PERGAMON

Neural Networks 15 (2002) 719–730

Neural Networks

www.elsevier.com/locate/neunet

2002 Special issue

Acetylcholine in cortical inference

Angela J. Yu*, Peter Dayan

Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR, UK

Received 5 October 2001; accepted 2 April 2002

Abstract

Acetylcholine (ACh) plays an important role in a wide variety of cognitive tasks, such as perception, selective attention, associative learning, and memory. Extensive experimental and theoretical work in tasks involving learning and memory has suggested that ACh reports on unfamiliarity and controls plasticity and effective network connectivity. Based on these computational and implementational insights, we develop a theory of cholinergic modulation in perceptual inference. We propose that ACh levels reflect the *uncertainty* associated with top-down information, and have the effect of modulating the interaction between top-down and bottom-up processing in determining the appropriate neural representations for inputs. We illustrate our proposal by means of an hierarchical hidden Markov model, showing that cholinergic modulation of contextual information leads to appropriate perceptual inference.

Keywords: Acetylcholine; Perception; Neuromodulation; Representational inference; Hidden Markov model; Attention

1. Introduction

Neuromodulators such as acetylcholine (ACh), serotonine, dopamine, norepinephrine, and histamine play two characteristic roles. One, most studied in vertebrate systems, concerns the control of plasticity. The other, most studied in invertebrate systems, concerns the control of network responses. For instance, a single, recurrently connected assembly of neurons can exhibit multiple dynamical modes (Pflüger, 1999), as neuromodulators alter the excitabilities of individual neurons and the amplitudes of synaptic potentials (Marder, 1998). These two roles have also been brought together, notably in the theoretical and experimental studies of Hasselmo and his colleagues, into the neuromodulatory control of plasticity in recurrently connected neural networks (Hasselmo, 1995; Hasselmo & Bower, 1993). This work sits with that on dopamine (e.g. Schultz, Dayan and Montague, 1997) in proposing *computationally* specific roles for neuromodulation.

Hasselmo and his colleagues (Hasselmo, 1995; Hasselmo & Bower, 1993) focused on cholinergic neuromodulatory influences over learning and memory in the hippocampus and cortex. ACh is delivered to the cortex and hippocampus from a small number of nuclei in the basal forebrain (BF): medial septum (MS), diagonal band of Broca (DBB), and nucleus basalis (NBM). Physiological studies on ACh indicate that its neuromodulatory effects at the cellular level are diverse, causing synaptic facilitation and suppression as well as direct hyperpolarization and depolarization, all within the same cortical area (Kimura, Fukuda, & Tsumoto, 1999). Behavioral experiments indicate that ACh is involved in a wide variety of cognitive functions such as perception, selective attention, associative learning, and memory (Everitt & Robbins, 1997; Hasselmo, 1995; Holland, 1997).

Hasselmo and colleagues proposed that cholinergic (and perhaps other) neuromodulation controls read-in to, and read-out from, recurrently-connected, attractor-like memories, such as that in area CA3 of the hippocampus. Such attractor networks (Amit, 1989) fail if the recurrent connections are operational during storage, since new memories lose their specific identity by being forced to map onto existing memories retrieved through the recurrent dynamics. Hasselmo and colleagues suggested, and collected direct experimental evidence, that cholinergic neuromodulation during storage could selectively suppress but plasticize the recurrent connections (and perhaps the perforant path connections) onto CA3 cells and selectively boost the feedforward mossy fiber inputs from the dentate gyrus. During recall, the recurrent connections should play a fuller part, being comparatively boosted through a lower level of ACh, allowing associative retrieval. The degree of

* Corresponding author. Tel.: +44-20-7679-1195; fax: +44-20-7679-1173.

E-mail addresses: feraina@gatsby.ucl.ac.uk (A.J. Yu), dayan@gatsby.ucl.ac.uk (P. Dayan).

0893-6080/02/$ - see front matter
PII: S0893-6080(02)00058-8

ACh release would reflect the unfamiliarity of the input, and thereby act as a gate to learning. This mechanism has been widely adopted, for instance in our own work (Káli and Dayan, 2000) to help understand how spatial place cells in CA3 might result from a learned surface attractor network. Hasselmo and his colleagues have also demonstrated that ACh has similar physiological and functional effects in the piriform cortex (Linster & Hasselmo, 2001), which is important for olfactory memory.

Lesions studies in classical conditioning tasks provide additional insight into ACh's role in the cortex. Animals are known to learn faster about stimuli whose consequences remain uncertain (Pearce & Hall, 1980). Through an extensive series of selective lesion experiments in rats, Holland and his colleagues have demonstrated that the cholinergic projection from the nucleus basalis magnocellularis (nucleus basalis of Meynert in primates) to the parietal cortex is essential for this sort of faster learning (Holland, 1997; Holland & Gallagher, 1999). These data have been interpreted, using the theoretical viewpoint of statistical learning models, as implying that the ACh signal reports the unfamiliarity of the stimuli, or the *uncertainty* in its predictions (Dayan, Kakade, & Montague, 2000).

In this paper, we present a theory of cortical cholinergic function in perceptual inference based on combining the physiological evidence that ACh can differentially modulate synaptic transmission to control states of cortical dynamics, with theoretical ideas about the information carried by the ACh signal. Crudely speaking, perception involves inferring the most appropriate representation for sensory inputs. This inference is influenced by both top-down inputs, providing contextual information, and bottom-up inputs from sensory processing. We propose that ACh reports on the uncertainty associated with top-down information, and has the effect of modulating the relative strengths of these two input sources. Many cognitive functions affected by ACh levels can be recast in the conceptual framework of representational inference.

In Section 2, we present a simple hierarchical hidden Markov model that casts sensory perception in the theoretical framework of representational inference. As we demonstrate in Section 3, approximate inference in such a model could be mediated by cortical cholinergic innervation. A summary of relevant experimental data and proposals for new experiments is presented in Section 4.

2. Hidden Markov models and perceptual inference

Inferring appropriate representations for the constant stream of sensory inputs is a formidable task, largely because of the inherent ambiguity and noise in the sensory input. A vital source of information that helps resolve ambiguities comes from temporal and spatial context, and thus a key issue for perceptual inference is updating and maintaining this top-down contextual information, and using it correctly in concert with bottom-up information from the sensory input (Grenander, 1995; Helmholtz, 1896; Neisser, 1967).

For simplicity, we only consider one of the most basic forms of top-down contextual information, namely that coming from the recent past. That is, we consider a series of sensory inputs whose internal representations are individually ambiguous. Disambiguation comes via top-down information based on a slowly changing overall state of the environment. Here, only temporal context is relevant; there is no spatial context. The resulting model (see also Becker, 1999) is a form of Hidden Markov Model (HMM). The HMM captures the way that sensory inputs are generated or synthesized. We consider the inferential task of recognition or analysis in which the representation for each input is determined. We compare an approximate recognition model based on cholinergic neuromodulation, with the exact recognition model (Rabiner, 1989), in a case chosen so that the exact model is computationally tractable.

