

Bayesian inference and attentional modulation in the visual cortex

Rajesh P.N. Rao

Department of Computer Science and Engineering, University of Washington, Seattle, Washington, USA.

Sponsorship: Supported by grants from NSF, ONR, and the Packard Foundation.

Correspondence and requests for reprints to Rajesh P.N. Rao, PhD, Department of Computer Science and Engineering, Box 352350, University of Washington, Seattle, WA 98195-2350, USA
E-mail: rao@cs.washington.edu

Received 20 July 2005; accepted 23 August 2005

The responses of neurons in cortical areas V2 and V4 can be significantly modulated by attention to particular locations within an input image. We show that such effects emerge naturally when perception is viewed as a probabilistic inference process governed by Bayesian principles and implemented in hierarchical cortical networks. The proposed model can explain a rich variety of attention-related responses in cortical area V4 including multiplicative modulation of tuning curves, restoration of neural responses

in the presence of distracting stimuli, and influence of attention on neighboring unattended locations. Our results suggest a new interpretation of attention as a cortical mechanism for reducing perceptual uncertainty by combining top-down task-relevant information with bottom-up sensory inputs in a probabilistic manner. *NeuroReport* 16:1843–1848 © 2005 Lippincott Williams & Wilkins.

Keywords: attention, Bayesian inference, computational models, vision, visual cortex

Introduction

Visual attention plays a crucial role in the perception of objects in complex scenes [1,2]. By selecting specific scene locations or objects for preferential processing, attentional mechanisms allow efficient use of the brain's limited computational resources. Given its importance in perception, several studies have sought to unravel the neural mechanisms of attention. McAdams and Maunsell [3] showed that the tuning curve of a neuron in cortical area V4 is multiplied by an approximately constant factor when the monkey focuses attention on a stimulus within the neuron's receptive field. Reynolds *et al.* [4] have shown that focusing attention on a target in the presence of distractors causes the response of a V2 or V4 neuron to closely approximate the response elicited when the target appears alone. Finally, a study by Connor *et al.* [5] demonstrated that responses to unattended stimuli can be affected by spatial attention to nearby locations. In all three studies, neural responses are determined by combining a task-relevant constraint (such as a prespecified spatial location) with sensory information (the input image).

A rigorous approach to combining sensory evidence with prior constraints is to use Bayesian models [6–8]. In such models, prior knowledge and sensory information are probabilistically combined according to Bayes' rule. In this article, we show that a Bayesian model for visual processing can explain all three classes of attentional effects discussed above. These effects emerge as a consequence of Bayesian inference in a hierarchical network of neurons encoding the probabilistic interaction between visual features and their

spatial locations. We additionally show that the responses of model neurons become invariant to changes in the spatial position of features, mimicking the responses of neurons in the occipitotemporal pathway [9]. The model and simulation results together suggest a new interpretation of cortical neurons as probabilistic integrators of top-down and bottom-up information in hierarchical networks, and of attention as a cortical mechanism for reducing uncertainty in complex scenes.

Methods

Probabilistic generative model

Bayesian models typically assume a probabilistic generative model specifying conditional dependencies between inputs and their causes. Figure 1a shows a simple three-level generative model incorporating two variables relevant to V4 neurons: spatial location L , which is assumed to be one of n values L_1, L_2, \dots, L_n , and visual feature F , which can be one of m different values F_1, F_2, \dots, F_m . The 'intermediate representation' node C denotes different combinations of features and locations, each of its values C_1, C_2, \dots, C_p encoding a specific feature at a specific location. Representing all possible combinations is infeasible but it is sufficient to represent those that occur frequently and to map each feature–location (L, F) combination to the closest C_i using an appropriate distribution $P(C_i | L, F)$ (see Simulation details). An image with a specific feature at a specific location is generated according to the image likelihood $P(I | C)$.

