

Spatial Attention and Temporal Expectation Under Timed Uncertainty Predictably Modulate Neuronal Responses in Monkey V1

Jitendra Sharma^{1,3,†}, Hiroki Sugihara^{1,†}, Yarden Katz², James Schummers^{1,4}, Joshua Tenenbaum² and Mriganka Sur^{1,2}

¹Picower Institute for Learning and Memory, ²Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 01239, USA,

³Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA 02129, USA and ⁴Max Planck Florida Institute for Neuroscience, Jupiter, FL 33458, USA

Address correspondence to Mriganka Sur, Simons Center for the Social Brain, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Room 46-6237, Cambridge, MA 02139, USA. Email: msur@mit.edu

[†]Jitendra Sharma and Hiroki Sugihara contributed equally to this study.

The brain uses attention and expectation as flexible devices for optimizing behavioral responses associated with expected but unpredictably timed events. The neural bases of attention and expectation are thought to engage higher cognitive loci; however, their influence at the level of primary visual cortex (V1) remains unknown. Here, we asked whether single-neuron responses in monkey V1 were influenced by an attention task of unpredictable duration. Monkeys covertly attended to a spot that remained unchanged for a fixed period and then abruptly disappeared at variable times, prompting a lever release for reward. We show that monkeys responded progressively faster and performed better as the trial duration increased. Neural responses also followed monkey's task engagement—there was an early, but short duration, response facilitation, followed by a late but sustained increase during the time monkeys expected the attention spot to disappear. This late attentional modulation was significantly and negatively correlated with the reaction time and was well explained by a modified hazard function. Such bimodal, time-dependent changes were, however, absent in a task that did not require explicit attentional engagement. Thus, V1 neurons carry reliable signals of attention and temporal expectation that correlate with predictable influences on monkeys' behavioral responses.

Keywords: hazard rate, neural responses, reaction time, top-down influences

Introduction

Attention is a flexible spatio-temporal conduit that biases perceptual processing and prioritizes behavioral responses. Spatially directed attention is known to improve detection of change in stimulus attributes, such as the contrast or orientation of an element (Posner et al. 1982; Desimone and Duncan 1995; Carrasco and McElree 2001; Gutnisky et al. 2009). Attention in the temporal domain also has direct implications for perception and behavior (Coull and Nobre 1998; Gallistel and Gibbon 2000; Nobre 2001; Tse et al. 2004). For instance, it has been shown that stimulus processing is improved by prior knowledge of stimulus location, as well as timing of its appearance (or disappearance) (Coull and Nobre 1998; Nobre 2001; Ghose and Maunsell 2002; Hayden and Gallant 2005; Lakatos et al. 2007). A closely related, but a less explored, internal construct is expectation, which may also have spatial and temporal implications. It is generated by likelihood of an event to take place and gets updated based on previous occurrences in space or in time (Mauk and Buonomano 2004; Oswal et al. 2007). The degree of temporal expectation of an event, such as the probability of change, manifests as a marked improvement

in behavioral performance, which scales with time. Expectation may thus improve detection accuracy and response timing, ultimately leading to behavioral optimization (Nobre 2001; Oswal et al. 2007). However, the relationship between attention and expectation and their relative influence on early sensory processing remains unclear and even disputed (Nobre et al. 2007; Summerfield and Egner 2009).

Our primary aims were 2-fold: To investigate whether single neurons in V1 were influenced by spatial attention; and whether neural responses varied in time with expectation of a change at the attended location in monkeys performing an attention task. Past studies have provided conflicting evidence regarding the influence of attention in primate V1. This is quite in contrast to higher visual areas such as V4 and MT, as well as parietal and frontal cortices, where single neurons readily exhibit robust attentional modulation (Goldman-Rakic 1995; Luck et al. 1997; Goldberg et al. 2006; Maunsell and Treue 2006; Squire et al. 2013). Several studies found minimal, if any, influence of spatially selective attention in primate V1 (Moran and Desimone 1985; Luck et al. 1997). Whereas Motter (1993) observed a significant increase in responses in about 25% of V1 neurons when monkeys attended to the cued location. Importantly, he also found a sizable population, roughly 10% of neurons, to exhibit suppression in responses with spatial attention. It appears that the key requirement is to employ tasks with sufficiently high attentional demand and an innovative use of stimulus configurations, specifically tailored to the properties of V1 neurons (Roelfsema et al. 1998; Vidyasagar 1998; Ito and Gilbert 1999; Sengpiel and Hubener 1999). In addition, consistently locating the target stimulus and distracters in close proximity, ideally within the receptive field (RF), may be critical, but difficult to achieve within the restricted size of the RF of V1 neurons (Luck et al. 1997; Klein 2000; Ling and Carrasco 2006; reviewed in Maunsell and Treue 2006).

Similar to spatial attention, neurons in hierarchically higher areas have also been shown to be influenced by the temporal flow of events that reflect internal state changes such as expectation or anticipation of an upcoming event. For example, neurons in the parietal cortex seem to follow a monkeys' subjective assessment of time in a manner which is specific to the learned schedules of the task (Janssen and Shadlen 2005). Similarly, responses in area V4 have been shown to correlate with monkeys' anticipation of a change in stimulus properties, borne out of previous training, and follow the hazard rate, or the conditional probability that an event would occur provided it has not happened already (Ghose and Maunsell 2002). However, neural underpinnings of expectation (or lack of it) in

V1 have been the subject of considerable debate (Sirotin and Das 2009; Kleinschmidt and Muller 2010; Vanzetta and Slovín 2010; Handwerker and Bandettini 2011). Sirotin and Das (2009) have claimed that hemodynamic signals, and not neural responses in V1, are time-locked to a monkey's anticipation of a visual stimulus. Others have argued that the mismatch may simply be a methodological artifact (Kleinschmidt and Muller 2010; Vanzetta and Slovín 2010). Regardless of whether neuronal activity correlates with hemodynamic responses, the question as to whether expectation or anticipation of an impending event (leading to a behavioral report) influences neural responses in V1, and whether attention plays a role in guiding temporal expectation, remains unresolved.

Our main challenge was to devise a behavioral task, which would allow us to have explicit control over spatial attention and temporal expectation. We hypothesized that the two could be disambiguated by sequentially manipulating the requirement to attend to the location of a target and then to the time when the target change was most likely to occur. As the target change occurred at an unpredictable time, the subjects had to implicitly track time. The key was to alternate high demand for attention to target location and then to time, that is, when the target change was expected to take place. In between was a period of relatively low attention demand that served to separate the 2 aspects of attention. Accordingly, we trained 2 monkeys in a covert attention task—while fixating at a central fixation spot they continuously monitored a peripheral attention spot. When it disappeared, they had to release a lever within a fixed time window to earn liquid reward. The time between attention spot appearance and its extinction was divided into 2: An initial “fixed” time window, in which the attention spot appeared and stayed on for a fixed duration, and a “variable” time window when the attention spot could disappear any time, prompting a behavioral response by the monkey. There were no cues given with regard to timing of fixed and variable periods, and monkeys had to make a subjective assessment of the elapsed time to correctly predict when the attention spot would extinguish. While the monkeys performed the task, we recorded from single neurons in V1 and subsequently analyzed their response dynamics. Our results show that change in V1 responses was significantly correlated with the task structure and monkeys' behavioral performance. Specifically, their reaction times (RTs) and performance improved with increasing trial duration. Neural responses showed a bimodal change—an early, moderate, but transient increase was followed by a sustained, late increase, beginning just before the time the monkeys expected a change to occur. Importantly, the late increase inversely correlated with RT, demonstrating that the early, largely spatial attention was later modulated by temporal expectation. To further confirm that bimodal response modulation was indeed due to time-dependent change in the attentional requirement, we recorded from the same area while the monkeys performed a simple fixation task. Here, no attentional engagement was necessary and the animals were rewarded for holding fixation within a window for fixed periods. We go on to show that in this task, there was no late, time-dependent increase in V1 responses, consistent with the lack of internally generated temporal expectation.

Methods

All experiments were performed under protocols approved by MIT's Animal Care and Use Committee and conformed to NIH guidelines.

Two rhesus monkeys (6–7 kg) underwent initial chair training and were fitted with titanium head-posts under anesthesia (Dragoi et al. 2002; Sharma et al. 2003). After allowing 6–8 weeks for wound healing and head-post stabilization, monkeys were initially trained in standard eye-fixation and color change detection tasks. Training in experiment-specific tasks (see details below) followed where monkeys were exposed to increasing levels of difficulty as their learning progressed. After several months of training, the monkeys were able to maintain fixation on a small red dot (0.1°) for up to 5 s within a tight fixation window of 0.5° (monkey AB) and 0.75° (monkey AT) and were able to complete 3000–4000 trials in daily sessions at better than 95% accuracy. Importantly, the monkeys were trained for much longer fixation durations and completed twice as many trials as were required for the actual experiments. This was done to a keep constant motivational level and stable performance at the time of data recording. An infra-red eye tracker (ISCAN, Inc., Burlington, MA, USA), with 240 Hz sampling at $\pm 0.1^\circ$ Resolution, was used for monitoring eye position. Monkeys were required to hold stable fixation throughout the trial; fixational instabilities $>0.25^\circ$ aborted the trial. The eye tracker was calibrated before each experiment using a five-point calibration procedure, in which the animal was required to fixate on each 1 of 5 points (1 in the center, 2 in the vertical, and 2 in the horizontal axes or the diagonals) in steps of 5° and 12° from the central fixation spot. Gains were adjusted for linear horizontal and vertical eye deflections. Fixation patterns were carefully analyzed to rule out any systematic biases and inconsistencies under different experimental conditions [e.g., Attend-Toward (Attend-To) or Attend-Away].