Our HMM (Fig. 1(A) and (B)) consists of three pieces. One, z_t, is the overall state of the environment at time t, which we also call the context. Changes to z_t are stochastically controlled by a *transition matrix* $\mathcal{T}_{z_{t-1}z_t}$, whose entries ensure that the context changes rarely. The second piece is y_t, which is determined stochastically on each time-step, in a way that depends on the current state of the environment. The third piece, the observed input $\mathbf{x}_t$, depends stochastically on y_t. The inferential task is to represent inputs $\mathbf{x}$ in terms of the y values that were responsible for them. However, the relationship between y_t and $\mathbf{x}_t$ is such that this is ambiguous, so top-down information from the likely states of z_t, i.e. the likely context, is important to find the correct representation for $\mathbf{x}_t$. Fig. 1(A) shows the probabilistic contingencies among the variables. Fig. 1(B) shows the same contingencies in a different way, and specifies the particular setting of parameters used to generate the examples found in the remainder of the paper.

More formally, the context is a discrete, hidden, random variable z_t, whose stochastic temporal dynamics are described by a Markov chain with transition matrix $\mathcal{T}_{z_{t-1}z_t}$ as

$$P[z_t|z_{t-1}] \equiv \mathcal{T}_{z_{t-1}z_t} = \begin{cases} \gamma & \text{if } z_t = z_{t-1}, \\ \dfrac{1-\gamma}{n_z - 1} & \text{otherwise,} \end{cases} \tag{1}$$

where n_z is the number of all possible states of z, and γ is the probability of persisting in one text. When γ is close to 1, as is the case in the example of Fig. 1(B), the context tends to remain the same for a long time. When γ is close to 0, the context tends to switch among the different states of z rapidly and randomly. The state of the second hidden layer, y is generated from z with the mapping $O_{z_t y_t}$, which specifies $P[y_t|z_t]$, and controls which of a set of circular two dimensional Gaussians is used to generate the observations $\mathbf{x}_t$ via the densities $p[\mathbf{x}|y]$. The y_t that was actually involved

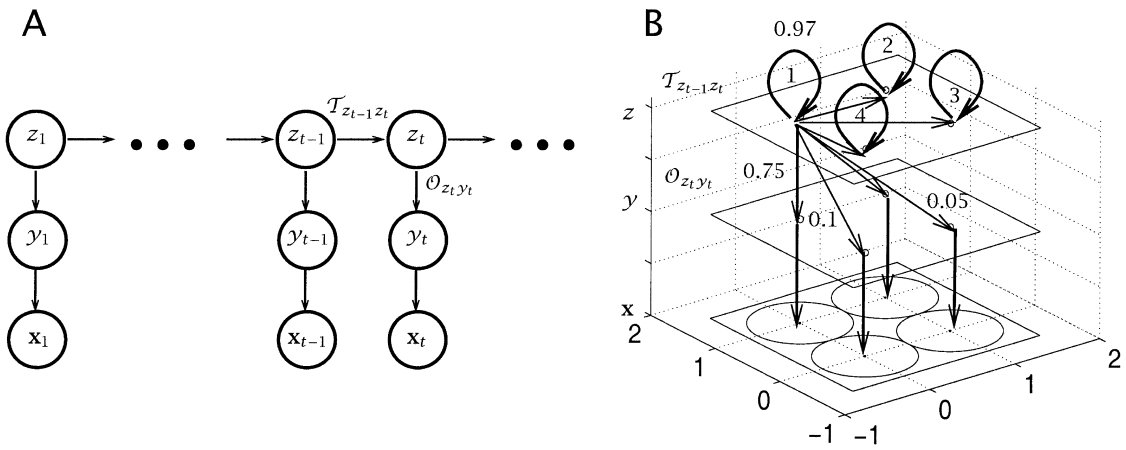

Fig. 1. Hierarchical HMM. (A) Three-layer model, with two hidden layers, z and y, and one observed layer, $\mathbf{x}$. The temporal dynamics are captured by the transition matrix $\mathcal{T}_{z_{t-1}z_t}$ in the z layer, and the observations $\mathbf{x}$ are generated from y and, indirectly, from z. (B) Example parameter settings: $z \in \{1-4\} \Rightarrow y \in \{1-4\} \Rightarrow \mathbf{x} \in \Re^2$ with dynamics ($\mathcal{T}$) in the z layer ($P[z_t = z_{t-1}] = 0.97$), a probabilistic mapping (O) from $z \rightarrow y$ ($P[y_t = z_t|z_t] = 0.75$), and a Gaussian model $p[\mathbf{x}|y]$ with means at the corners of the unit square and standard deviation $\sigma = 0.5$ in each direction. Only some of the links are shown to reduce clutter.

in generating $\mathbf{x}_t$ is also called the model's (true) *representation* of $\mathbf{x}_t$. The means of the Gaussians $p[\mathbf{x}|y]$ are at the corners of the unit square, as shown in Fig. 1(B), and the variances of these Gaussians are $\sigma^2 I$. The parameters in the model are the prior distribution of z, its temporal dynamics $\mathcal{T}_{z_{t-1}z_t}$, the conditional distributions O_{zy}, and the emission densities $p[\mathbf{x}|y]$. It is assumed that all the parameters have already been correctly learned at the outset of the inference problem.

Fig. 2 shows an example of a sequence of 400 states generated from the model. The state in the z layer stays the same for an average of about 30 time steps, and then switches to one of the other states, chosen with equal probability. The key inference problem is to determine the posterior distribution over y_t, that best explains the observation $\mathbf{x}_t$, given the past experiences

$$\mathcal{D}_{t-1} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}\}.$$

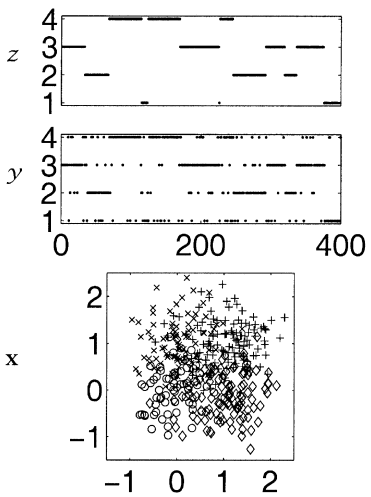

Fig. 2. Generative model. A sample sequence involving 400 time steps, generated from the model shown in Fig. 1(B). Note the slow dynamics in z, the stochastic mapping into y, and substantial overlap in $\mathbf{x}$s generated from the different ys (different symbols correspond to different Gaussians shown in Fig. 1(B)).

Inference of the 'true' posterior distribution, $P[y_t|\mathbf{x}_t, \mathcal{D}_{t-1}] = \mathcal{P}[y_t|\mathcal{D}_t]$, uses temporal contextual information, consisting of existing knowledge built up from past observations, as well as the new observation $\mathbf{x}_t$. Fig. 3(A) shows the structure of the standard HMM inference model, where the posterior distributions $P[y_t|\mathcal{D}_t]$ and $P[z_t|\mathcal{D}_t]$ can be computed using a procedure that is equivalent to the forwards part of the forwards–backwards algorithm (Rabiner, 1989). The adaptation to include the y layer is straightforward.