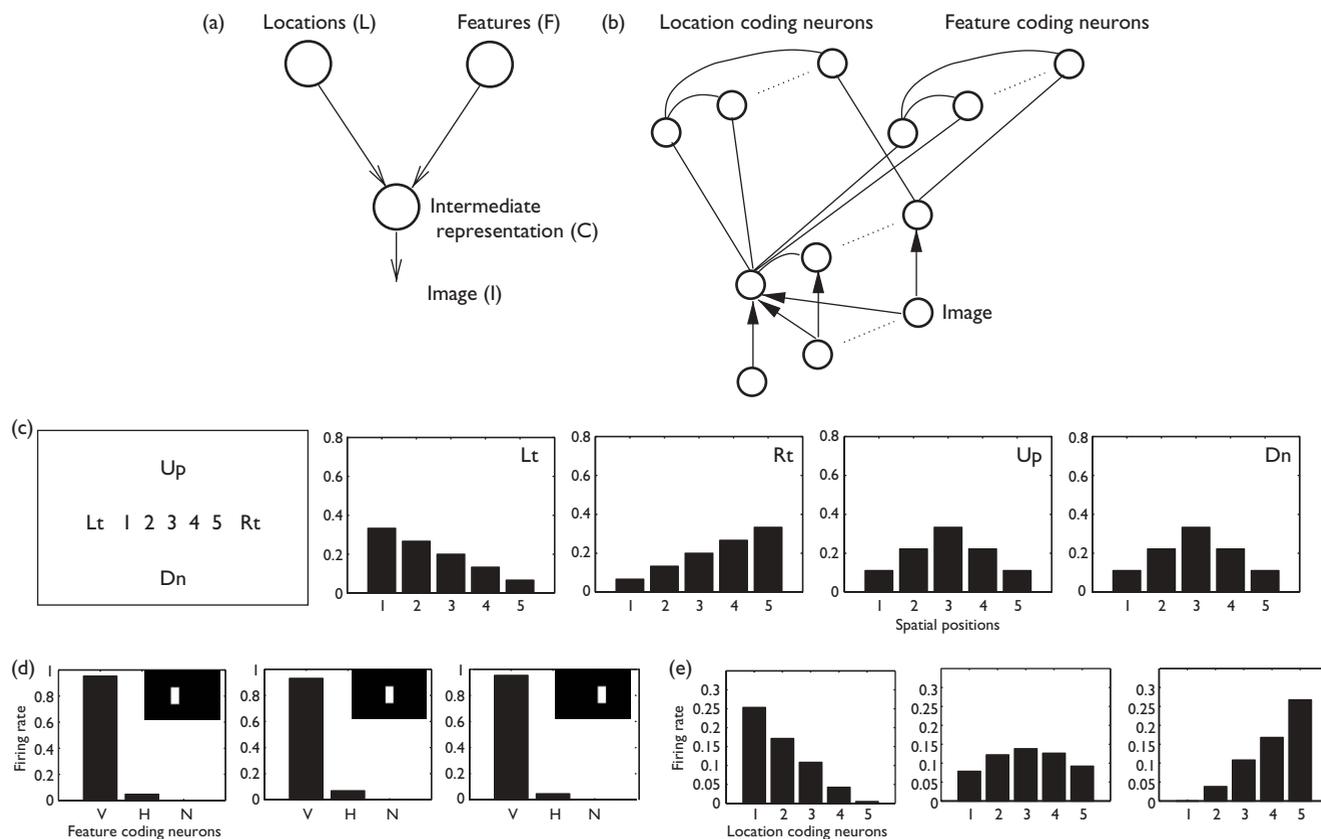


Fig. 1 Model network and invariant responses. (a) Probabilistic generative model that factors images into spatial locations and features. (b) Three-level network for implementing Bayesian inference for the generative model in (a). Circles represent neurons, arrows represent feedforward connections, and lines represent bidirectional connections. (c) Left panel: example image locations (labeled 1–5 and Up, Dn, Lt, and Rt for up, down, left, and right) relevant to the experiments discussed in the paper. Right panel: each bar plot shows $P(C_i | L, F)$ for a fixed value of L ($=Lt, Rt, Up,$ or Dn) and for an arbitrary fixed value of F . Each bar represents the probability for the feature–location combination C_i encoding one of the locations 1–5. (d) Each plot represents the responses of three model neurons, the first coding for vertical features (V), the second for horizontal features (H), and the third for the null (blank) stimulus (N), when the input is a vertical bar flashed at three different locations (dark inset images). The vertical feature-coding neuron’s firing rate remains high, invariant to bar position. (e) Information about the vertical bar’s current location is represented by the location-coding neurons, which compute the posterior probability $P(L | I = I')$.

Bayesian inference

We interpret perception as estimating the posterior probabilities of features (more generally, objects or object parts) and their locations in an input image. This can be done for arbitrary generative models using a well known Bayesian algorithm called ‘belief propagation’ [10]. Given the model in Fig. 1a and a specific input image $I = I'$, belief propagation prescribes that the following ‘messages’ (probabilities) be transmitted from one node to another, as given by the arrows in the subscripts:

$$m_{L \rightarrow C} = P(L), \quad m_{F \rightarrow C} = P(F), \quad m_{I \rightarrow C} = P(I = I' | C), \quad (1)$$