Experiments 1 and 2

Figure 1*a,b* shows the task structure and the temporal sequence of events. Trials were initiated when the monkeys pressed down on a lever and a red fixation spot appeared. Once they acquired stable fixation for 300 ms, an attention spot (green, 0.1°) and an identical, isoluminant, gray distracter spot appeared on either side of fixation spot. In the Attend-To condition, the attention spot appeared toward the neurons' RF, whereas it appeared at a matched location on the contralateral side in the Attend-Away condition. The attention spot could appear in 1 of 4 locations chosen randomly for each recording—2 on the side of the RF, separated by at least 90° polar angle; the other 2 were mirror symmetric locations (Fig. 1*c* shows one such configuration). Introducing randomization in location of the attention spot had 2 advantages: first, it helped normalize any location specific bias in attention modulation with respect to a neuron's RF; secondly, it increased the task difficulty, requiring greater attentional engagement by the monkey. In experiment 1, the attention spot remained on for a fixed period of 900 ms, after which it could disappear anytime between 900 and 2300 ms. Monkeys had to wait for additional 150 ms before releasing the lever (within 500 ms) to earn their liquid reward. The mandatory delay of 150 ms was introduced to discourage monkeys from making a reflexive release, thereby minimizing the influence of preparatory motor signals from affecting the neural responses. No other cues were given at any time during the trial and the monkeys had to learn and implicitly follow the trial structure. While the probability of attention-spot disappearance was distributed evenly over time, its conditional probability increased with time (Luce 1986; Ghose and Maunsell 2002). Since the attention spot stayed on for the entire trial duration, monkeys had to continuously monitor the spot and were allowed to break fixation only after releasing the lever, thus controlling for oculomotor influences on neural responses. In experiment 2, the protocol was similar with the exception that the time schedules were very different. The total duration of each trial was shortened to 1550 ms while the initial, fixed time window was longer and lasted 1250 ms instead of 900 ms, thus substantially compressing the variable period when the attention spot could disappear to 300 ms (1250–1550 ms). In experiment 2, we wished to test whether changing the fixed time window would affect the duration and magnitude of early, temporally limited attention modulation. We also wanted to assess the influence of duration of the “variable” period and thereby changing the slope of the hazard rate would bring a commensurate change in the late attention modulation. Ideally, this could be done keeping the maximum trial

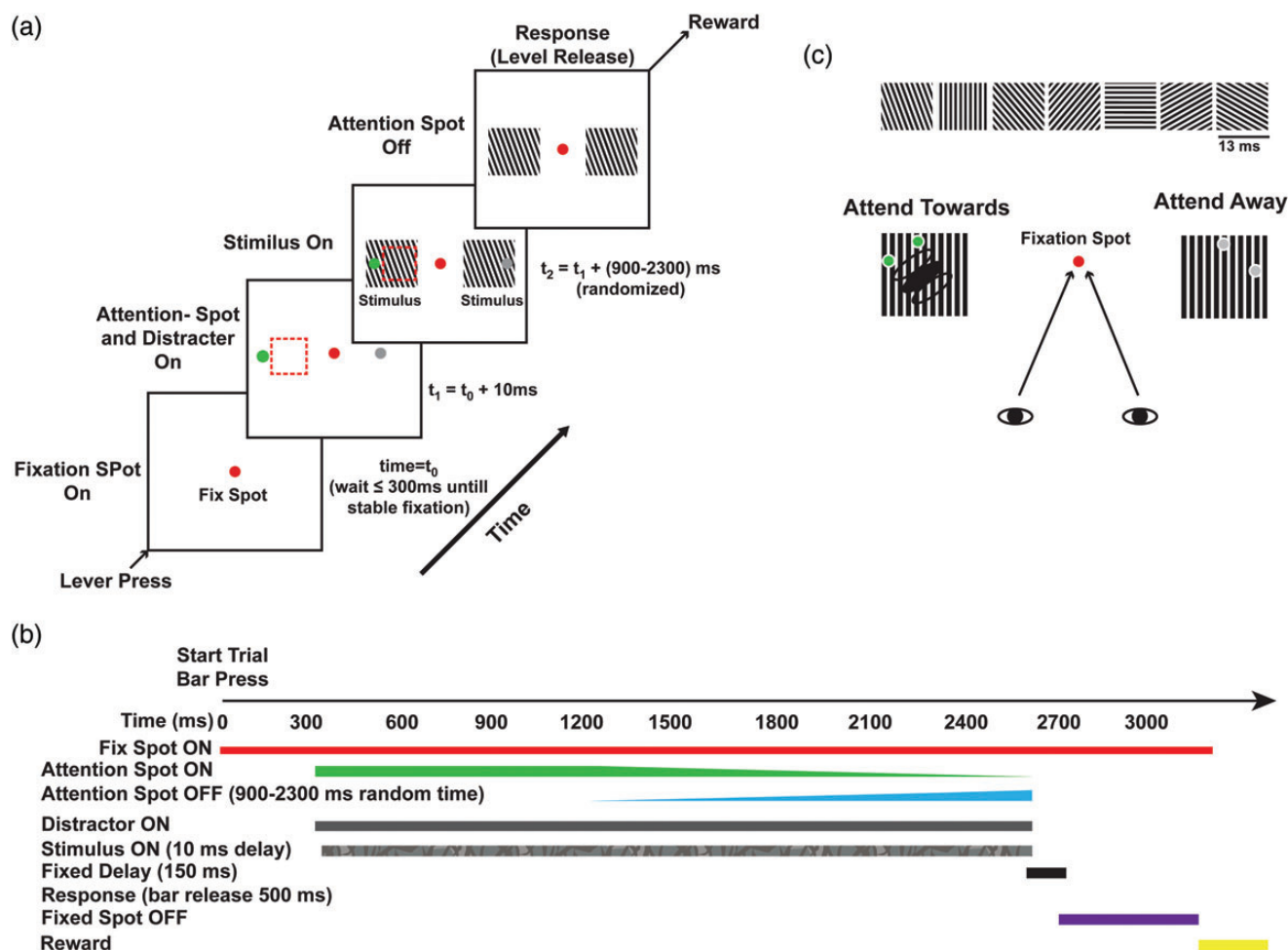


Figure 1. The behavioral task and RT measurements. (a) Schematic of behavioral task: Each frame represents a temporal element of the task in sequence. Monkeys initiated the task by holding down a lever when a red fixation spot appeared in the center at time t . Time (t_1) comprised a maximum period of 100 ms within which the monkey had to achieve fixation and hold it for 200 ms when an attention-spot (shown in green) and distracter (gray) appeared simultaneously, one toward the RF of a recorded neuron and the other in the opposite hemifield. Identical stimuli appeared on either side of the fixation spot within 10 ms thereafter. The trial continued for a “fixed” duration of 900 ms (t_2) followed by a “variable” duration (t_3) from 900 to 2300 ms, when the attention-spot was extinguished at anytime with equal probability over the duration. Monkeys waited for an additional 150 ms (t_4) before releasing the lever within 500 ms to earn a juice reward. (b) The top panel shows the sequence of events and their timing during the trial. (c) The stimulus consisted of patches of sinusoidal gratings, comprising 8 different orientations, each flashed for 13 ms in pseudorandom order but presented with equal probability during a trial. From trial to trial, the attention-spot and a distracter would appear at any location just outside the RF and at mirror symmetric location in the opposite hemifield. The data from all locations in both the Attend-To and Attend-away conditions were pooled for analysis.

length the same but increasing the fixed period and shortening the variable period. However, during training, monkeys were exposed to different lengths of “variable” schedules of the same trial duration. We therefore suspected that unless we changed the trial duration itself, just changing the variable period would not sufficiently change their previously learned behavior and might confound the results. The monkeys were also trained in durations much longer than were actually used for data recording to ensure consistent motivational levels. So, the option of increasing the trial duration also posed limitations. Shortening the total duration on the other hand, while increasing the fixed period and substantially compressing the variable period, we believe, achieved the same results as would have been done by keeping the overall duration constant, with an added advantage that monkeys were now exposed to an entirely different temporal structure of the task that had minimal overlap with the previously learned time schedules.

Stimulus Configuration for Attention Tasks (Experiments 1 and 2)

Stimuli were presented on a gamma-corrected monitor placed at a distance of 60 cm from the animal’s eyes. Two identical patches of gray

sinusoidal gratings, at 18% contrast ($2 \times 2^\circ$; 8 randomly interleaved orientations), were presented at the monitor refresh rate of 75 Hz (13 ms each frame; Fig. 1c)—one covering the RF, and another at a matched location in the opposite hemifield. Each trial cycle consisted of 112 stimulus presentations and 8 blank conditions, for a total of 120 conditions. To minimize the influence of physical attributes on neural responses, the attention spot was always placed outside the classical RF of the recorded neuron, yet close enough to elicit attentional influence in its vicinity. Similar configurations have been used previously for examining attention-related responses in V4 (Connor et al. 1997). The RF of the recorded neuron was first mapped with manual mouse-controlled stimuli and later with patches of randomly oriented gratings in a reverse correlation paradigm while the monkeys performed the fixation task. The extent of the classical RF was ascertained by minimum and maximum response field criteria. Stimulus patterns of either expanding center or converging annuli were placed at the center of the mapped RF—a significant decrease with expanding center or a significant increase with converging annulus gave a fairly precise estimate of the extent. At visual eccentricities of $4-8^\circ$ from the fovea, the average size of the RF’s encountered varied between 0.67° and 1.14° . To account for fixational instabilities from potentially making the

attention-spot drift in and out of the RF, thereby affecting the neural response, it was placed 0.25° outside the edge of the RF. As trials were automatically aborted if a monkey's fixation drifted >0.25°, this strategy effectively ensured that the attention spot never entered the RF. In addition, the attention spot appeared randomly in 1 of 2 locations, ≥90° (polar angle) apart outside the RF, and at matched symmetric locations in the opposite hemifield (also see Discussion).