In each time step t, the top-down information is communicated by z_t, while the bottom-up information is carried by $\mathbf{x}_t$. The prior distribution over z_t

$$P[z_t|\mathcal{D}_{t-1}] = \sum_{z_{t-1}} P[z_{t-1}|\mathcal{D}_{t-1}]\mathcal{T}_{z_{t-1}z_t}, \tag{2}$$

distills the contextual information from past experiences $\mathcal{D}_{t-1}$. This information is propagated to the representational units y by

$$P[z_t, y_t|\mathcal{D}_{t-1}] = P[z_t|\mathcal{D}_{t-1}]O_{z_t y_t}. \tag{3}$$

The bottom-up information, $P[y_t|\mathbf{x}_t]$, is proportional to the likelihood, $p[\mathbf{x}_t|y_t]$, and interacts with the top-down information, $P[z_t, y_t|\mathcal{D}_{t-1}]$ in the conditioning step

$$P[z_t, y_t|\mathcal{D}_t] \propto P[z_t, y_t|\mathcal{D}_{t-1}]p[\mathbf{x}_t|y_t], \tag{4}$$

where the constant of proportionality normalizes the full conditional distribution. From the joint posterior distribution over z and y, we can then compute the marginalized posterior distribution of y_t, which gives the relative belief in each of the states of y_t having generated the current

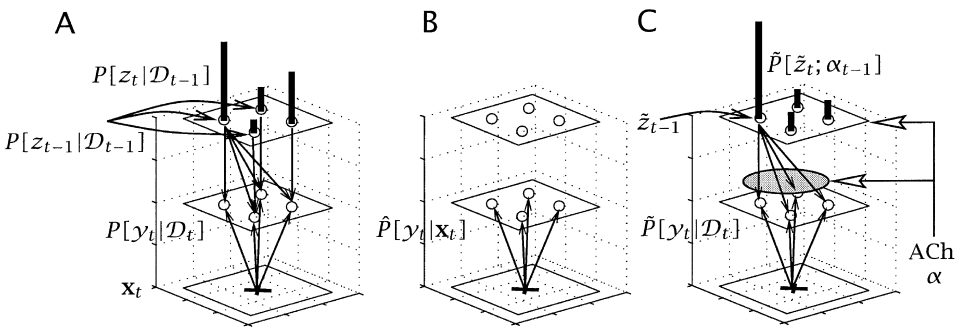

Fig. 3. Recognition models. (A) Exact recognition model. $P[z_{t-1}|\mathcal{D}_{t-1}]$ is propagated to provide the prior $P[z_t|\mathcal{D}_{t-1}]$ (shown by the lengths of the thick vertical bars), and thus the prior $P[y_t|\mathcal{D}_{t-1}]$. This is combined with the likelihood term from the data $\mathbf{x}_t$ to give the true $P[y_t|\mathcal{D}_t]$. (B) Bottom-up recognition model uses only a generic prior over y_t, which conveys no information, so the likelihood term dominates. (C) ACh model. A single estimated state $\tilde{z}_{t-1}$ is used, in conjunction with its uncertainty α_{t-1}, reported by cholinergic activity, to produce an approximate prior $\tilde{P}[\tilde{z}_t;\alpha_{t-1}]$ over z_t (which is a mixture of a delta function and a uniform), and thus an approximate prior over y_t. This is combined with the likelihood to give an approximate $\tilde{P}[y_t|\mathcal{D}_t]$, and a new cholinergic signal α_t is calculated.

observation $\mathbf{x}_t$, in the context of past experiences:

$$P[y_t|\mathcal{D}_t] = \sum_{z_t} P[z_t, y_t|\mathcal{D}_t]. \tag{5}$$

This distribution, henceforth referred to as the exact posterior, is the fullest possible representation of $\mathbf{x}_t$. One can also create the updated contextual information

$$P[z_t|\mathcal{D}_t] = \sum_{y_t} P[z_t, y_t|\mathcal{D}_t], \tag{6}$$

which is propagated forward to the next time step $(t+1)$ as in Eq. (2).

Figs. 4 and 5 show various aspects of inference in the HMM for a particular run. The true contextual states $\{z_1^*, z_2^*, \ldots\}$, the true representational states $\{y_1^*, y_2^*, \ldots\}$, and the observations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$, are generated from the model with the parameters given in Fig. 1(B). The posterior distributions over z_t and y_t given $\mathcal{D}_t$, that is all the observations up to, and including time t, are computed at each time step using the algorithm detailed earlier. If the algorithm is working properly, then we would expect to see a high correspondence between the true contextual state z_t^* and the inferred, most likely state $z_t^0 = \text{argmax}_{z_t} P[z_t|\mathcal{D}_t]$. Fig. 4(A) and (B) shows that z_t^0 mostly replicates z_t^* faithfully. One quirk of inference in HMMs is that these individually most likely states z_t^0 do not form a most likely state *sequence* as, for instance, found by the Viterbi algorithm.

Fig. 5(A) and (B) show histograms of the representational posterior probabilities of the true states y_t^* and all the other possible states $\bar{y}_t^*$, respectively, computed by the exact inference algorithm. As one might hope, the former are generally large and cluster around 1, while the latter are generally small and cluster around 0.

The exact inference algorithm that we have described achieves good performance. However, one may well ask whether it is computational feasible for the brain to perform the complete, exact inference in all its mathematical complexity. Viewed abstractly, the most critical problem seems that of maintaining and manipulating simultaneously the information about all possible contexts ($P[z_t|\mathcal{D}_t]$). This is particularly difficult in the face of population coding, for which the activity patterns of one or a few populations of units in relevant cortical areas are used to represent all possible contexts. Of course, in our simple example, there are only four possible contexts. However, in general, there are potentially as many contexts as known environments, a huge number.

A 'naive' solution to the complexity problem is to use only the likelihood term, $p[\mathbf{x}_t|y_t|]$, in the inference about the current representational states y_t, and ignore the top-down contextual information altogether. This is actually one traditional model of inference for unsupervised

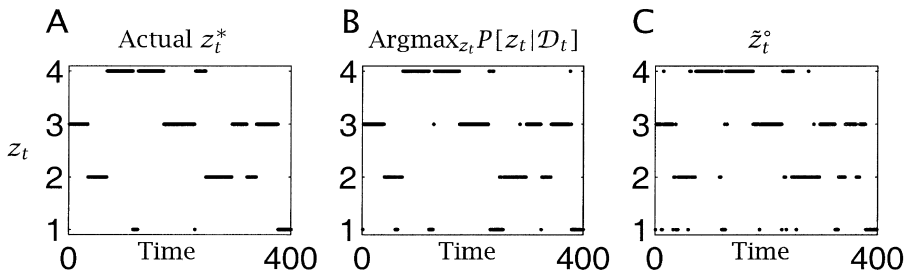

Fig. 4. Contextual representation in exact inference. (A) Actual z_t states. (B) Highest probability z_t states from the exact posterior distribution. (C) Most likely $\tilde{z}_t$ states from the ACh-mediated approximate inference model.