$$m_{C \rightarrow L} = \sum_F \sum_C P(C | L, F) P(F) P(I = I' | C), \quad (2)$$

$$m_{C \rightarrow F} = \sum_L \sum_C P(C | L, F) P(L) P(I = I' | C). \quad (3)$$

The first two messages in Equation (1) are prior probabilities encoding beliefs about locations and features before a sensory input becomes available. These are used in forming the messages in Equations (2) and (3). The posterior

probabilities of the unknown variables $C, L,$ and $F,$ given the input image $I',$ are calculated by combining messages at each node as follows:

$$P(C | I = I') = \frac{1}{\alpha} m_{I \rightarrow C} \sum_F \sum_L P(C | L, F) m_{L \rightarrow C} m_{F \rightarrow C}, \quad (4)$$

$$P(L | I = I') = \frac{1}{\beta} m_{C \rightarrow L} P(L), \quad (5)$$

$$P(F | I = I') = \frac{1}{\gamma} m_{C \rightarrow F} P(F), \quad (6)$$

where $\alpha, \beta,$ and γ are normalization constants that make each of the above probabilities sum to 1. Note how the prior $P(L)$ multiplicatively modulates the posterior probability of a feature in Equation (6) via Equation (3). As described below, we simulate spatial attention by increasing $P(L)$ for a desired location.

Modeling cortical neurons and networks

We use the standard leaky integrator model. At any given hierarchical level, the instantaneous firing rate v_i of neuron i

in a recurrently connected network is given by

$$\tau \frac{dv_i}{dt} = -v_i + f \left(\sum_j w_{ij} I_j \right) + g \left(\sum_j u_{ij} v_j \right), \quad (7)$$

where τ is a time constant, f and g are dendritic filtering functions, I_j is the firing rate of input neuron j , w_{ij} is the strength of the synapse from input j to recurrent neuron i , and u_{ij} is the recurrent synaptic strength from neuron j to neuron i . Spikes can be generated by interpreting Equation (7) as the dynamics of the membrane potential v_i , rather than the firing rate and using [11]

$$P(\text{neuron } i \text{ spikes at time } t) = k \exp[(v_i(t) - T)/c], \quad (8)$$

where k and c are constants, and T is the membrane threshold.