Experiment 3

In this experiment, there was no attention condition. Trials commenced when a red fixation spot appeared in the center of the stimulus monitor. Monkeys were simply required to acquire fixation within 300 ms and hold it steady within the prescribed window till the spot disappeared after 1500 ms. Notably in this task, there were no attention cues or any behavioral report required. Once the monkeys acquired stable fixation, stimuli consisting of randomly oriented sinusoidal gratings (8 orientations; spatial frequency: 2 cycles/deg; temporal frequency: 1 Hz, at 28% contrast) covering the RF were presented. The question we wished to ask was whether temporal expectation-related late modulation of neural responses as seen in experiments 1 and 2 was contingent upon attention. In other words, by stripping the task from requirements of attention to time or to space, we could disambiguate the influence of attention from time-dependent changes in stimulus-evoked neural responses. During the initial recordings, we used either 1 or 2 identical stimulus patches on either side of the fixation spot, which was similar to the configuration used in attention tasks of experiments 1 and 2. However, as the responses for the single or the dual stimulus configurations were similar, we later switched to one stimulus condition for all subsequent recordings.

Neuronal Recording and Data Analysis

We made transdural recordings in area V1, used a Crist grid (Crist Instruments, Baltimore, MD) to advance tungsten microelectrodes (1–2 MΩ at 1 kHz; FHC, Inc., ME, USA) via stainless-steel guide tubes. Neural signals were recorded using the Multichannel Acquisition Processor system (MAP; Plexon, Inc., Dallas, TX, USA). Single units were amplified, filtered, and viewed on an oscilloscope, and heard through a speaker. Spike waveforms were sorted using an off-line spike sorter program (Plexon, Inc., Dallas, TX, USA) and later analyzed with custom software written in Matlab (Mathworks, Natick, MA, USA). Responses to individual frames were assessed with the reverse correlation technique and were averaged over the entire duration of a single trial (Ringach et al. 1997). The probability distribution of stimulus-evoked response was calculated after subtracting responses during the stimulus blank condition for each stimulus. Only the responses until attention-spot offset in each stimulus condition were used to make a composite response histogram for all trial durations.

Reaction Time and Attention Modulation Index Calculations

The RTs (in ms) were calculated by taking the difference between time of attention-spot offset (plus an additional fixed time of 150 ms) and release of the response lever. In Experiment 1, for example, the earliest lever release in a correct trial could occur only after 1050 ms after attention-spot onset; hence the RT distributions were computed starting at 1100 ms (see Fig. 2a). Change in RT (Δ RT) was analyzed as a function of trial duration. Influence of attention on neural responses was quantified by calculating an attention modulation index (AMI) by averaging responses of each neuron in 2 attention conditions for the entire trial duration, as: [Response (Attend-To – Attend-Away)/Response (Attend-To + Attend-Away)]. AMI varies between –1 and 1, where negative values signify suppression of responses and positive values indicate facilitation in the Attend-To condition when compared with the Attend-Away condition. AMI for individual neurons was averaged in 50 ms bins using a sliding window with 10 ms overlap to capture the dynamics of attentional modulation over the course of the trial. To test whether neural responses modulated by attention over the entire trial period exhibited a bimodal distribution—an early spatial attention modulation followed by late, temporal expectation—we used Hartigan's Dip test (Hartigan and Hartigan 1985). The Dip test measures maximum difference between the empirical distribution function and a

unimodal distribution to test for multimodality. The Dip test code in Matlab was adapted from the code originally written by F. Mechler (Mechler and Ringach 2002) and is available online.

Results

Two monkeys (AB and AT) performed a sustained attention task in experiments 1 and 2—they covertly monitored an attention spot that appeared in 1 of 4 possible locations and waited till it disappeared before releasing a lever within 500 ms to obtain a reward. Maintaining stable attention to a small spot of low to medium saliency presented within a patch of rapidly changing gratings of different orientations made the task attentionally demanding. Consequently, their average stable performance after several months of training was on the moderate side 72% (± 5.8 SD) for AB and 76% (± 4.1 SD) for AT.

Behavioral Data: Performance Measures

The monkeys also made significantly fewer performance errors in longer compared with shorter duration trials, suggesting progressively greater allocation of attention in the longer trials ($P < 0.05$; Student's t -test, for all errors including break fixation, early release, delayed release, and no release). Monkeys RT improved and performance errors systematically decreased with the length of the trials. Supplementary Figure 1 shows the histogram of RT and error rate in short, medium, and long duration trials for all data from both monkeys. Since the attention spot remained unchanged for a minimum fixed duration before a change occurred, at highly variable times, we examined whether behavioral performance reflected awareness of this temporal dichotomy. We found that monkeys tended to make far fewer errors during the initial “fixed” time window when compared with later in the trial when the transition was made to the “variable” period of attention-spot extinction. The “early release errors” in each trial session were calculated by taking the ratio of errors in the fixed and variable periods from the total errors. Of 5673 total errors in 53 sessions, there were 109 and 461 early release errors in fixed and variable periods, respectively (arcsin 0.17 and 0.3; $P < 0.0002$ two-tailed t -test, session-by-session average). This is consistent with the monkeys' awareness of elapsed time and their internal expectation of when the attention spot would disappear after trial onset.

Reaction Times and Model

RT data were analyzed from all sessions in which monkeys completed at least 3 trial cycles (typically 4–5 cycles) of 120 stimulus conditions per trial (Methods), and the same principal single units could be held for the entire duration. The mean RT (53 recording sessions: 32 from monkey AB, 21 from monkey AT, $n = 25$ 808 correct trials) showed strong inverse correlation with the trial duration ($r = -0.93$, $P < 0.001$, Spearman's test; Fig. 2a). This systematic decrease in RT with increasing trial duration indicates that monkeys were sensitive to elapsed time and implicitly tracked the likelihood of attention-spot extinction as it increased with each passing moment after a certain point in the trial. Consequently, their expectation of an impending response also increased with time, resulting in faster RTs. Similar expectation-related response changes have been shown in human psychophysics and nonhuman primate studies (cf. Coull and Nobre 1998; Nobre 2001; Oswal et al. 2007; Lima et al. 2011). Figure 2a also shows mean RT as a function of trial length in the Attend-Away condition; there was