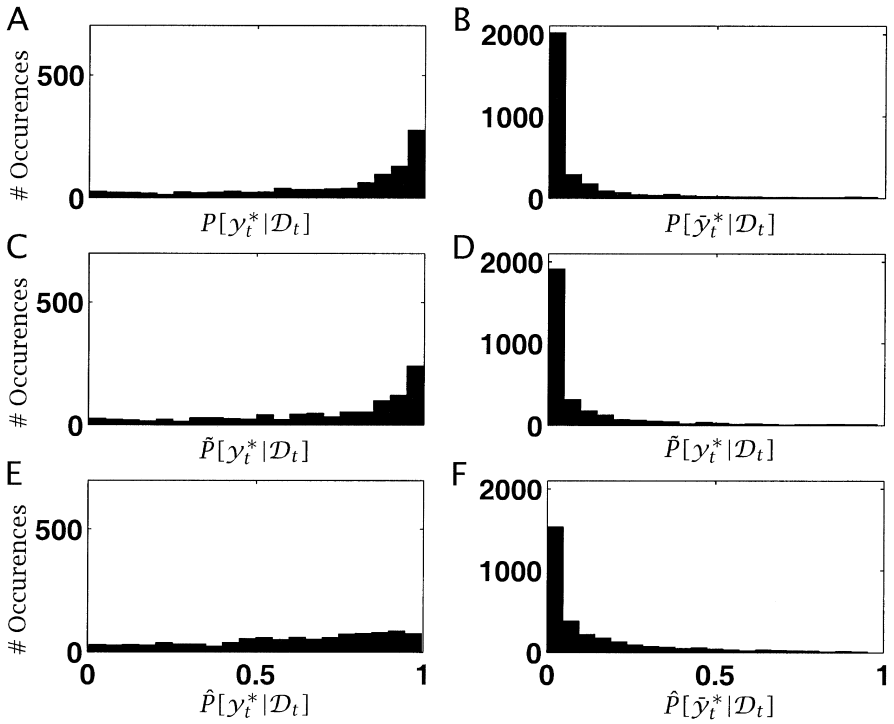

Fig. 5. Quality of exact inference. Histograms of the posterior distributions of the true state y_t^* (A, C, and E) and all other possible states $y_t \neq y_t^*$ (B, D, and F), tallied over 1000 time steps of a run. A and B are based on the exact inference algorithm, C and D on the ACh-mediated inference algorithm, and E and F on the bottm-up inference algorithm. The x-axis is divided into bins of $P[y_t|\mathcal{D}_t]$ ranging from 0 to 1, and the y-axis refers to the number of occurrences that probability accorded to y_t^* or $\bar{y}_t^*$ falls into each of the binned probability intervals in the posterior distributions. Note that the histograms in the right column have larger entries than those in the left, because at each time steps, only the true state contributes to the histogram on the left, while the other three contribute to the right. The differential degrees of similarity between the histograms produced by the ACh algorithm compared to the exact algorithm, and by the bottom-up algorithm compared to the exact algorithm, are an indication of their respective quality of representational inference.

analysis-by-synthesis models (e.g. Hinton & Ghahramani, 1997). Fig. 3(B) shows the structure of a purely bottom-up model, where the approximate posterior is computed by $\hat{P}[y_t|\mathbf{x}_t] = p[\mathbf{x}_t|y_t]/Z$, where Z is a normalization factor. Purely bottom-up inference solves the problem of high computational costs: there is no need to carry any information from one time step to the next. However, the performance of this algorithm is likely to be poor, whenever the probability distribution of generating $\mathbf{x}$ for the different values of y overlap substantially, as is the case in our example. This is just the ambiguity problem described earlier.

Fig. 6(A) shows the representational performance of this model, through a scatter-plot of $\hat{P}[y_t|\mathbf{x}_t]$ against the exact posterior $P[y_t|\mathcal{D}_t]$. If bottom-up inference were perfectly correct, then all the points would lie on the diagonal line of equality. The bow-shape shows that purely bottom-up inference is relatively poor. The particularly concentrated upper and lower boundaries indicate that when the true posterior distribution assigns a very high or very low probability to a state of y, the corresponding distribution inferred from bottom-up information alone tends to assign a much more neutral probability to that state. This tendency highlights the loss of the contribution of the disambiguating top-down signal in the bottom-up model. With only the bottom-up information, it rarely happens that one can say with confidence that a state of y_t is either definitely the one, or definitely not the one, that generated $\mathbf{x}_t$. The exact shape of the envelope is determined by the extent of overlap in the densities $p[\mathbf{x}|y]$ for the various values of y, but we do not analyze this relationship in detail here.

3. ACh-mediated approximate inference

A natural compromise between the exact inference model, which is representationally and computationally expensive, and the naive inference model, which has poor performance, is to use a model that captures useful top-down information at a realistic computational cost. The intuition we gain from exact inference is that top-down expectations can resolve bottom-up ambiguities, permitting better processing. However, in the face of contextual uncertainty, top-down information is just generic. Thus, we consider a model in which just a single contextual state is represented in the activity of contextual units (presumably in pre-frontal areas), and ACh is used to report on the uncertainty of this contextual state and to control the balance between bottom-up and top-down inference. In exact inference, the notion of uncertainty is captured in the (entropy of) the posterior distribution of the contextual state

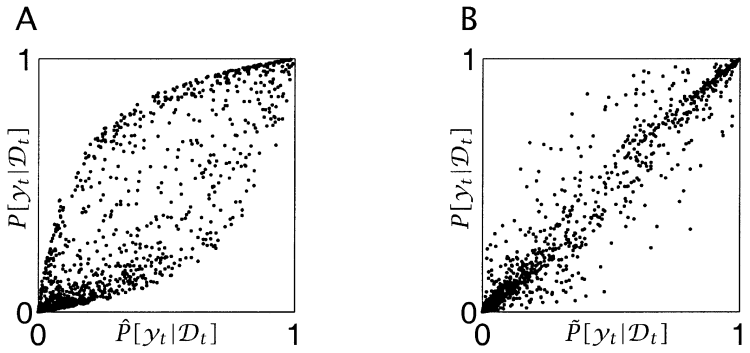

Fig. 6. Representational performance. Comparison of (A) the purely bottom-up $\hat{P}[y_t|\mathcal{D}_t]$ and (B) the ACh-based approximation $\tilde{P}[y_t|\mathcal{D}_t]$, with the true $P[y_t|\mathcal{D}_t]$ across all values of y_t. The ACh-based approximation is substantially more accurate.

$P[z_{t-1}|\mathcal{D}_{t-1}]$ in Eq. (2). This uncertainty determines the relative strength of the top-down information, $P[z_t, y_t|\mathcal{D}_{t-1}]$, compared with the information from the likelihood $p[\mathbf{x}_t|y_t]$, in Eq. (4).

More formally, in our ACh-mediated approximate inference model, only two pieces of information about the context are maintained over time: $\tilde{z}_{t-1}$, the most likely contextual *state* having seen $\mathcal{D}_{t-1}$, and α_{t-1}, the amount of uncertainty associated with that state. The idea is that α_{t-1} is reported by the level of ACh, and is used to control the extent to which top-down information based on $\tilde{z}_{t-1}$ is used to influence inference about y_t.

Fig. 3(C) shows a schematic diagram of the proposed approximate inference model. If we were given the full, exact posterior distribution $P[z_{t-1}, y_{t-1}|\mathcal{D}_{t-1}]$, then one natural definition for this ACh signal would be the uncertainty in the most likely contextual state

$$\alpha_{t-1} = 1 - \max_z P[z_{t-1} = z|\mathcal{D}_{t-1}] = 1 - P[z^\circ_{t-1}|\mathcal{D}_{t-1}]. \quad (7)$$

Fig. 7(B) shows the resulting ACh signal for one run with the actual sequence of contextual states shown in Fig. 7(A). As expected, the ACh level is generally high at times when the true state z^*_t is changing, and decreases during the periods that z^*_t is constant. During times of change, top-down information is confusing or potentially incorrect, and so the current context should be abandoned while a new context is gradually built up from a period of perception that is mainly dominated by bottom-up input. This switch in inferential strategy is just the putative inferential effect of ACh.