Bayesian inference in networks of neurons

Several models have been proposed for neural implementation of Bayesian inference (e.g. [12–18]). We extend the model in [19] to implement the belief propagation equations above in a neural circuit (Fig. 1b). Each V4 neuron is assumed to encode a feature F_i as its preferred stimulus. A separate group of neurons (e.g. in the parietal cortex) is assumed to encode spatial locations (and potentially other spatiotemporal transformations) irrespective of feature values. Lower level neurons (e.g. in V2 and V1) are assumed to represent the intermediate representations C_i . Note that the architecture in Fig. 1b mimics the division of labor between the ventral object processing ('what') stream and the dorsal spatial processing ('where') stream in the visual cortex [20].

The initial firing rates of location and feature-coding neurons represent prior probabilities $P(L)$ and $P(F)$, respectively, assumed to be set by task-dependent feedback from higher areas such as those in the prefrontal cortex. The input likelihood $P(I=I' | C)$ is set to $\sum_j w_{ij} I_j$, where the weights w_{ij} represent the attributes of C_i (specific feature at a specific location). For our simulations, we set each weight vector to a spatially localized oriented Gabor filter. Equations (2) and (3) are assumed to be computed by feedforward neurons in the location-coding and feature-coding parts of the network, with their synapses encoding $P(C | L, F)$. Taking the logarithm [19] of both sides of Equations (4)–(6), we obtain equations that can be computed using leaky integrator neurons as in Equation (7) (f and g are assumed to approximate a logarithmic transformation). Recurrent connections in Equation (7) are used to implement the inhibitory component corresponding to the negative logarithm of the normalization constants. Furthermore, as the membrane potential $v_i(t)$ is now proportional to the log of the posterior probability, i.e. $v_i(t) = c \log P(F | I=I') + T$ (and similarly for L and C), using Equation (8) we obtain

$$P(\text{feature-coding neuron } i \text{ spikes at time } t) = k \exp[(v_i(t) - T)/c] = k P(F | I = I'). \quad (9)$$

This provides a new interpretation of the spiking probability (or instantaneous firing rate) of a V4 neuron as representing the posterior probability of a preferred feature in an image region irrespective of spatial location.

Simulation details

To model the three primate experiments discussed above [3–5], we used horizontal and vertical bars that could appear at

nine different locations in the input image (Fig. 1c). All results were obtained using a network with a single set of parameters. $P(C | L, F)$ was chosen such that for any given value of L and F , say location L_j and feature F_k , the value of C closest to the combination (L_j, F_k) received the highest probability, with decreasing probabilities for neighboring locations (Fig. 1c).

Results

Invariant responses to changes in spatial position

As shown in Fig. 1d, the response of a model V4 neuron (here, a vertical feature-coding neuron representing $P(F | I=I')$ for the vertical feature) remains unchanged despite significant shifts in the location of a vertical bar stimulus. This is due to the summation over locations in Equation (3) which is used in Equation (6). Note that all locations are assumed equally probable; that is, $P(L)$ is assumed to be uniform. Information about the current location of the stimulus is not lost but is represented by the activities of the location-coding neurons in the network (Fig. 1e).

Multiplicative modulation of responses

We simulated the attentional task of McAdams and Maunsell [3] by presenting a vertical bar and a horizontal bar simultaneously in an input image. 'Attention' to a location L_i containing one of the bars was simulated by setting a high value for $P(L_i)$, corresponding to a higher firing rate for the neuron coding for that location.

Figure 2a depicts the orientation-tuning curves of the vertical feature-coding model V4 neuron in the presence and absence of attention (squares and circles, respectively). The plotted points represent the neuron's firing rate, encoding the posterior probability $P(F | I=I')$, F being the vertical feature. Attention in the model approximately multiplies the 'unattended' responses by a constant factor, similar to V4 neurons (Fig. 2b). This is due to the change in the prior $P(L)$ for locations between the attended and unattended modes, which affects Equations (3) and (6) multiplicatively.