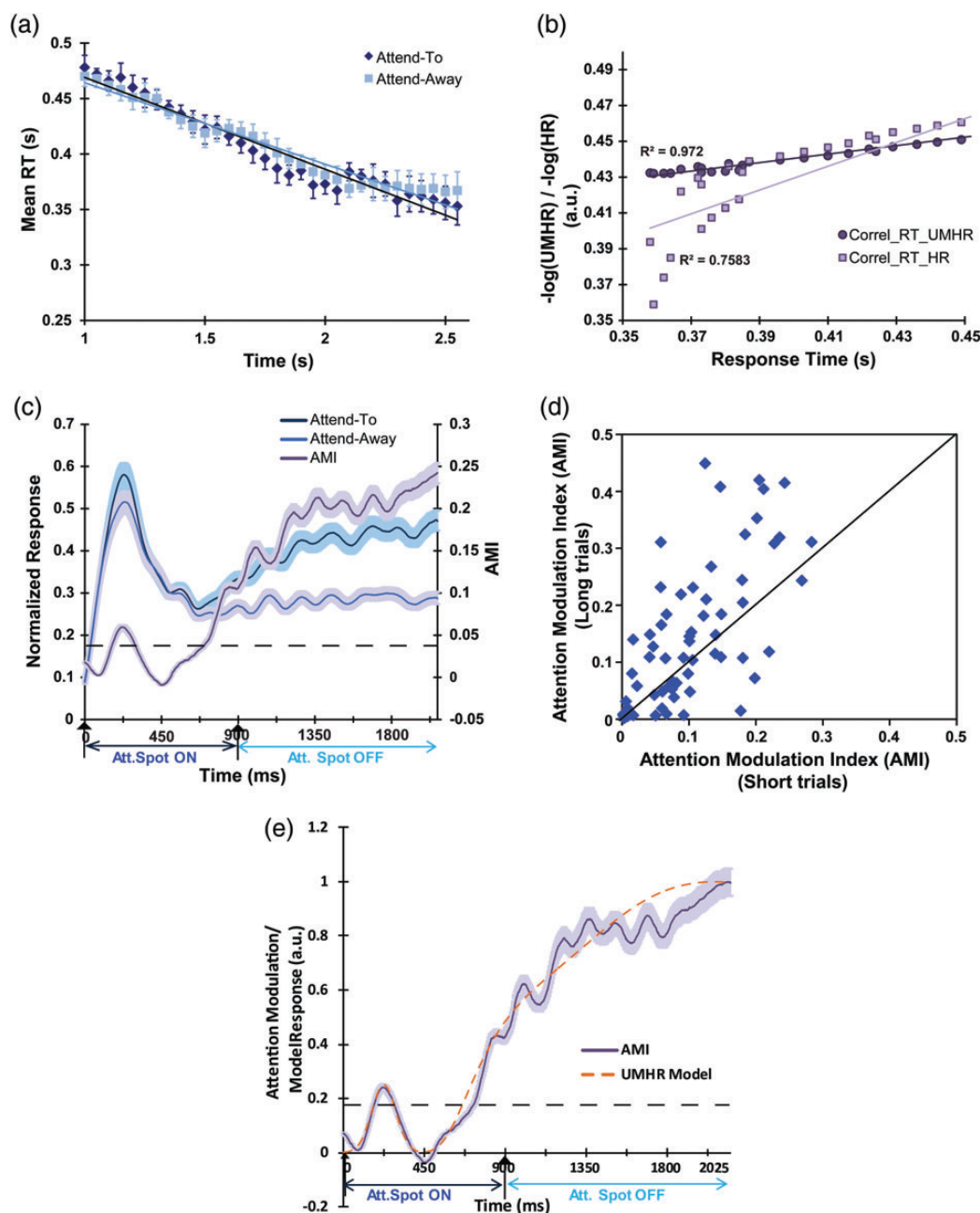


Figure 2. Behavioral data and neuronal data recorded from V1 in 2 monkeys during task performance in experiment 1. All data for experiment 1 in this and subsequent figures are shown in shades of blue. (a) RT decreased with increasing trial length. All RTs falling within a 50-ms time bin were pooled and averaged. There was no significant difference between monkeys' RT in Attend-To (dark blue) or Attend-Away conditions (light blue). Error bars denote mean \pm SD. The continuous line is the linear least-squares fit to the data ($R^2 = 0.945$). (b) Correlation between monkeys' RT data with model simulations for the uncertainty modulated hazard rate model (UMHR, filled circles) and the standard hazard rate model (HR, open squares). For calculating $h^*(t)$, actual trial times of 50 randomly chosen trials from each of 32 behavioral data sets were averaged. The continuous line is the least-squares fit to the model data ($R^2 = 0.972$). In comparison, the correlation between RT and standard hazard model data, though highly correlated ($R^2 = 0.758$), is significantly less so compared with the UMHR model ($P < 0.001$, Spearman's coefficient). (c) Left ordinate: Normalized cumulative response histogram of all neurons ($n = 98$) in the Attend-To condition (dark blue), and the Attend-away condition (light blue). Envelopes represent 90th percentile confidence levels. Initially (from attention-spot onset to several hundred milliseconds into the trial), there was a modest facilitation in the response of neurons when monkeys directed attention toward the RF (Attend-To condition). As the trial progressed, the responses between the 2 attention conditions converged before beginning to diverge, several tens of milliseconds (90 ± 40 ms) before the shortest time when the attention-spot could extinguish (900 ms). Thereafter, the response in the Attend-To condition continued to increase for the rest of the trial duration. Right ordinate: The time course of attention modulation (AMI) calculated for all neurons ($n = 98$) as a function of time (trial duration). The light envelope shows 90th percentile confidence levels. Dashed black line is the average AMI obtained by shuffling and performing a bootstrap analysis (see text for details). (d) Scatter plot of average AMI calculated for all neurons that showed attention facilitation in short duration trials (trial length < 400 ms) and long duration trials (> 2000 ms). An AMI was significantly higher for long duration trials compared with the short trials (41/98 showed significant attention facilitation; 5/98 showed suppression, not shown). The diagonal represents the unity slope. (e) Time course of change in AMI and its correlation with UMHR model data (brown dashed lines). The model had 2 temporally separable components: An early spatial attention component, empirically modeled as a temporally limited Gaussian function reflecting an initial moderate attentional facilitation, which levels-off around 450 ms after attention-spot onset, followed by an uncertainty modulated hazard rate (UMHR) component. The hazard rate increases with trial duration, signifying the subjective probability or a monkeys' expectation that the attention spot will extinguish.

no significant difference in RTs between Attend-To and Attend-Away conditions ($P = 0.78$, Spearman's test). This is consistent from a monkey's behavioral point of view as the 2 conditions were identical, requiring similar attentional resources; additionally they received the same amount of reward for correct responses irrespective of attention-spot location. The behavioral data show RTs monotonously decreasing with increasing trial duration as shown by a linear regression ($R^2 = 0.945$ and 0.958 for Attend-To and Attend-Away conditions, respectively). However, it is important to note that it would be unlikely for monkeys to continue to respond faster and faster, with increasing trial duration. This apparently seems to be the case, for a closer examination of RT data does show the tendency of RTs to asymptote toward the end. Consequently, RT data could as well be fitted to a log function rather than linear regression, though R^2 values in the current experiments were comparable for the 2 fits (0.93 and 0.95 , respectively) (Tsunoda and Kakei 2008).

These behavioral results could be qualitatively described by a model of probabilistic inference under temporal uncertainty. Task performance under time-dependent processes can be conceptualized as the hazard rate $b(t)$ (Luce 1986; Nobre and Shapiro 2006). Formally,

$$b(t) = \frac{p(t)}{1 - F(t)}, \quad (1)$$

where $p(t)$ is the probability density over the trial duration and $F(t)$ is the cumulative probability density function. A standard model (Luce 1986) predicts RT being proportional to the negative log of hazard rate, that is, $RT \sim -\log b(t)$. It is important to note that the RT not only reflects change in the subjective hazard rate but also the influence of a temporal uncertainty (Barlow 1964; Pike 1968; Lasley and Cohn 1981), making the duration of longer trials harder to predict when compared with the shorter trials (Eagleman et al. 2005). This feature was modeled as an "uncertainty modulated hazard rate" (UMHR) or $b^*(t) = b(t) * X$, where $b(t)$ is convolved by a Gaussian $X \sim N(t, \sigma)$, and σ is the Weber fraction for temporal estimation ($\sigma = 0.3$ in our calculations). The monkey's RT at time t was thus modeled as the negative log of the noisy hazard rate $b^*(t)$,

$$RT(t) \simeq -a \log(b^*(t)) \quad (2)$$

We used actual trial times from the experiments to calculate $b^*(t)$ and compared it with monkeys' RT data. The model provided an excellent fit to the RT data (Fig. 2*b*, correlation coefficient between $-\log b^*(t)$ and RT, $r = -0.95$; $P < 0.001$ Spearman's test). The inclusion of temporal uncertainty in the UMHR model provided a better description of behavioral performance than a standard hazard rate model (Supplementary Figs 2 and 3).

Neural Data: Analysis, Model, and Correlation with Behavioral Data

While monkeys engaged with the task, single-neuron activity was recorded in V1 (2 animals and 3 hemispheres). Our first priority was to ascertain if neural responses individually and at the population level were influenced by location of the attention spot, namely toward the RF (Attend-To) or away from it (Attend-Away). Stimulus-evoked responses at each time point were pooled in the 2 attention conditions and averaged for the

entire trial period. Figure 2*c* (left ordinate) shows the time course of normalized population response of 98 well-isolated single units. The population responses in the Attend-To conditions were moderately higher in the beginning and then tapered off to become non-different from the Attend-Away condition before starting to diverge consistently, a trend that continued for the rest of the trial period. Since each stimulus was presented with equal probability in each trial cycle, the combined activity over time had attention as the only variable. The AMI was calculated for individual neurons and averaged to obtain population AMI (Fig. 2*c*, right ordinate; see Methods for details). The population averages showed a transient increase in responses in the Attend-To condition, peaking around 140 ms (± 33 ms, mean \pm SD), and leveling off by 500 ms (± 80 ms). The dashed line indicates significance at $P < 0.05$, calculated by a bootstrap analysis where responses for individual units in the 2 attention conditions were shuffled and randomly picked 10 000 times to arrive at an unbiased estimate. As individual neurons showed considerable variability, we ran the same bootstrap analysis for the early period (first 600 ms after attention-spot onset) and plotted average AMI as well as peak AMI for individual units that showed either attentional facilitation or suppression (Supplementary Fig. 4; also see representative examples of individual neurons in Fig. 3*a*). After remaining close to the baseline for a few hundred milliseconds, the responses again diverged, beginning 810 ms (± 40 ms) after attention-spot onset and continuing until the end of the trial. We calculated mean responses for the 2 attention conditions for the period starting 800 ms after attention-spot onset to the end of the trials for an individual neuron. For the entire population, the average response during this period for the same neuron was significantly higher in the Attend-To compared with the Attend-Away condition [$F_{2,97} = 76.51$, $df = 2$, mean square error (MSE) = 0.002, $P < 0.014$, two-way ANOVAs]. Notably, this increase in response in the Attend-To condition and AMI preceded behavioral response at the earliest time point (or the shortest trial duration) by several tens of milliseconds (90 ± 40 ms). For further examination of change in AMI with trial length, we divided the trials into short (< 1400 ms) and long (> 1900 ms) duration trials (Fig. 2*d*). The average AMI was significantly higher in the longer compared with the shorter trials ($AMI_{\text{long}} = 0.16 \pm 0.03$ SEM; $AMI_{\text{short}} = 0.09 \pm 0.02$; $F_{2,97} = 29.51$, $df = 2$, MSE = 0.013, $P < 0.05$, two-way ANOVAs). When AMI was compared for the time windows where expectation-related changes were prominent, including 800–1400 ms (short trials; starting 100 ms before the earliest attention-spot extinction) and 800–2300 ms (long trials), the population average AMI for the long trials was again significantly higher than that in the short trials ($AMI_{\text{long}} = 0.23 \pm 0.02$ SD; $AMI_{\text{short}} = 0.12 \pm 0.03$ SD; $F_{2,97} = 97.51$, $df = 2$, MSE = 0.001, $P < 0.01$, two-way ANOVAs). We also compared the standard deviation (SD) in average AMI for short and long duration trials by taking the same number of short and long duration trials and calculated SD for average AMI. There was a trend toward SD for short trials (1000–1400 ms) to be lower compared with long trials (1900–2300). Taken together, the neural responses and attentional modulation during the same trial showed clear bimodality: An early, moderate increase followed by a sustained increase, beginning several milliseconds before the earliest attention-spot extinction. The AMI change over time was significantly bimodal—the significant dip occurred between 350 and 650 ms as confirmed by Hartigan's Dip test