The ACh signal of Fig. 7(B) was calculated assuming knowledge of the true posterior. This is, of course, unreasonable. The model of Fig. 3(C) includes the key approximation that the only information from $\mathcal{D}_{t-1}$ about the state of z, besides uncertainty signaled by ACh, is in the single choice of context variable $\tilde{z}_{t-1}$. As in the full inference model, the first computation at each time step t is to compute the prior distribution over z_t. However, since the full posterior distribution of z_{t-1} is no longer available, it has to be reconstructed from what is believed to be the most

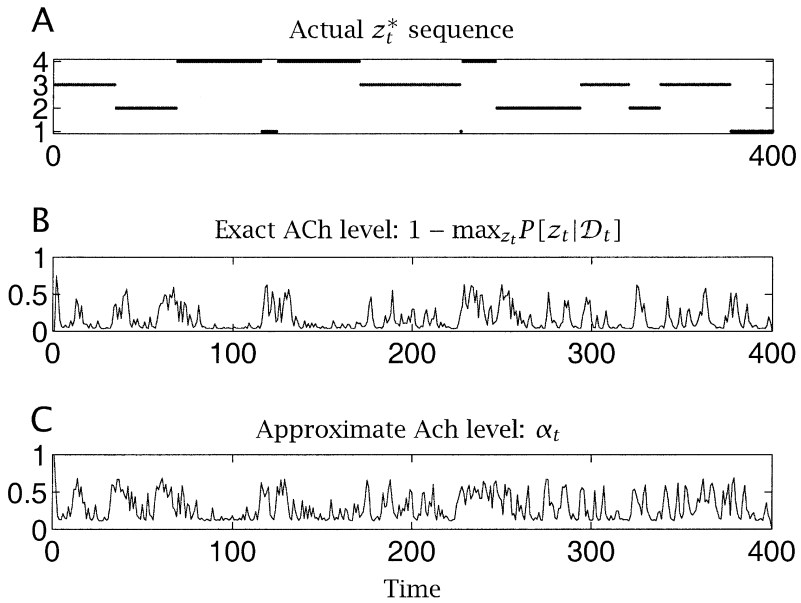

Fig. 7. ACh model. (A) Actual sequence of contextual states z_t for one run. (B) ACh level from the exact posterior in the same run. (C) ACh level α_t from the approximate model. Note the coarse similarity between A and B.

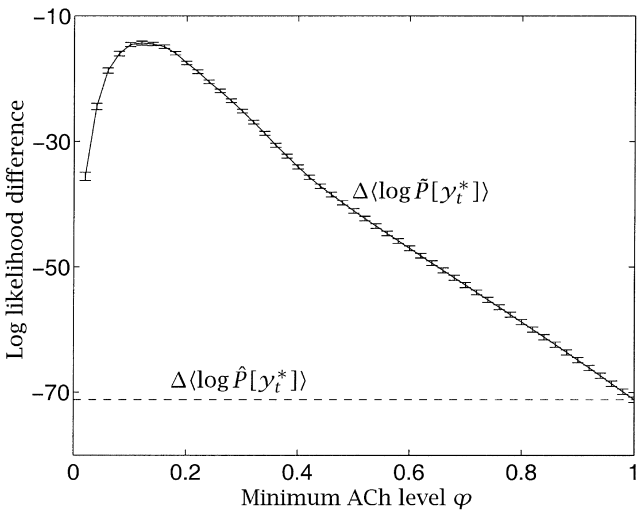

Fig. 8. Representational cost. Solid: the mean extra representational cost for the true state y_t^* over that in the exact posterior using the ACh model as a function of the minimum allowed ACh level φ. Dashed: the same quantity for the pure bottom-up model (which is equivalent to the approximate model for $\varphi = 1$), denoted $\Delta\langle\log\hat{P}[y_t^*]\rangle$ here. Errorbars show standard errors of the means over 1000 trials.

likely contextual state $\tilde{z}_{t-1}^0$, together with its associated uncertainty α_{t-1} :

$$\tilde{P}[\tilde{z}_{t-1};\alpha_{t-1}] = \begin{cases} (1-\alpha_{t-1}) + \dfrac{\alpha_{t-1}}{n_z} & \text{if } \tilde{z}_{t-1} = \tilde{z}_{t-1}^\circ, \\ \dfrac{\alpha_{t-1}}{n_z} & \text{otherwise.} \end{cases} \tag{8}$$

The approximation made is that all the non-explicitly modeled states of z_{t-1} equally share a fraction of the probability that $\tilde{z}_{t-1}$ was not the true context. As before, the contextual information is then propagated to z_t and y_t

$$\tilde{P}[\tilde{z}_t;\alpha_{t-1}] = \sum_{z_{t-1}} \tilde{P}[\tilde{z}_{t-1};\alpha_{t-1}]\mathcal{T}_{z_{t-1}z_t}, \tag{9}$$

$$\tilde{P}[y_t,\tilde{z}_t;\alpha_{t-1}] = \tilde{P}[\tilde{z}_t;\alpha_{t-1}]O_{\tilde{z}_t y_t}, \tag{10}$$

and the new observation is incorporated into the inference in the conditioning step:

$$\tilde{P}[y_t,\tilde{z}_t|\mathcal{D}_t] \propto \tilde{P}[y_t,\tilde{z}_t;\alpha_{t-1}]p[\mathbf{x}_t|y_t]. \tag{11}$$

The new posterior distributions are computed, as before, by marginalizing the joint posterior distribution:

$$\tilde{P}[y_t|\mathcal{D}_t] = \sum_{z_t} \tilde{P}[y_t,\tilde{z}_t|\mathcal{D}_t], \tag{12}$$

$$\tilde{P}[\tilde{z}|\mathcal{D}_t] = \sum_{y_t} \tilde{P}[y_t,\tilde{z}_t|\mathcal{D}_t]. \tag{13}$$

In addition, we compute the contextual information that is to be propagated to the next time step:

$$\tilde{z}_t^0 = \text{argmax}_{z_t}\tilde{P}[\tilde{z}_t|\mathcal{D}_t] \text{ most likely contextual state}, \tag{14}$$

$$\alpha_t = 1 - \tilde{P}\left[\tilde{z}_t^0|\mathcal{D}_t\right] \text{ Ach level}. \tag{15}$$

The crucial differences between this approximate inferential algorithm and the exact one detailed before are the use of ACh as a scalar measure of uncertainty (Eq. (15)) and the reconstruction of an estimate of the posterior distribution from this scalar estimate (Eq. (8)). In general, we would not expect to be able to represent this distribution either. Rather, ACh would control the way that information about just $\tilde{z}_{t-1}$ would be used to make inferences at time t. If α_{t-1} (i.e. the ACh level) is high, then the input stimulus-bound likelihood term dominates in the conditioning process (Eq. (11)); if α_{t-1} (i.e. the ACh level) is low, then the temporal context ($\tilde{z}_{t-1}$) and likelihood terms are appropriately balanced. These computations are local and straight forward, except for the representation and normalization of the joint distribution over y_t and $\tilde{z}_t$.

One potentially dangerous aspect of this inference procedure is that it might get unreasonably committed to a single state: $\tilde{z}_{t-1} = \tilde{z}_t = \cdots$. Because the probabilities accorded to the other possible values of z_{t-1} given $\mathcal{D}_{t-1}$ are not explicitly represented from one time step to the next, there is little chance for uncertainties about a context to build up, a condition necessary for inducing a context switch. A natural way to avoid this is to bound the ACh level from below by a constant, φ, making approximate inference slightly more stimulus-bound than exact inference. Thus, in practice, rather than using Eq. (15), we use

$$\alpha_t = \varphi + (1-\varphi)\left(1 - \tilde{P}\left[\tilde{z}_t^0|\mathcal{D}_t\right]\right). \tag{16}$$

Larger values of φ lead to a larger *guaranteed* contribution of the bottom-up, stimulus-bound likelihood term to inference.

