the decision that a stimulus belongs to a particular category could be identified with selection of the target location by visually responsive neurons in FEF, SC, and LIP, the decision to act requires additional processing by movement-related neurons (Schall, 2001). Our model formalizes this distinction.

Strategic adjustments in speed and accuracy
Threshold adjustments seem to provide a simple and intuitive account of speed–accuracy trade-offs. However, we found that variation of gating inhibition can produce equivalent speed–accuracy trade-offs by either promoting or delaying the accumulation of salience as has been observed in another study (Pouget et al., 2011). This model predicts that changes in gating inhibition will produce a qualitatively different pattern of movement neuron dynamics than changes in threshold when tested during speed–accuracy trade-offs. This also suggests that a neuron that implements the gating inhibition will modulate its firing rate inversely with speed stress.

Other recent models have indicated that speed–accuracy trade-offs can be implemented by a time-varying multiplicative gain or “urgency” signal applied to the inputs to stochastic accumulators (Churchland et al., 2008; Cisek et al., 2009). Increasing noise and variance are intrinsic properties of our neurophysiologically derived model input (Fig. 2B). Our model predicts that this could take place via changes in the firing rates of visually responsive neurons without assuming an explicit multiplicative gain parameter. Altogether, our model makes clear predictions to discriminate alternative accounts of the neural basis of speed–accuracy trade-offs.

Neurophysiological implications
FEF, SC, and LIP neurons have anatomical connectivity to perform the functions proposed by our model. These areas receive inputs from diverse regions of extrastriate visual cortex and prefrontal areas to compute stimulus salience (Blatt et al., 1990; Lu et al., 1995; Schall et al., 1995). Neurons representing visual salience reside in both supragranular and infragranular layers of FEF (Thompson et al., 1996) and intermediate layers of SC (White and Munoz, 2011). Movement neurons reside in layer 5 of FEF and the intermediate layers of SC and project to oculomotor brainstem nuclei that control saccade initiation (Moschovakis et al., 1988; Segraves, 1992; Pouget et al., 2009). Visually responsive LIP neurons could drive SC and FEF movement neurons to threshold via known projections (Blatt et al., 1990; Paré and Wurtz, 2001). The constant presaccadic firing rate observed in FEF and SC movement neurons may correspond to the fixed degree of excitation necessary to tip the balance from gaze holding to gaze shifting via excitation of long-lead burst neurons and inhibition of omnipause neurons (Munoz et al., 2000; Scudder et al., 2002).

Our model assumes that a gating mechanism intervenes between visually responsive and movement neurons. Gating inhibition is a simple operation that can be realized through several different neural mechanisms. First, gating could be implemented by inhibitory interneurons within FEF and SC that can adjust the level of excitation necessary to initiate spiking in individual neurons (Somogyi, 1977; Markram et al., 2004). Simple neural networks that balance interneuron inhibition with self-excitation exhibit bistable dynamics that limit integration until inputs exceed a critical level (Simen and Cohen, 2009). Second, gating could be mediated by FEF and SC fixation neurons, which fire tonically during fixation but cease firing before a saccade at a time coinciding with the onset of movement neuron firing across conditions (Munoz and Wurtz, 1993; Everling et al., 1998; Hanes et al., 1998). Finally, gating inhibition could also be implemented via basal ganglia circuitry as posited by other models (Brown et al., 2004; Frank, 2006). FEF neurons project to the caudate nucleus, which sends inhibitory projections to the substantia nigra pars reticulata (SNpr) (Hikosaka et al., 2000). SNpr neurons cease firing around the time of target selection during search (Basso and Wurtz, 2002). SNpr sends direct projections to SC and to the mediodorsal (MD) and ventroanterior (VA) nuclei (Parent and Hazrati, 1995). MD and VA are thalamic nuclei with saccade-
related responses (Tankaka, 2007) and reciprocal connections with FEF (Huerta et al., 1987).

Conclusions

We showed that a feedforward cascade of salience into a network of multiple competing stochastic accumulators can account for visual search performance and replicate the dynamics of neuron activity related to presaccadic movement. By forgoing the details of spike generation in exchange for the ability to relate specific network mechanisms to comprehensive behavioral data, this approach allowed us to reject plausible model architectures and identify key mechanisms necessary and sufficient to explain how the brain decides among multiple stimuli. We can further test our model’s premise of a feedforward visual-to-motor cascade by using visually responsive neuron physiology recorded in the context of additional tasks as model input to our framework. Factorial manipulations of search efficiency and set size, variations in stimulus–response mapping, and tasks in which animals are cued to trade speed for accuracy will provide critical tests of the basic model assumptions.

References