Effects of attention on responses in the presence of distractors

To simulate the experiments of Reynolds *et al.* [4], a single vertical bar ('reference') was presented in the input image and the responses of the vertical feature-coding model neuron were recorded over time. As seen in Fig. 2c (upper panel, dotted line), the neuron's firing rate reflects a posterior probability close to 1 for the vertical stimulus. When a horizontal bar ('probe') alone is presented at a different location, the neuron's response drops dramatically (solid line) as its preferred stimulus is a vertical bar, not a horizontal bar. When the horizontal and vertical bars are simultaneously presented ('pair'), the firing rate drops to almost half the value elicited for the vertical bar alone (dashed line). Signaling increased uncertainty about the stimulus compared with the reference-only case. When 'attention' is turned on by increasing $P(L)$ for the vertical bar location (Fig. 2c, lower panel), the firing rate is restored to its original value and a posterior probability close to 1 is signaled (dot-dashed line). Thus, attention acts to reduce uncertainty about the stimulus, given a location of interest.

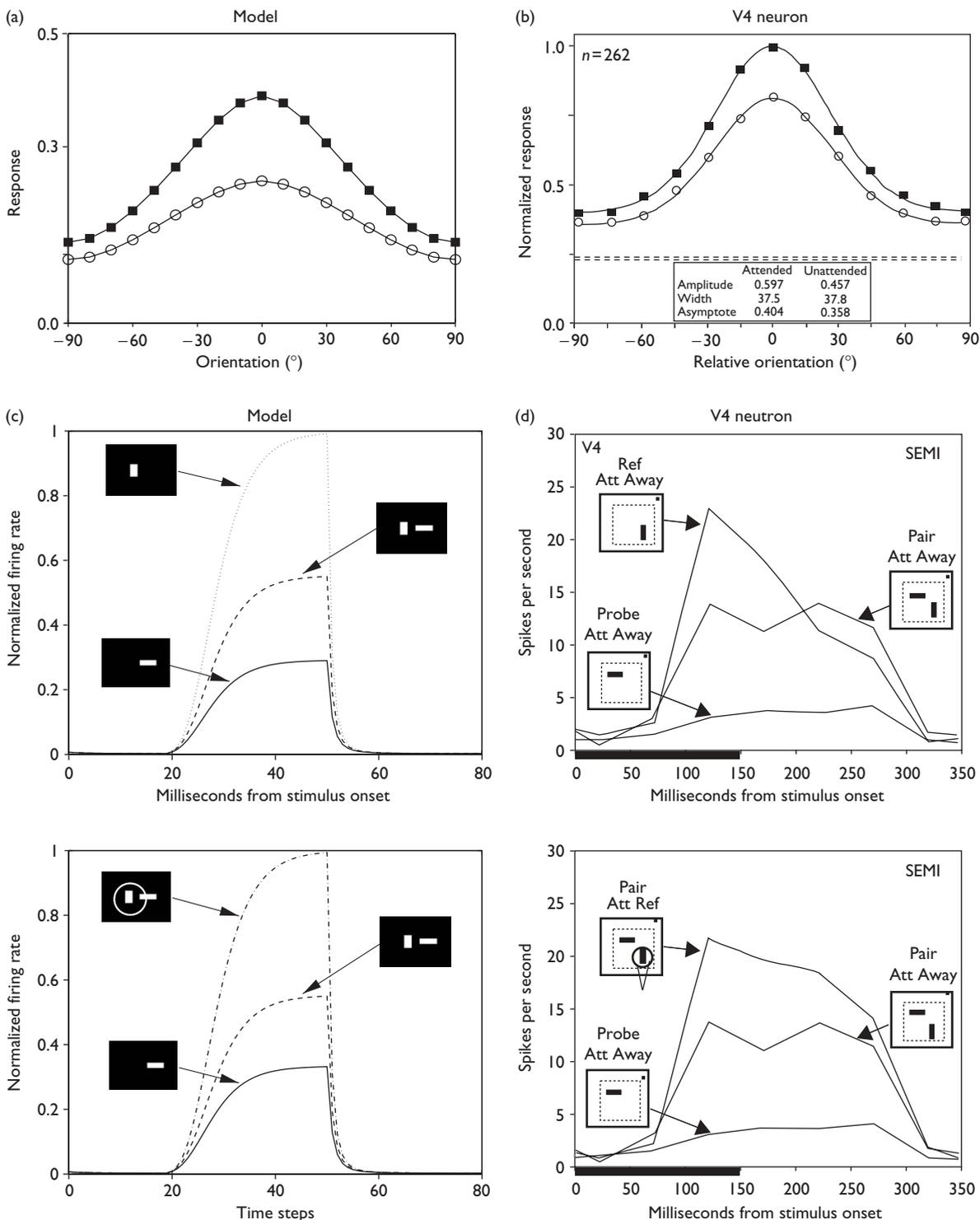


Fig. 2 Multiplicative modulation and response restoration. (a) Orientation-tuning curve of a feature-coding model neuron with a preferred stimulus orientation of 0° with (filled squares) and without (unfilled circles) attention. (b) Normalized orientation-tuning curves for a population of V4 neurons with (filled squares) and without attention (unfilled circles) (reproduced from [3], copyright 1999 by the Society of Neuroscience). (c) Top panel: the three line plots represent the vertical feature-coding neuron's response to a vertical bar ('reference', Ref), a horizontal bar at a different position ('probe'), and both bars presented simultaneously ('pair'). In each case, the input lasted 30 time steps, beginning at time step 20. Bottom panel: when 'attention' (Att) is focused on the vertical bar, the firing rate for the pair stimulus approximates the firing rate obtained for the reference alone. (d) Top panel: responses from a V4 neuron without attention (reproduced from [4], copyright 1999 by the Society of Neuroscience). Bottom panel: responses from the same neuron when attending to the vertical bar (see condition Pair Att Ref) (reproduced from [4], copyright 1999 by the Society for Neuroscience).