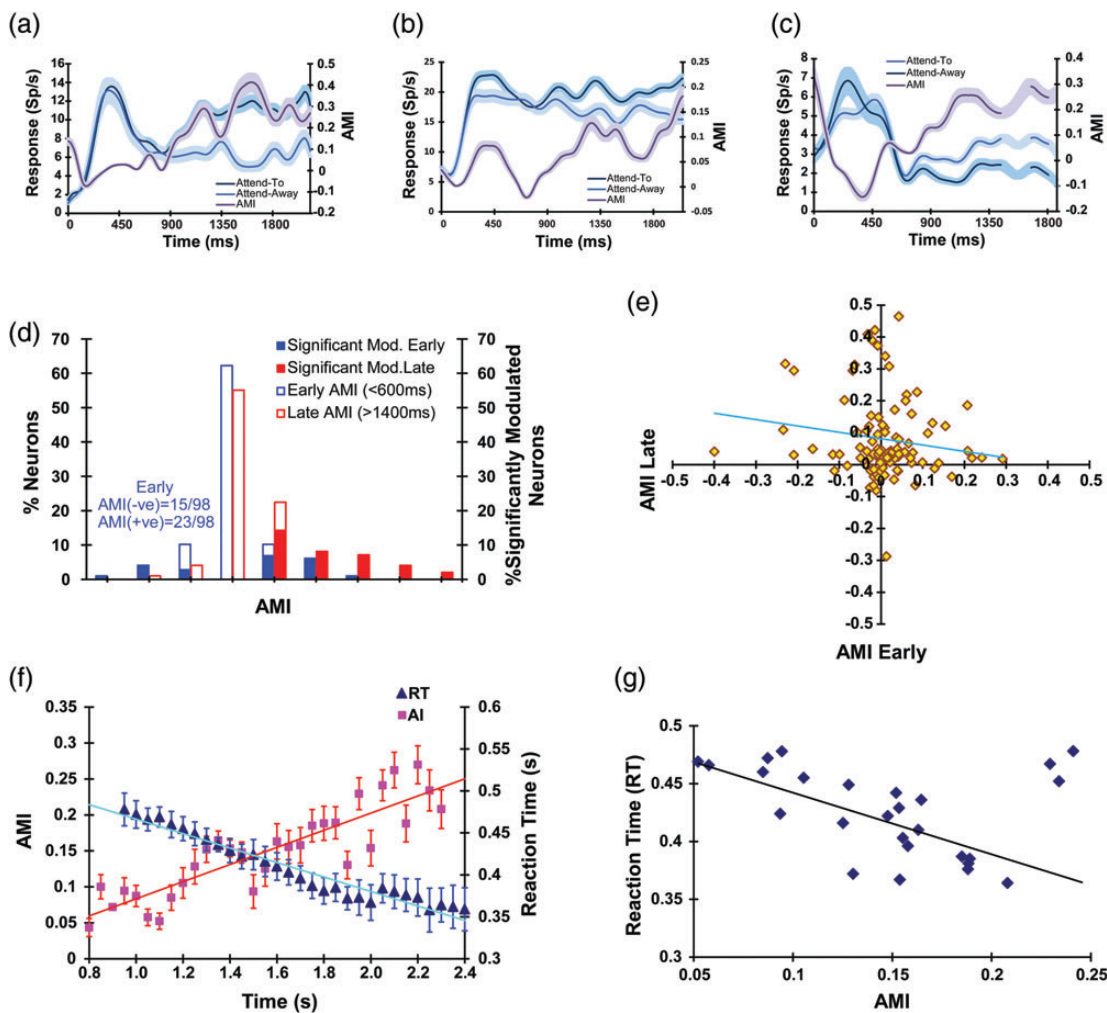


Figure 3. Diversity in attentional modulation and its relationship to monkeys' RT. (a–c) Individual V1 neurons show sequential spatial and temporal response modulation. Top row shows 3 examples. Response histograms in Attend-To (dark blue) and Attend-Away (light blue) can be divided into 2 distinct phases: An early response phase and a late period (~900 ms after attention-spot onset). In the early phase, responses in Attend-To and Attend-away conditions are quite diverse, while in the late phase, responses show striking similarity: Responses in the Attend-To condition increase and diverge from responses in the Attend-away condition. (d) Attention modulation in V1 neurons has distinct signatures of the task, including sequential spatial and temporal modulation. The AMI was calculated for early (first 600 ms; blue) and late periods (1400 ms – 2300; red) for each cell ($n = 98$). Left ordinate: Percentage of neurons in the overall population (no-fill bars); the attention index binned in values of 0.1 are shown on the abscissa, where negative AMI shows suppression in responses in the Attend-To condition and positive AMI signifies facilitation. Right ordinate: Percentage of neurons with a significant change in AMI (> ±0.15; solid-fill bars) for both early (blue bars) and late modulation (red bars) in AMI. As seen in data from individual cells above (Fig. 3b), at the population level also there was diversity in average AMI in the early attention period: 23/98 showed attention facilitation and 15/98 showed suppression. However in the Attend-To case, the same neurons showed a significantly different trend later in the trial, 48/98 (41/98 significant positive AMI) were facilitated and only 5/98 (3/98 significant negative AMI) showed attention-related suppression ($P < 0.01$, two-tailed t -test). (e) Early and late attentional modulation for the same neuron is uncorrelated. The AMI was calculated for each cell in early (first 600 ms) and late phases (> 1400 ms) of a trial. The plot shows the two to be highly uncorrelated ($R^2 = 0.0221$), indicating possible independence of origins and mechanisms of attention modulation in the 2 phases of the same trial. (f) Population-averaged AMI ($n = 98$ neurons) and respective RT plotted as a function of trial duration. Each data point for AMI and RT was calculated from AMI and RT values in successive 50 ms time bins. Error bars represent (S.E.M.). The solid lines represent linear least squares fits to the data. (g) Population distribution of AMI versus RT shows significant inverse correlation. The solid line, linear least-squares fit ($R^2 = 0.64$). Each point in the plot represents average values of all RTs and average AMI taken from within the same 100-ms bins for all correct trials from both monkeys.

(Dip = 0.029; $P = 0.002$; see Methods for details), demonstrating significant variability in attention that seemed to follow the monkeys' subjective assessment of the trial structure.

Extending our earlier conception of time-dependent, uncertainty modulated hazard rate (UMHR) to neural data, we modeled the response of cell i at time t as a function of the conditional probability that a target will appear in r_i , the region in the vicinity of cell i 's RF, at time t or a near-future time (up to $t + \delta$). We assumed that the monkey treated spatial and temporal components of the target's appearance and disappearance as independent events conditioned on its experience. This, in fact, describes the true structure of the task: first, the

target appears indicating “where” the attention needs to be directed; subsequently, the temporal dynamics of the task provides evidence about “when” the target will disappear. In practice, any single neuron's response will not be a function of these exact probabilities, but only estimates of these probabilities obtained by sampling stochastic evidence in the form of input from other neurons. Our model of neural activity thus takes the form of a linear combination of 2 components:

$$R = w_s \times p(r_i|x(t)) + w_t \times b \times (t) \quad (3)$$

The first term, $p(r_i|x(t))$, simply adds a Gaussian with an empirically set delay to the predicted response as soon as the

monkey receives evidence about where the attention spot will appear. The rationale derives from the proposal that the initial deployment of attention toward a salient or standout region, if sustained, is quickly suppressed by internal inhibitory influences, thereby prompting an attentional shift to a new locus (e.g., Koch and Ullman 1985). This manifests in a time-dependent interplay of bottom-up facilitation generated by external saliency, and intracortical inhibition that limits this facilitation (Klein 2000; Itti and Koch 2001). Indeed, it has been shown that sustained attention over time can impair perceptual sensitivity to the attended stimulus (Ling and Carrasco 2006), possibly due to direct inhibitory effects that follow attentional gain over time, especially in early sensory processing (Lou 1999). Similar inhibitory processes may underlie the well-known psychophysical phenomena variously described as “inhibition of return” (Posner et al. 1982) or “attentional blink” (Raymond et al. 1992). We emphasize that our choice of temporally limited Gaussian envelope as a physiologically plausible descriptor of the initial deployment of attention and its decay over time is not critical for the model; one can as well use a simple step function of the same duration without affecting the overall results. The second term, $b^*(t)$, is the uncertainty modulated hazard rate (UMHR). We also posit that different circuits underlie computation of the spatial and temporal evidence terms, with different effective strengths or confidence values; consequently, the neural responses are represented by a 2-component model with differential weights w_s and w_t for the spatial and temporal terms, respectively. The model output and population AMI showed highly significant correlation ($r=0.92$, $P<0.001$, Spearman’s test; Fig. 2e). The model simulations thus support the hypothesis that within the course of the same trial, the modulation of neuronal responses carries different weights depending on the task—that is, an initial modulation weighted more by a spatial component and a later component emphasizing time-dependent change in attention, relating closely to the expectation of an attention-spot extinction (hazard rate) and contingent behavioral response.

Next, we examined if responses of individual neurons also reflected the spatial and temporal variations inherent in the task. Previous studies have shown that V1 neurons exhibit considerable response variability depending on their laminar location and subtype specificity, indicating their diverse roles in processing bottom-up and top-down influences (Gilbert et al. 1996; McAlonan et al. 2008). The time course of responses for 2 attention conditions and attention modulation for 3 sample neurons are shown in Figure 3a–c and represent the diversity of neural populations in our data. Focusing first on the early period, soon after attention-spot onset, Neuron 1 (top left) exhibited virtually no difference between the Attend-To and Attend-Away conditions and no attentional modulation. Neuron 2 (top middle) showed response facilitation in the Attend-To condition and significant attentional modulation, whereas Neuron 3 (top right) displayed response suppression in the Attend-To condition, demonstrating attentional suppression. Interestingly, the situation was much different later in the trial, when, regardless of their early preferences, a significantly greater proportion of neurons exhibited prominent attentional facilitation (Fig. 3a–c, right ordinate). This was further confirmed by analyzing each neuron’s AMI in early and late phases for the entire population (Fig. 3d). In the early phase, there was no significant difference, with 22% (23/98) showing facilitation and 15% (15/98) suppression in the Attend-To condition. However, in the later phase, a

significantly higher proportion of neurons (47/98, or 48%) were facilitated, with 40/98 showing a significant increase in AMI >0.15 ($P<0.01$, Student’s *t*-test), whereas only 5/98 exhibited suppression in response during the same period. Additionally, there was no correlation between AMI in early and late phases of the same trial (Fig. 3e). This suggests that distinct sources of top-down attention may come into play during early and late periods of a trial. Next, we examined whether change in the AMI and RT was correlated with the task timing. The AMI from individual neurons and RT data from the same trial sets were averaged in segments of 50 ms, starting 800 ms after the attention-spot onset. There was a significant positive correlation between AMI and trial duration and a significant negative correlation between mean RT and trial duration ($r=0.9$ and -0.93 , respectively; $P<0.001$, Spearman’s test, Fig. 3f); on direct comparison, AMI and RT (Fig. 3g) showed a significant negative correlation ($r=-0.7$; $P<0.001$, Spearman’s test). These results demonstrate that trial length significantly influences changes in AMI and covaries with behavioral responses.