Fig. 7(C) shows the approximate ACh level for the same case as in Fig. 7(A) and (B), using $\varphi = 0.1$. Although the detailed value of this signal over time is clearly different from that arising from an exact knowledge of the posterior probabilities in Fig. 7(B), the gross movements are quite similar. Note the effect of preventing the ACh level from dropping to 0. Fig. 5(C) and (D) show that the ACh-mediated approximate inferences has the same tendency as the exact algorithm (Fig. 5(A) and (B)) to accord high probabilities to the true sequence of states, y_t^*, and low probabilities to all the other states, $\bar{y}_t^*$. In comparison, the bottom-up model performs much worse (Fig. 5(E) and (F)), tending in general to give the true states, y_t^*, lower probabilities. Fig. 6(B) shows that the ACh-based approximate posterior values $\tilde{P}[y|\mathcal{D}]$ are much loser to the true values than for the purely bottom-up model, particularly for values of $P[y_t|\mathcal{D}_t]$ near 0 and 1, where most data lie. Fig. 4(C) shows that inference about z_t^* is noisy, but the pattern of true values is certainly visible.

Fig. 8 shows the effects of different φ on the quality of inference about the true states y_t^*. What is plotted is the difference between approximate and exact log probabilities of the true states y_t^*, averaged over 1000 cases. The average log likelihood for the exact model is -210. If $\varphi = 1$, then

inference is completely stimulus-bound, just like the purely bottom-up model.

Note the poor performance for this case. For values of φ slightly less than 0.2, the approximate inference model does well, both for the particular setting of parameters described in Fig. 1(B) and for a range of other values (not shown here). An upper bound on the performance of approximate inference can be calculated in three steps by: (i) using the exact posterior to work out $\tilde{z}_t$ and α_t, (ii) using these values to approximate $P[\tilde{z}_t; \alpha_t]$ as in Eq. (8), and (iii) using this approximate distribution in Eq. (10) and the remaining equations. The average resulting cost (i.e. the average resulting difference from the log probability under exact inference) is -3.5 log units. Thus, the ACh-based approximation performs well, and much better than purely bottom-up inference.

4. Experimental data

We have argued that accurate reporting of top-down uncertainty, signaled by ACh level, is important for achieving an *appropriate* balance between top-down and bottom-up processing in perceptual inference. To validate this statement, two main properties of cholinergic modulation need to be verified. One is the modulatory effect of ACh on the relative processing of top-down and bottom-up information. The other is that it is uncertainty that controls ACh release in relevant areas of the cortex. We know of no experiment that directly tests our theory. However, data from various behavioral and physiological studies lend some support to its elements. For example, abnormal ACh levels, due to either pharmacological manipulations or neurological diseases, lead to characteristic deficits in attentional perceptual tasks (Sarter & Bruno, 1998) and general behavioral symptoms such as hallucination (Perry & Perry, 1995). Also, physiological data indicate that the cellular and network effects of ACh activation on the processing of sensory stimuli are *facilitatory*, as we would expect, although there is not yet data showing what specific effects ACh has on top-down information.

Unfortunately, there is little detailed information on what drives the activity of the cholinergic cells. It is known that endogenous, task-related release of ACh occurs shortly before the presentation of stimuli, as measured by microdialysis (Fadel, Sarter, & Bruno, 2001) and single-cell recordings (Richardson & DeLong, 1991), indicating at a minimum that ACh release cannot be a simple consequence of ongoing perceptual processing. Further, some candidate anatomical pathways have been identified (such as from the central nucleus of the amygdala, Bucci, Holland and Gallagher, 1998), but not at the level of detail required to refute or support the theory.

After summarizing the relevant, existing data, we will propose some experiments to investigate aspects of ACh-mediated perception that are less well understood.

4.1. General behavioral effects of Ach

ACh involvement has been found in a variety of attentional tasks, such as versions of sustained attention, selective spatial attention, and divided attention. The attentional tasks studied in association with ACh can generally be viewed as top-down/bottom-up inferential tasks with elements of uncertainty. First, in the case of sustained attention in signal detection (Sarter, Givens, & Bruno, 2001), the rare occurrence of signals may lead to a top-down bias of not detecting a signal. Thus, abnormally low levels of cortical ACh, according to our theory, would lead to undue confidence in non-detection when signal does appear, and excessive ACh release would result in over-processing of sensory information, leading to higher rates of false alarms but not misses. Pharmacological manipulations of ACh in monkeys induce behavioral deficits consistent with this view (Holley, Turchi, Apple, & Sarter, 1995; McGaughy, Kaiser, & Sarter, 1996; Turchi & Sarter, 2001). Second, in Posner's task, a version of selective spatial attention, the cue provides top-down information about the location of the stimulus (Jackson, Marrocco, & Posner, 1994; Muir, Dunnett, Robbins, & Everitt, 1992), and uncertainty can be induced by presenting the stimulus at the unexpected location. As would be consistent with the theory, abnormal ACh levels, induced by pharmacological manipulations (Muir et al., 1992), BF lesions (Muir, Everitt, & Robbins, 1994), and neurological diseases (Parasuraman, Greenwood, Haxby, & Grady, 1992), result in longer reaction times and sometimes lower accuracy. Third, in a version of a divided attention task with modality uncertainty, each session of signal detection is either bimodal (mixed auditory and visual cues) or unimodal (only auditory or only visual) (Turchi & Sarter, 1997). The session serves as an implicit source of contextual information about the modality of the signal, thus the uncertainty level is inherently higher in the bimodal session than the unimodal one. After a BF lesion (Turchi & Sarter, 1997), rats have longer reaction time than controls in the bimodal case, as they are no longer capable of giving bottom-up information higher priority in the face of top-down uncertainty.

A further source of behavioral data on cholinergic modulation comes from patients with neurological diseases. Hallucination is common among patients diagnosed with Lewy Body Dementia, Parkinson's disease, and Alzheimer's disease, all of which are accompanied by some degree of cortical cholinergic deficit. In the context of our model, this route to hallucination might reflect over-processing of top-down information due to an ACh deficit. The cholinergic nature of hallucination is supported by the observed correlation between the severity of hallucination and the extent of cholinergic depletion (Perry et al., 1993). Consistent with the notion that hallucination is antagonistic to sensory processing, hallucinatory experiences induced by plant chemicals containing anti-muscarinic agents such as scopolamine and atropine (Schultes & Hofmann, 1992) are

enhanced during eye closure and suppressed by visual input (Fisher, 1991). Many patients with Lewy Body Dementia and Alzheimer's disease also exhibit pereidolias, or the discernment of images such as faces or animals in wallpaper, curtains, or clouds (Perry & Perry, 1995), a condition ameliorated by the administration of physostigmine, an ACh reuptake-inhibitor (Cummings, Gorman, & Shapira, 1993). Of course, there are also other routes to hallucinations, notably due to hyperactivities of cholinergic neurons in the pedunculopontine nucleus and dorsolateral tegmental nucleus (Perry & Perry, 1995), as well as via serotonin receptors (e.g. Jacobs, 1978).