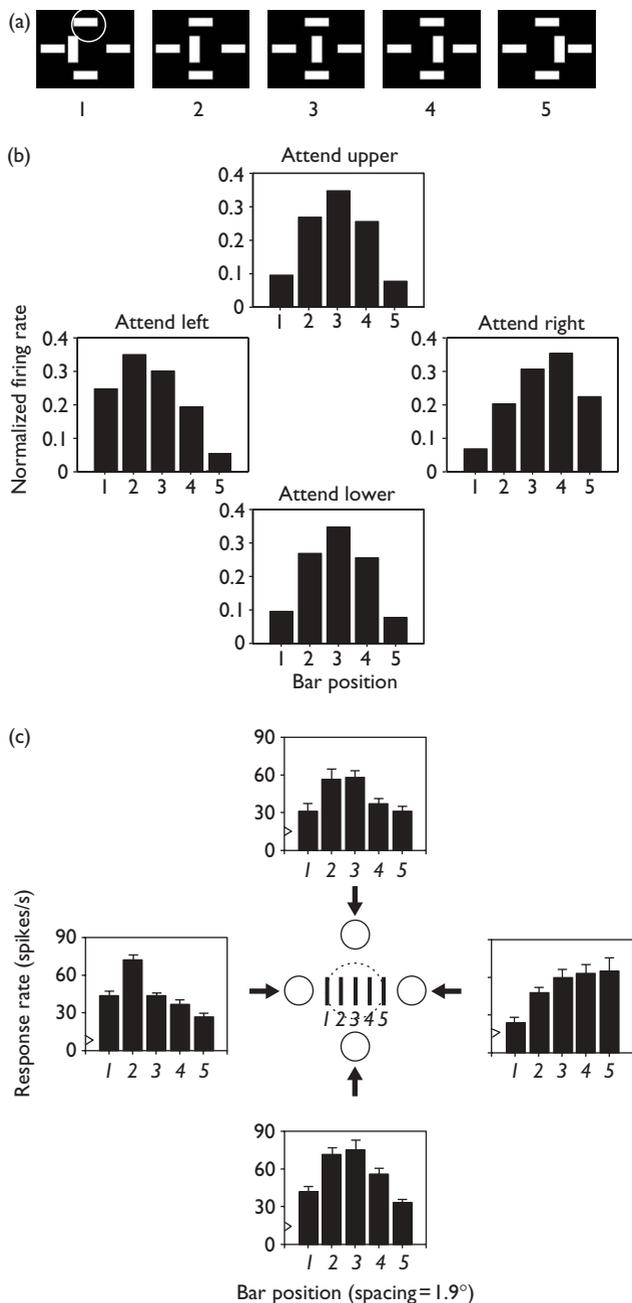


Fig. 3 Influence of attention on neighboring spatial locations. (a) Example trial based on Connor *et al.*'s experiments [5] showing five images, each containing four horizontal bars and one vertical bar. Attention was focused on a horizontal bar (e.g. upper bar, circled) while the vertical bar's position was varied. (b) Responses of the vertical feature-coding model neuron. Each plot shows five responses, one for each location of the vertical bar, as attention was focused on the upper, lower, left, or right horizontal bar. (c) Responses from a V4 neuron (reproduced from [5], copyright 1997 by the Society for Neuroscience).

Such behavior closely mimics the effect of spatial attention in areas V2 and V4 [4] (Fig. 2d).

Effects of attention on neighboring spatial locations

We simulated the experiments of Connor *et al.* [5] using an input image containing four fixed horizontal bars at four

different locations (see Fig. 3a). A vertical bar was flashed at one of five different locations in the center (Fig. 3a, 1–5). Each bar plot in Fig. 3b shows the responses of the vertical feature-coding model V4 neuron as a function of vertical bar location (bar positions 1 through 5) when attention is focused on one of the horizontal bars (left, right, upper or lower). Attention was again simulated by assigning a high prior probability for the location of interest.

As seen in Fig. 3b, there is a pronounced effect of proximity to the locus of attention: the unattended stimulus (vertical bar) produces higher responses when it is closer to the attended location than further away (see, e.g., 'attend left'). This is due to the spatial spread in the conditional probability $P(C|L, F)$ (see Fig. 1c), and its effect on Equations (3) and (6). The larger responses near the attended location reflect a reduction in uncertainty at locations closer to the focus of attention compared with locations further away. For comparison, the responses from a V4 neuron are shown in Fig. 3c (from [5]).

Discussion

The model we have proposed makes several predictions. First, as feedback connections from higher cortical areas are assumed to convey prior probabilities [$P(L)$ and $P(F)$], the model predicts that stimulation of these feedback axons should mimic the effects of attention, producing, for example, a multiplicative modulation of orientation-tuning curves as in Fig. 2a. A second prediction is that the relationship between a neuron's membrane potential and the instantaneous firing rate is exponential. Emerging evidence shows such a transformation in certain visual neurons [21,22]. Finally, the model predicts that attention to a particular feature (e.g. a vertical bar) should also influence the processing of similar 'neighboring' features (e.g. a bar at 80°) due to uncertainty in feature space, similar to the uncertainty in locations as in Fig. 1c.

Our model bears some similarities to a recent Bayesian model of attention proposed by Yu and Dayan [23] (see also [24,25]). They use a five-layer neural architecture and a logarithmic probability encoding scheme [19] to model reaction time effects and multiplicative response modulation (Fig. 2a). Their model, however, does not use an intermediate representation to factor input images into separate feature and location attributes (Fig. 1a). It therefore cannot explain effects such as the influence of attention on neighboring unattended locations [5].