From experiment 1, we were able to confirm that even in a sustained attention task, monkeys’ attention varied with time and followed its internal assessment of task timing. There was a bimodal modulation in attention: An early temporally limited change was followed by a sustained increase that scaled with trial duration. We wondered if such bimodal changes in AMI would be replicable in trials with completely different time schedules. In experiment 2, the task was essentially the same as in experiment 1, with the following differences: The total trial duration was shorter, the initial “fixed” time window was longer, and the “variable” time window, when the attention spot disappeared, was shorter (see Methods for rationale and details). Figure 4a shows the population response of 32 single neurons (left ordinate), and the change in attention modulation (right ordinate), with the abscissa showing the time of the trials. The dashed line indicates the bootstrap averaged attention modulation at the $P<0.05$ significance level. Qualitatively, the responses in the 2 attention conditions as well as AMI followed similar dynamics as in experiment 1: An initial, moderate, time-limited increase that tapered off was followed by a late, consistent increase in responses in the Attend-To condition relative to the Attend-Away condition. It is noteworthy that the late increase in AMI occurred later in experiment 2, around 1180 (± 35) ms after attention-spot onset, compared with 810 (± 50) ms in experiment 1. This demonstrates that the late increase did follow task timing and monkeys allocated attentional resources in accordance with their expectation of a change in relation to target onset as well as its offset. In addition, the slope of AMI was significantly different than in experiment 1. Thus, late attentional modulation in V1 responses not only reflect their estimation of time as to “when” the expected change was to take place but also, the “duration” within which it was most likely to occur.

Finally, to examine if the pattern and time course of response modulation was indeed an indication of the monkeys’ internal state contingent on its behavioral response, we ran a third experiment where the animal’s behavioral task required no attentional control or behavioral response. Thus, in experiment 3, they had to simply hold fixation at a central spot while gratings were presented in the RF for a fixed duration, at the end of which they received a liquid reward. Normalized population responses from roughly the same sites, as in previous experiments, showed a robust, stimulus-evoked increase that

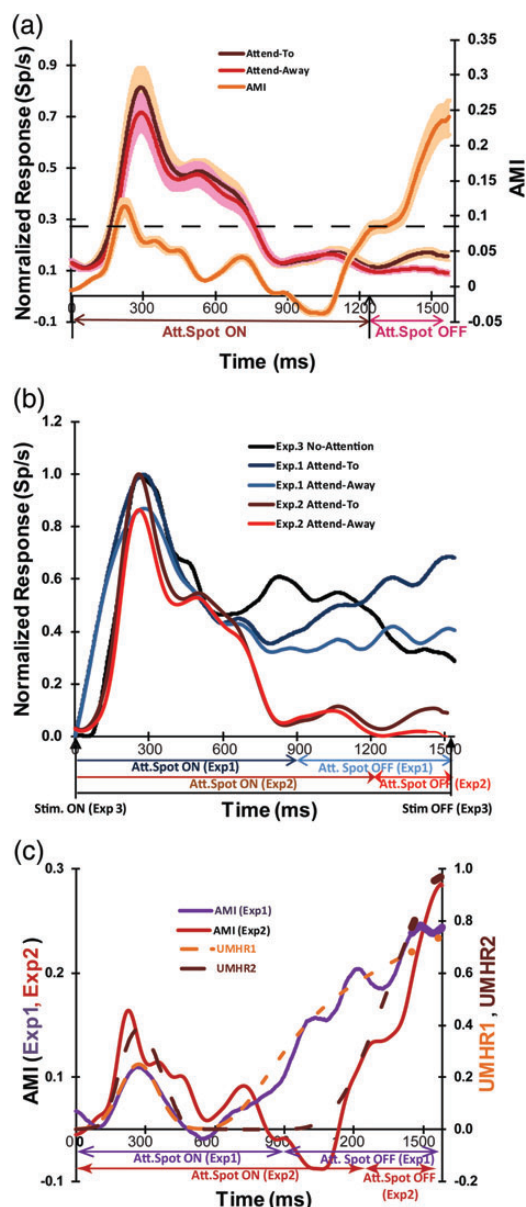


Figure 4. Neuronal responses are predictably influenced by changes in latency of attention-spot extinction and hazard rate. All data for experiment 2 are shown in shades of red. (a) Normalized histogram of responses of V1 neurons ($n = 32$) recorded from the same monkeys performing the task of experiment 2. Population responses in Attend-To (dark traces) and Attend-Away (light traces) are shown along with changes ± 90 th percentiles (envelope). The colored screen signifies the time (1250–1550 ms) within which the attention spot could extinguish at any moment. This leads to very different hazard latency and rate compared with experiment 1. Consequently, the time course of AMI (right ordinate, orange trace) shows a brief inflection early on, after attention-spot onset, that tapers down before rising again roughly 100 ms before the earliest attention-spot offset, but then quickly rises until the end of the trial. Envelope traces are 90th percentiles. Dashed black line is the average AMI obtained by shuffling and performing a bootstrap analysis (see text for details). (b) Comparison of experiments 1 and 2 with experiment 3 (No-attention experiment) in which the behavioral task did not require any attentional control or response to earn reward (see text for details). The black trace shows population response histogram ($n = 26$ V1 neurons). All trials for experiment 3 are aligned to the Attention-spot onset (left arrow) from experiments 1 and 2. The initial response profiles are qualitatively similar in all 3 experiments. However, in the 2 attention controlled experiments (1 and 2), the responses diverge later in the trial. Importantly, this divergence starts at strikingly different points in time depending on when the monkey starts expecting the attention spot to extinguish. The right arrows show the earliest offset times; time window for experiment 1 was truncated to 1500 ms to compare with experiment 2 whose offset time window is shown in pink. (c) The dashed lines represent output of the UMHR model for experiments 1 (UMHR1) and 2 (UMHR2).

gradually declined as the trial progressed. Figure 4b shows the average population response from 33 single units in experiment 3 (black trace). For comparison, also shown are the average responses in the 2 attention conditions from experiment 1 (blue), and experiment 2 (red). In experiment 3 and unlike experiments 1 and 2, we observed no late, time-dependent change of V1 responses. This indicates that, in the absence of attention and contingent behavioral responses, V1 neurons do not exhibit response modulations due to attentional engagement and temporal expectation-related changes (Fig. 4b). The bimodality in responses was completely absent in the absence of spatially localized attention and task-dependent behavioral requirements.

The proposal that late attentional modulation may derive from temporal demands of the task is supported by the significantly different rates of change of attention modulation in experiments 1 and 2 (Fig. 4c; see legend for details). The task-dependent change in AMI and the model simulation (normalized for comparison, Fig. 4c) shows the time courses of attention modulation in experiments 1 and 2, including an early modest increase followed by a monotonic increase following task-dependent latency. The UMHR model for RT data and AMI neural data also showed significant correlation ($r = 0.92$ and 0.96 for experiments 1 and 2, respectively; $P < 0.001$, Spearman's test).

Discussion

Most previous studies on the cortical locus of attention have focused on attention that enhances perceptual sensitivity to a location or feature of a relevant stimulus. Robust visuo-spatial attentional modulation has been described throughout the early visual pathway by manipulating distracter load, increasing (or decreasing) saliency, or spotlighting stimulus features such as color, orientation etc. (Spitzer et al. 1988; Gilbert et al. 1996; Connor et al. 1997; Vidyasagar 1999; Hayden and Gallant 2005; Chen et al. 2008; McAlonan et al. 2008). A number of underlying mechanisms have been proposed: Enhanced saliency by contextual-collinearity; biased-competition that prioritizes task-relevant information; scaling of neural responses or response normalization within the locus of attention; and selective increase of contrast of the attended location or object (Desimone and Duncan 1995; Gilbert et al. 1996; Maunsell and Treue 2006; Reynolds and Heeger 2009). Others have proposed intra- and interareal synchrony in specific frequency bands of local field potentials (Schroeder et al. 2001; Coe et al. 2002; Corbetta and Shulman 2002), or neuromodulators such as acetylcholine (Herrero et al. 2008; Goard and Dan 2009) as the principal drivers of spatially specific attention. In contrast, surprisingly few studies have examined temporal aspects of attention. Studies on human subjects attending selectively to different points in time have shown enhancement in behavioral performance (Carrasco and McElree 2001; Tse et al. 2004), an effect that seems to be principally mediated by fronto-parietal systems (Corbetta and Shulman 2002). Other studies have explored anticipatory changes in oculomotor preparation in parietal cortex, frontal eye fields, and superior colliculus (Dorris et al. 1997; Janssen and Shadlen 2005; Zhou and Thompson 2009), cerebral blood flow changes in preparation for trial onset (Sirotin and Das 2009), and responses of lateral intraparietal cortex (LIP) neurons signaling elapsed time relative to a remembered duration (Leon and Shadlen 2003). Thus, while the existing literature provides clues to temporal

signals in the brain, few previous studies have examined the representation of temporal attention by cortical neurons.

Our results demonstrate that a substantial proportion of V1 neurons are significantly modulated by attention. While the primary task involved sustained spatial attention alone, clearly attentional responses do undergo change over time. Our results suggest that this time-dependent change does not follow a monotonous increase or decrease, but seems to be guided by the spatial and temporal conditions of the task. A more basic question involves whether modulation in neural responses could be due to “low”-level stimulus effects such as pop-out or figure-ground interactions. Let us recapitulate some noteworthy aspects of our experimental design: in experiments 1 and 2, the attention spot was located close to but always outside the RF of a neuron. In addition, the stimulus gratings provided no cues and were irrelevant to the monkeys’ task performance. This effectively dissociated the monkeys’ behavioral task from the neural “task,” the stimuli used for eliciting neural responses. Placing the attention spot outside the RF allowed us to record neural modulation as a result of spatially directed attention, unadulterated by characteristics of the attended spot. Furthermore, responses during stimulus “blank” trials, when only the attention spot was presented, were subtracted from stimulus-driven responses, so that the residual responses were solely stimulus evoked. For all these reasons, the response difference in the 2 attention conditions would most likely be due to attentional state of the animal. Finally, a sustained, time-dependent increase in responses, late in the trial, as seen in the present data, is highly unlikely to be due to low-level sensory influences.