4.2. Physiological effects of ACh

Physiological studies, though traditionally focusing only on the effects of ACh on bottom-up, stimulus-bound processing, have suggested a much more specific set of cholinergic effects than behavioral studies. A large body of physiological data from both anesthetized and awake animals supports the notion that BF ACh activation enhances stimulus processing across sensory cortices (Metherate, Asche, & Weinberger, 1990; Sillito & Murphy, 1987; Tremblay, Warren, & Dykes, 1990). For example, tetanic stimulation in the NBM increases cortical responsiveness by facilitating the ability of synaptic potentials in thalamo-cortical connections to elicit action potentials in the rat auditory cortex (Hars, Maho, Edeline, & Hennevin, 1993; Metherate, Asche, & Weinberger, 1993), an effect blocked by the application of atropine. Similarly, iontophoretic application of ACh in somatosensory cortex (Donoghue & Carroll, 1987; Metherate, Tremblay, & Dykes, 1987) and visual cortex (Sillito & Kemp, 1983) enhances stimulus-evoked discharges and short-term potentiation without concomitant loss in selectivity.

At the network level, ACh seems selectively to promote the flow of information in the feed-forward pathway over that in the top-down feedback pathway. Existent data suggest that ACh selectively enhances thalamo-cortical synapses via nicotinic receptors (Gil, Conners, & Amitai, 1997) and strongly suppresses intracortical synaptic transmission in the visual cortex through pre-synaptic muscarinic receptors (Kimura et al., 1999). In separate studies, ACh has been shown selectively to suppress synaptic potentials elicited by the stimulation of layer I, which contains a high percentage of feedback synapses, while having no effect on synaptic potentials elicited by the stimulation of layer IV, which has a high percentage of what we would consider as feed-forward synapses (Hasselmo & Cekic, 1996). Collectively, these data suggest ACh activation enables the stimulus-bound input to have a dominant effect in sensory processing.

4.3. New experiments

From the perspective of our model, the two areas in which data on cortical cholinergic modulation are particular lacking are the effects of ACh on top-down processing relative to bottom-up processing and the control of ACh activation. Investigating these issues requires a perceptual task in which the contextual and sensory inputs are clearly distinct, and both they and their inferential implications can be carefully controlled.

One possible paradigm would be a modification of that described in Ress et al. (Ress, Backus, & Heeger, 2000), which uses fMRI techniques to measure visual cortical activity during a stimulus detection task. Their data show a strong, stimulus-independent, spatially selective (for the region in which the stimulus is likely to appear), and putatively top-down signal correlated with detection performance. In terms of our model, this top-down signal would correspond to something like $\tilde{P}[y_t, \tilde{z}_t; \alpha_{t-1}]$ or $\tilde{P}[y_t; \alpha_{t-1}]$. To ascertain the effects of top-down uncertainty on this signal, it is possible to manipulate uncertainty by changing the average ratio of signal to non-signal trials in a given session. Or, more radically, something closer to the selective spatial attention tasks as described earlier could be introduced in the task. When uncertainty is high, the signal should be weak; when uncertainty is low, the signal should be strong. If this signal indeed varies as a function of uncertainty as expected, then the next step would be to investigate the relationship between ACh level, this fMRI signal, and the subject's behavioral performance. The ACh level can be manipulated with the administration of ACh receptor agonists/antagonists, ACh re-uptake inhibitors, or other drugs with known effects on cortical ACh. We predict that elevated ACh levels would lead to a weak top-down signal, and selectively impaired performance with increased false alarms but not misses. In contrast, lowered ACh level would lead to a stronger top-down signal and impairment in performance reflected mainly in the lengthening of reaction time.

A direct relationship between top-down uncertainty and ACh level is difficult to establish using the Ress paradigm, as the available techniques for measuring ACh level—single-cell recordings and microdialysis—are inappropriate for fMRI studies. However, this relationship can be separately investigated by adapting the same task to animal models. The top-down signal, presumably similar to that for humans, would no longer be monitored. However, it would be possible to measure the level of ACh using either single-cell recordings in the NBM, the main source of cortical ACh, or microdialysis in the visual cortex itself. Top-down uncertainty could again be manipulated by varying the average frequency of signal trials relative to non-signal trials. Note that, due to the lack of data, it is difficult to hypothesize a priori the precise relationship between uncertainty and the concentration of ACh. In our mathematical model, we implicitly assumed a linear relationship and ignored the fact that ACh release and effects have various components at different time scales (Hasselmo, 1995; Sarter & Bruno, 1997). It will be very interesting to

find out the exact relationship between uncertainty and ACh level, and how this relationship differs for the tonic and phasic components of ACh signal.

5. Discussion

We have suggested that one role for ACh in cortical processing is to report contextual uncertainty in order to control the balance between stimulus-bound, bottom-up processing, and context-bound, top-down processing. This is an extension of computational and implementational ideas about the involvement of cholinergic modulation in learning in attractor networks. We used the example of a hierarchical HMM, in which representational inference for a middle layer correctly reflects such a balance, and showed that a simple model of the drive and effects of ACh leads to competent inference.

Our suggestion bears an interesting relationship to a crucial mechanism in adaptive resonance theory (ART; Carpenter & Grossberg, 1991). In ART, a representation of an input is chosen via pattern matching, and its appropriateness is assessed using top-down signals. If the representation is found to be wanting, then a neuromodulatory mechanism is used to reset the units in the representational layer, thus allowing a different representation to be selected. The models have in common the idea that neuromodulation should alter network state or dynamics based on information associated with top-down processing. However, the nature of the information that controls the neuromodulatory signal, and the effect of neuromodulation on cortical inference are quite different in the two models.

The mathematical model we have used to illustrate our general theory on ACh is overly simple in several respects. For example, it uses a localist representation for the state z, so that exact inference is feasible. Also, only a two-level hierarchy was modeled, whereas sensory systems in the brain are known to involve many levels of processing. It would be more biologically realistic to consider distributed representations at each of many levels in a hierarchy, and in which only one or a very few contexts, presumably stored in the pre-frontal cortex, could be entertained at once. Also, it is necessary to modify the steps in Eqs. (10) and (11), since it would be difficult to represent the joint uncertainty over representations at multiple levels in the hierarchy. Further, despite the limited physiological evidence mentioned above, exactly which sets of cortical connections should be modulated by ACh is not completely clear (see also Dayan, 1999). For instance, to what extent should recurrent connections within a cortical column, or within a cortical hypercolumn be influenced in the same way as long-range horizontal connections between hypercolumns, or top-down connections from cortical areas higher in the cortical hierarchy?

A strong assumption made in using a HMM is that the sequence of contextual states obeys the Markov property: the context at any particular time step only depends on the context in the preceding step and not on any of the previous ones. However, perceptual inference in real systems has the potential of using top-down information from arbitrarily distant past, stored in long-term memory. A more sophisticated mathematical model would be needed to capture the contribution of multiple and longer term temporal dependencies.

In the HMM example, we also did not consider sources of top-down information that are distinct from temporal context. Clearly, processes such as inter-modality interactions and spatial contextual integration could also exert a top-down influence on perceptual inference. By comparison with multiple timescales, it would be relatively straightforward to build a more complete mathematical model that captures these other sources of information too.

In this work, we have discussed ACh in the context of perceptual inference in isolation, independent of other processes modulated by ACh. However, ample data at both the systems (e.g. Hasselmo & Bower, 1993) and cellular (e.g. Sillito & Kemp, 1983) levels indicate ACh plays an important role in cortical plasticity and learning (Dimyan & Weinberger, 1999; Hasselmo & Bower, 1993; Kilgard & Merzenich, 1998; Weinberger, 1998). Studies of cholinergic modulation in conditioning (Dayan et al., 2000; Holland, 1997; Holland & Gallagher, 1999) suggest there are strong interactions between learning and inference, both of which are modulated by ACh. As in the work that inspired ours, the association of ACh with uncertainty or unfamiliarity makes it an ideal signal for controlling learning.