Conclusion

Our results suggest that attentional modulation in the primate visual cortex may arise as a consequence of Bayesian inference in hierarchical cortical networks. Our study provides a new interpretation of attention as a cortical mechanism for reducing uncertainty during the perception of complex scenes.

References

1. James W. *The principles of psychology*. New York: Holt; 1890.
2. Itti L, Rees G, Tsotsos J, editors. *Neurobiology of attention*. New York: Academic Press; 2005.
3. McAdams CJ, Maunsell JHR. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 1999; 19:431–441.

4. Reynolds JH, Chelazzi L, Desimone R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 1999; **19**:1736–1753.
5. Connor CE, Preddie DC, Gallant JL, Van Essen DC. Spatial attention effects in macaque area V4. *J Neurosci* 1997; **17**:3201–3214.
6. Knill DC, Richards W. *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press; 1996.
7. Rao RPN, Olshausen BA, Lewicki MS. *Probabilistic models of the brain: perception and neural function*. Cambridge, Massachusetts: MIT Press; 2002.
8. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 2004; **27**:712–719.
9. Van Essen DC. Functional organization of primate visual cortex. In: Peters A, Jones EG, editors. *Cerebral cortex*. Vol. 3. New York: Plenum; 1985. pp. 259–329.
10. Pearl P. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, California: Morgan Kaufmann; 1988.
11. Gerstner W. Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Comput* 2000; **12**:43–89.
12. Simoncelli EP. *Distributed representation and analysis of visual motion*. PhD Thesis, Department of Electrical Engineering and Computer Science, MIT; 1993.
13. Anderson CH, Van Essen DC. Neurobiological computational systems. In: Zurada JM, Marks RJ 2nd, Robinson CJ, editors. *Computational intelligence: imitating life*. New York: IEEE Press; 1994. pp. 213–222.
14. Rao RPN, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nat Neurosci* 1999; **2**:79–87.
15. Gold JI, Shadlen MN. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci* 2001; **5**:10–16.
16. Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A* 2003; **20**:1434–1448.
17. Deneve S. Bayesian inference in spiking neurons. In: Saul LK, Weiss Y, Bottou L, editors. *Advances in neural information processing systems*. Vol. 17. Cambridge, Massachusetts: MIT Press; 2005. pp. 353–360.
18. Zemel RS, Huys QJM, Natarajan R, Dayan P. Probabilistic computation in spiking populations. In: Saul LK, Weiss Y, Bottou L, editors. *Advances in neural information processing systems*. Vol. 17. Cambridge, Massachusetts: MIT Press; 2005. pp. 1609–1616.
19. Rao RPN. Bayesian computation in recurrent neural circuits. *Neural Comput* 2004; **16**:1–38.
20. Mishkin M, Ungerleider LG, Macko KA. Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 1983; **6**:414–417.
21. Gabbiani F, Krapp HG, Hatsopoulos N, Mo CH, Koch C, Laurent G. Multiplication and stimulus invariance in a looming-sensitive neuron. *J Physiol Paris* 2004; **98**:19–34.
22. Priebe N, Ferster D. The origin of cross-orientation suppression in primary visual cortex. *Presentation at Computational and Systems Neuroscience (CoSyNe) Conference*; 2005.
23. Yu A, Dayan P. Inference, attention, and decision in a Bayesian neural architecture. In: Saul LK, Weiss Y, Bottou L, editors. *Advances in neural information processing systems*. Vol. 17. Cambridge, Massachusetts: MIT Press; 2005. pp. 1577–1584.
24. Dayan P, Zemel RS. Statistical models and sensory attention. In: Willshaw D, Murray A, editors. *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. London: IEE Press; 1999. pp. 1017–1022.
25. Rao RPN. Hierarchical Bayesian inference in networks of spiking neurons. In: Saul LK, Weiss Y, Bottou L, editors. *Advances in neural information processing systems*. Vol. 17. Cambridge, Massachusetts: MIT Press; 2005. pp. 1113–1120.