Strikingly, similar attentional modulation early in the trials in experiments 1 and 2, despite having distinctly different “fixed” periods during which the attention spot appeared and stayed (900 and 1250 ms, respectively), suggests that attentional resources get continuously titrated according to the task demands. During this early phase, the monkeys simply had to locate the relevant spot; subsequently, the high attentional requirement soon dissipated, though monkeys had to continue to monitor the spot. There is some evidence that attention, sustained over time, interferes with perceptual processing in the early sensory pathway (Lou 1999; Sherman and Guillery 2002; Ling and Carrasco 2006). In addition, top-down attention is resource limited (Lavie and Tsal 1994). This implies that unless and until the demand is high, intracortical circuits actively inhibit the use of the precious resource and deploy it elsewhere for more emergent needs, as is probably the case for inhibition of return and attentional blink conditions (Posner et al. 1982; Raymond et al. 1992). Our finding of an early, temporally limited modulation in attention is novel and departs from other similar studies. Studies done in LIP and V4 (Gallistel and Gibbon 2000; Janssen and Shadlen 2005) have shown no early modulation of neural response, suggesting such interactions in attentional processing may be characteristic of V1, which includes both intracortical and cortico-thalamic circuits (Sherman and Guillery 2002). These interactions, along with specific inhibitory mechanisms, may control both the magnitude and duration of attention deployment depending on behavioral contingencies. Recent evidence indicates these inhibitory processes to be quite dynamic; they may increase attentional facilitation via disinhibition in response to increasing resource demands, as in the case of uncertainty (Yu and Dayan 2005) and task difficulty (Castel et al. 2005), or conversely,

they may inhibit and/ or redeploy resources during low demand, as seen here, and in tasks requiring sustained attention (Ling and Carrasco 2006). Inhibitory mechanisms in V1, particularly those involving calretinin-positive neurons, are attractive substrates for implementation of such dynamic gating of top-down excitatory influences (Meskenaite 1997; Xu and Callaway 2009).

Our results also provide predictable and at least partially separable signatures of spatial and temporal attention. Several lines of evidence support this contention. First, as discussed above, at the beginning of a trial, attentional elevation is transient and is limited to locating the relevant spot. Hence, it seems fully devoted to space while remaining agnostic to time. Later in the trial, sustained elevation that scales with trial duration indicates transition from space only to time as well, because the expectation of a change also requires attention to time. Secondly, an increase in neuronal response in the Attend-To condition begins just before the earliest attention-spot disappearance and is temporally shifted in experiments 1 and 2, from 900 to 1250 ms (fixed delay). This demonstrates task-dependent internal state changes. Thirdly, the rate of change of late attentional modulation is significantly different in experiments 1 and 2 (Fig. 4c), signifying exquisite sensitivity of neuronal modulation not only to the time when the attention-spot was expected to disappear, but also the duration within which the expected change would occur. Fourthly, the magnitudes of early and late attention modulation are significantly different and have no correlation (Fig. 3d,e). These findings, combined with a highly significant inverse correlation between monkeys’ RT and trial duration, provide evidence for an overriding influence of attention to space in the early phase and to time in the later phase of the trials, both in guiding monkeys’ behavior and in their V1 responses. This argument is further supported by the results from experiment 3, where no such temporally guided late response modulation is discerned, confirming that dynamic changes in V1 responses in “attentionally controlled” experiments are indeed dependent on internal states that override “normal” neuronal responses and that change in a task-dependent fashion during the course of the same trial. The exquisite sensitivity of monkeys to trial duration is also evident in their surprisingly faster RTs during occasional repeats of trials of similar duration (see Supplementary Material for full description). Their behavioral data from experiment 1 show a significant improvement in RT, especially in the short duration trials (Supplementary Fig. S5). The monkeys seem to be sensitive to such randomly occurring repeat sequences, indicating their conscious awareness of such epochs, which they use to optimize their behavioral performance.

Prior to our work, 2 studies in awake, behaving monkeys, one in V4 (Ghose and Maunsell 2002) and the other in LIP (Janssen and Shadlen 2005), have shown similar task-dependent changes that correlate with the hazard rate of attention change. Both of these studies changed the shape of the hazard function to show that attention-guided saccadic responses covary with task timing and monkeys’ anticipation of stimulus change probability. However, the V4 study (Ghose and Maunsell 2002) did not test whether the responses correlated with the behavioral report of the monkeys. The LIP study (Janssen and Shadlen 2005) on the other hand did not explicitly use attention as a probe and a modulator of anticipatory change in their primary task. In the present study, we have not only focused on the earliest possible cortical area, V1, but have also

varied the hazard function by manipulating the trial duration to shape behavior. Also, instead of using saccade as an indicator of performance, which may confound results from an area where neurons can be sensitive to saccade preparation, our task required monkeys to respond by grabbing or releasing a lever, and that too after a fixed delay of 150 ms following attention-spot extinction, with maintained fixation till response, effectively minimizing reflexive and oculomotor influences (Dorris et al. 1997; Moore et al. 1998; Coe et al. 2002; Corbetta and Shulman 2002). As reward contingency and its timing have been shown to affect even V1 responses (Shuler and Bear 2006), we did not manipulate the reward to encourage faster RTs, and the animal received a fixed amount of reward for each correct response regardless of whether the attention spot appeared towards the RF or away from it. However, neural responses related to reward expectation are difficult to disambiguate from attention (Maunsell 2004), because most experimental designs requiring a behavioral report from primates (and other mammals) use reward as the primary motivator. Similar confounds exist in interpretation of other cognitive signals where reward timing is contingent on behavioral response and does not occur at a fixed time point during the trial. To the best of our knowledge, no clear demonstration of reward-related facilitation of single-neuron responses in V1 has been made, especially in a cognitive task where response modulation is directly related to whether the animal attended toward the RF or away from it. Thus, we believe our experiments capture internal signals principally related to the attentional state of the animal combined with expectation of a behavioral response.

The neural responses we describe provide insights into the dynamics and mechanisms of spatio-temporal attention. During the initial phase of the trials and responses, attentional resources are required to locate the attention spot (spatial uncertainty related to “where” the spot would appear). Here, no attention to time is necessary, because animals learn that there is no behavioral response to be made for a certain period of time after the trial onset. Conversely, in the later period, the attentional modulation seems to be wholly of temporal origin (temporal uncertainty related to “when” the spot would disappear). Here, no further attention to spatial processing, aside from awareness of the spot, appears necessary, resulting in considerably reduced attentional load (Chelazzi et al. 1998). This phase of attentional processing directly links with behavioral response, in this case releasing the lever whenever the attention spot is extinguished. The single-neuron responses (Fig. 3*a–c*) capture these 2 distinct phases, revealing an early spatial phase that varies between neurons and a later temporal phase, which is invariant across neurons. It is tempting to relate the attention-related responses to space and time to their respective uncertainties (Yoshor et al. 2007) arising from task structure. Dissipation in attention with increasing predictability has been proposed as a means to achieve cognitive economy (Pearce and Hall 1980; Yu and Dayan 2005). Conversely, the greater the uncertainty, the greater the attentional load. Initial uncertainty (“where” the attention spot will appear) followed by predictability (no change will occur for a period of time) might explain the early enhancement and dissipation of spatial attention signals in our task. Later in the trial, temporal uncertainty or expectation (“when” the spot will disappear) increases with each passing moment and requires heightened attention to time for successful completion of a trial and subsequent reward. Our finding of sequential spatial attention and temporal expectation signals in

V1 indicates that V1 may act both as a spatially selective gate for bottom-up inputs and an integrator of top-down influences, likely through feedback connections impinging on local and long-range intracortical and interareal networks (Gilbert et al. 1996; Das and Gilbert 1999; Lamme and Roelfsema 2000; Schroeder et al. 2001; Angelucci and Bressloff 2006; Hung et al. 2007; Womelsdorf and Fries 2007; Sommer and Wurtz 2008). Regardless of mechanism, robust modulation of neurons in early visual cortex, even when no overt visual cues are provided and in the absence of any eye movements, unequivocally demonstrates the widespread representation of behaviorally relevant time in the cortex.

Author's Contribution

J.S. and M.S. conceived the experiment and wrote the paper. J.S., H.S., and J.Sch trained monkeys and performed recordings. Y.K., J.B.T., J.S., and H.S. carried out the modeling. J.S., H.S., and Y.K. analyzed the data.

Supplementary Material

Supplementary material can be found at <http://www.cercor.oxfordjournals.org/> online.

Funding

Supported by grants from the US National Institutes of Health and the Simon Foundations to M.S.

Notes

The authors thank Dr R.P. Marini for excellent veterinary care and editorial advice; Travis Emery for technical help; Anubhav Jain for help with some figures; Samvaran Sharma and Sunanda Sharma for text editing, and members of the Sur lab especially Michael Goard and Nathan Wilson for discussions and comments. *Conflict of Interest:* The authors declare no competing financial interest.

References

- Angelucci A, Bressloff PC. 2006. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog Brain Res*. 154:93–120.
- Barlow JS. 1964. Evoked responses in relation to visual perception and oculomotor reaction times in man. *Ann N Y Acad Sci*. 112:432–467.
- Carrasco M, McElree B. 2001. Covert attention accelerates the rate of visual information processing. *Proc Natl Acad Sci USA*. 98:5363–5367.
- Castel AD, Pratt J, Chasteen AL, Scialfa CT. 2005. Examining task difficulty and the time course of inhibition of return: detecting perceptually degraded targets. *Can J Exp Psychol*. 59:90–98.
- Chelazzi L, Duncan J, Miller EK, Desimone R. 1998. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol*. 80:2918–2940.
- Chen Y, Martinez-Conde S, Macknik SL, Bereshpolova Y, Swadlow HA, Alonso JM. 2008. Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nat Neurosci*. 11:974–982.
- Coe B, Tomihara K, Matsuzawa M, Hikosaka O. 2002. Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. *J Neurosci*. 22:5081–5090.
- Connor CE, Preddie DC, Gallant JL, Van Essen DC. 1997. Spatial attention effects in macaque area V4. *J Neurosci*. 17:3201–3214.

- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:201–215.
- Coull JT, Nobre AC. 1998. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J Neurosci*. 18:7426–7435.
- Das A, Gilbert CD. 1999. Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*. 399:655–661.
- Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci*. 18:193–222.
- Dorris MC, Pare M, Munoz DP. 1997. Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J Neurosci*. 17:8566–8579.
- Dragoi V, Sharma J, Miller EK, Sur M. 2002. Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nat Neurosci*. 5:883–891.
- Eagleman DM, Tse PU, Buonomano D, Janssen P, Nobre AC, Holcombe AO. 2005. Time and the brain: how subjective time relates to neural time. *J Neurosci*. 25:10369–10371.
- Gallistel CR, Gibbon J. 2000. Time, rate, and conditioning. *Psychol Rev*. 107:289–344.
- Ghose GM, Maunsell JH. 2002. Attentional modulation in visual cortex depends on task timing. *Nature*. 419:616–620.
- Gilbert CD, Das A, Ito M, Kapadia M, Westheimer G. 1996. Spatial integration and cortical dynamics. *Proc Natl Acad Sci USA*. 93:615–622.
- Goard M, Dan Y. 2009. Basal forebrain activation enhances cortical coding of natural scenes. *Nat Neurosci*. 12:1444–1449.
- Goldberg ME, Bisley JW, Powell KD, Gottlieb J. 2006. Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. *Prog Brain Res*. 155:157–175.
- Goldman-Rakic PS. 1995. Anatomical and functional circuits in prefrontal cortex of nonhuman primates. Relevance to epilepsy. *Adv Neurol*. 66:51–63; discussion 63–55.
- Gutnisky DA, Hansen BJ, Iliescu BF, Dragoi V. 2009. Attention alters visual plasticity during exposure-based learning. *Curr Biol*. 19:555–560.
- Handwerker DA, Bandettini PA. 2011. Hemodynamic signals not predicted? Not so: a comment on Sirotnin and Das (2009). *Neuroimage*. 55:1409–1412.
- Hartigan JA, Hartigan PM. 1985. The Dip test of unimodality. *Ann Stat*. 13:70–84.
- Hayden BY, Gallant JL. 2005. Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron*. 47:637–643.
- Herrero JL, Roberts MJ, Delicato LS, Gieselmann MA, Dayan P, Thiele A. 2008. Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature*. 454:1110–1114.
- Hung CP, Ramsden BM, Roe AW. 2007. A functional circuitry for edge-induced brightness perception. *Nat Neurosci*. 10:1185–1190.
- Ito M, Gilbert CD. 1999. Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron*. 22:593–604.
- Itti L, Koch C. 2001. Computational modelling of visual attention. *Nat Rev Neurosci*. 2:194–203.
- Janssen P, Shadlen MN. 2005. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci*. 8:234–241.
- Klein RM. 2000. Inhibition of return. *Trends Cogn Sci*. 4:138–147.
- Kleinschmidt A, Muller NG. 2010. The blind, the lame, and the poor signals of brain function—a comment on Sirotnin and Das (2009). *Neuroimage*. 50:622–625.
- Koch C, Ullman S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*. 4:219–227.
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE. 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*. 53:279–292.
- Lamme VA, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*. 23:571–579.
- Lasley DJ, Cohn T. 1981. Detection of a luminance increment: effect of temporal uncertainty. *J Opt Soc Am*. 71:845–850.
- Lavie N, Tsai Y. 1994. Perceptual load as a major determinant of the locus of selection in visual attention. *Percept Psychophys*. 56:183–197.
- Leon MI, Shadlen MN. 2003. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*. 38:317–327.
- Lima B, Singer W, Neuenschwander S. 2011. Gamma responses correlate with temporal expectation in monkey primary visual cortex. *J Neurosci*. 31:15919–15931.
- Ling S, Carrasco M. 2006. When sustained attention impairs perception. *Nat Neurosci*. 9:1243–1245.
- Lou L. 1999. Selective peripheral fading: evidence for inhibitory sensory effect of attention. *Perception*. 28:519–526.
- Luce RD. 1986. Response times. London: Oxford University Press.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R. 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol*. 77:24–42.
- Mauk MD, Buonomano DV. 2004. The neural basis of temporal processing. *Annu Rev Neurosci*. 27:307–340.
- Maunsell JH. 2004. Neuronal representations of cognitive state: reward or attention? *Trends Cogn Sci*. 8:261–265.
- Maunsell JH, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci*. 29:317–322.
- McAlonan K, Cavanaugh J, Wurtz RH. 2008. Guarding the gateway to cortex with attention in visual thalamus. *Nature*. 456:391–394.
- Mechler F, Ringach DL. 2002. On the classification of simple and complex cells. *Vision Res*. 42:1017–1033.
- Meskenaitte V. 1997. Calretinin-immunoreactive local circuit neurons in area 17 of the Cynomolgus monkey, *Macaca fascicularis*. *J Comp Neurol*. 379:113–132.
- Moore T, Tolias AS, Schiller PH. 1998. Visual representations during saccadic eye movements. *Proc Natl Acad Sci USA*. 95:8981–8984.
- Moran J, Desimone R. 1985. Selective attention gates visual processing in the extrastriate cortex. *Science*. 229:782–784.
- Motter BC. 1993. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J Neurophysiol*. 70:909–919.
- Nobre A, Correa A, Coull J. 2007. The hazards of time. *Curr Opin Neurobiol*. 17:465–470.
- Nobre AC. 2001. Orienting attention to instants in time. *Neuropsychologia*. 39:1317–1328.
- Nobre AC, Shapiro KL. 2006. Other dimensions of attention. *Neural Netw*. 19:1450–1452.
- Oswal A, Ogden M, Carpenter RH. 2007. The time course of stimulus expectation in a saccadic decision task. *J Neurophysiol*. 97:2722–2730.
- Pearce JM, Hall G. 1980. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev*. 87:532–552.
- Pike AR. 1968. Latency and relative frequency of response in psychophysical discrimination. *Br J Math Stat Psychol*. 21:161–182.
- Posner MI, Cohen Y, Rafal RD. 1982. Neural systems control of spatial orienting. *Philos Trans R Soc Lond B Biol Sci*. 298:187–198.
- Raymond JE, Shapiro KL, Arnell KM. 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform*. 18:849–860.
- Reynolds JH, Heeger DJ. 2009. The normalization model of attention. *Neuron*. 61:168–185.
- Ringach DL, Hawken MJ, Shapley R. 1997. Dynamics of orientation tuning in macaque primary visual cortex. *Nature*. 387:281–284.
- Roelfsema PR, Lamme VA, Spekreijse H. 1998. Object-based attention in the primary visual cortex of the macaque monkey. *Nature*. 395:376–381.
- Schroeder CE, Mehta AD, Foxe JJ. 2001. Determinants and mechanisms of attentional modulation of neural processing. *Front Biosci*. 6:D672–D684.
- Sengpiel F, Hubener M. 1999. Visual attention: spotlight on the primary visual cortex. *Curr Biol*. 9:R318–R321.
- Sharma J, Dragoi V, Tenenbaum JB, Miller EK, Sur M. 2003. V1 neurons signal acquisition of an internal representation of stimulus location. *Science*. 300:1758–1763.
- Sherman SM, Guillery RW. 2002. The role of the thalamus in the flow of information to the cortex. *Philos Trans R Soc Lond B Biol Sci*. 357:1695–1708.

- Shuler MG, Bear MF. 2006. Reward timing in the primary visual cortex. *Science*. 311:1606–1609.
- Sirotin YB, Das A. 2009. Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature*. 457:475–479.
- Sommer MA, Wurtz RH. 2008. Brain circuits for the internal monitoring of movements. *Annu Rev Neurosci*. 31:317–338.
- Spitzer H, Desimone R, Moran J. 1988. Increased attention enhances both behavioral and neuronal performance. *Science*. 240:338–340.
- Squire RF, Noudoost B, Schafer RJ, Moore T. 2013. Prefrontal contributions to visual selective attention. *Annu Rev Neurosci*. 36:451–466.
- Summerfield C, Egner T. 2009. Expectation (and attention) in visual cognition. *Trends Cogn Sci*. 13:403–409.
- Tse PU, Intriligator J, Rivest J, Cavanagh P. 2004. Attention and the subjective expansion of time. *Percept Psychophys*. 66:1171–1189.
- Tsunoda Y, Kakei S. 2008. Reaction time changes with the hazard rate for a behaviorally relevant event when monkeys perform a delayed wrist movement task. *Neurosci Lett*. 433:152–157.
- Vanzetta I, Sloviter H. 2010. A BOLD Assumption. *Front Neuroenergetics* 2:139–142.
- Vidyasagar TR. 1998. Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *Neuroreport*. 9:1947–1952.
- Vidyasagar TR. 1999. A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Res Brain Res Rev*. 30:66–76.
- Womelsdorf T, Fries P. 2007. The role of neuronal synchronization in selective attention. *Curr Opin Neurobiol*. 17:154–160.
- Xu X, Callaway EM. 2009. Laminar specificity of functional input to distinct types of inhibitory cortical neurons. *J Neurosci*. 29:70–85.
- Yoshor D, Ghose GM, Bosking WH, Sun P, Maunsell JH. 2007. Spatial attention does not strongly modulate neuronal responses in early human visual cortex. *J Neurosci*. 27:13205–13209.
- Yu AJ, Dayan P. 2005. Uncertainty, neuromodulation, and attention. *Neuron*. 46:681–692.
- Zhou HH, Thompson KG. 2009. Cognitively directed spatial selection in the frontal eye field in anticipation of visual stimuli to be discriminated. *Vision Res*. 49:1205–1215.