We took from the animal conditioning studies that ACh might report uncertainty. However, we have not modeled a particularly interesting aspect of those studies, namely that cholinergic modulation might differentially affect learning for different stimuli that are present simultaneously. If the same applies for different possible contexts, this could significantly enrich the model.

Another important aspect of cholinergic modulation we have omitted is non-BF cholinergic modulation, i.e. mainly subcortical innervation by the pedunculopontine nucleus, the cuneiform nucleus, and the laterodorsal tegmental nucleus. ACh released by these nuclei has been implicated in modulating REM sleep (Jewett & Nortan, 1986; Lavie, Pratt, Scharf, Peled, & Brown, 1984; Velazquez-Moctezuma, Shiromani, & Gillin, 1990) and saccadic eye movement (Aizawa, Kobayashi, Yamamoto, & Isa, 1999), among other processes. It is not yet clear what, if any, similarities or interactions exist in the drive and effects of cortical ACh released by the BF and by the other sources.

A final important aspect of cholinergic modulation that we have not yet addressed is the interaction between ACh and other neuromodulators. For example, there is evidence that dopaminergic afferents from the nucleus accumbens modulates the activity of cholinergic neurons in the BF, and furthermore, that this dopaminergic modulation could be

associated with the cholinergic impairments in schizophrenics (Sarter & Bruno, 1998). It has also been suggested that ACh and norepinephrine play complementary roles in cortical developmental plasticity (Bear & Singer, 1986; Kirkwood, Rozas, Kirkwood, & Perez, 1999).

Acknowledgments

We are very grateful to Mike Hasselmo, David Heeger, Sham Kakade, Szabolcs Káli, Yair Manor, Martin Sarter for helpful comments. Funding was provided by the Gatsby Charitable Foundation and the NSF Graduate Research Fellowship Program.

References

Aizawa, H., Kobayashi, Y., Yamamoto, M., & Isa, T. (1999). Injection of nicotine into the superior colliculus facilitates occurrence of express saccades in monkeys. *Journal of Neurophysiology*, *82*(3), 1642–1646.

Amit, D. J. (1989). *Modelling brain function: The world of attractor neural networks*. Cambridge: Cambridge University Press.

Bear, M. F., & Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature*, *320*(6058), 172–176.

Becker, S. (1999). Implicit learning in 3d object recognition: The importance of temporal context. *Neural Computation*, *11*(2), 347–374.

Bucci, D. J., Holland, P. C., & Gallagher, M. (1998). Removal of cholinergic input to rat posterior parietal cortex disrupts incremental processing of conditioned stimuli. *Journal of Neuroscience*, *18*, 8038–8046.

Carpenter, G., & Grossberg, S. (Eds.), (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.

Cummings, J. L., Gorman, D. G., & Shapira, J. (1993). Physostigmine ameliorates the delusions of Alzheimer's disease. *Biological Psychiatry*, *33*, 536–541.

Dayan, P. (1999). Recurrent sampling models for the helmholtz machine. *Neural Computation*, *11*, 653–677.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, 1218–1223.

Dimyan, M. A., & Weinberger, N. M. (1999). Basal forebrain stimulation induces discriminative receptive field plasticity in the auditory cortex. *American Psychological Association, US*, *113*(4), 691–702.

Donoghue, J. P., & Carroll, K. L. (1987). Cholinergic modulation of sensory responses in rat primary somatic sensory cortex. *Brain Research*, *408*, 367–371.

Everitt, B. J., & Robbins, T. W. (1997). Central cholinergic systems and cognition. *Annual Review in Psychology*, *48*, 649–684.

Fadel, J., Sarter, M., & Bruno, J. P. (2001). Basal forebrain glutamatergic modulation of cortical acetylcholine release. *Synapse*, *39*(3), 201–212.

Fisher, C. M. (1991). Visual hallucinations on eye closure associated with atropine toxicity. *Canadian Journal of Neurological Sciences*, *18*, 18–27.

Gil, Z., Conners, B. W., & Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron*, *19*(3), 679–686.

Grenander, U. (1995). *Elements of pattern theory*. Baltimore, MD: Johns Hopkins University Press.

Hars, B., Maho, C., Edeline, J. M., & Hennevin, E. (1993). Basal forebrain stimulation facilitates tone-evoked responses in the auditory cortex of awake rat. *Neuroscience*, *56*, 61–74.

Hasselmo, M. E. (1995). Neuromodulation and cortical function: Modeling the physiological basis of behavior. *Behavioural Brain Research*, *67*(1), 1–26.

Hasselmo, M. E., & Bower, J. M. (1993). Acetylcholine and memory. *Trends in Neuroscience*, *16*(6), 218–222.

Hasselmo, M. E., & Cekic, M. (1996). Suppression of synaptic transmission may allow combination of associative feedback and self-organizing feedforward connections in the neocortex. *Behavioural Brain Research*, *79*, 153–161.

Helmholtz, H. von. (1896). Visual perception: Essential readings. *Physiological optics*, *3*(26), 1–36.

Hinton, G. E., & Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society of London B*, *352*, 1177–1190.

Holland, P. (1997). Brain mechanisms for changes in processing of conditioned stimuli in pavlovian conditioning: Implications for behavior theory. *Animal Learning and Behavior*, *25*(4), 373–399.

Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences*, *3*(2), 65–74.

Holley, L. A., Turchi, J., Apple, C., & Sarter, M. (1995). Dissociation between the attentional effects of infusions of a benzodiazepine receptor agonist and an inverse agonist into the basal forebrain. *Psychopharmacology*, *120*, 99–108.

Jackson, S. R., Marrocco, R., & Posner, M. I. (1994). Networks of anatomical areas controlling visuospatial attention. *Neural Networks*, *7*(6–7), 925–944.

Jacobs, B. L. (1978). Dreams and hallucinations: A common neurochemical mechanism mediating their phenomenological similarities. *Neuroscience and Behavioral Reviews*, *2*(1), 59–69.

Jewett, R. E., & Nortan, S. (1986). Effect of some stimulant and depressant drugs on sleep cycles of cats. *Experimental Neurology*, *15*, 463–474.

Ka'li, S., & Dayan, P. (2000). The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model. *Journal of Neuroscience*, *20*(19), 7463–7477.

Kilgard, M. P., & Merzenich, M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, *279*(5357), 1714–1718.

Kimura, F., Fukuda, M., & Tsumoto, T. (1999). Acetylcholine suppresses the spread of excitation in the visual cortex revealed by optical recording: Possible differential effect depending on the source of input. *European Journal of Neuroscience*, *11*(10), 3597–3609.

Kirkwood, A., Rozas, C., Kirkwood, J., & Perez, F. (1999). Modulation of long-term synaptic depression in visual cortex by acetylcholine and norepinephrine. *Journal of Neuroscience*, *19*(5), 1599–1609.

Lavie, P., Pratt, H., Scharf, B., Peled, R., & Brown, J. (1984). Localized pontine lesion; nearly total absence of REM sleep. *Neurology*, *34*, 118–120.

Linster, C., & Hasselmo, H. E. (2001). Neuromodulation and the functional dynamics of piriform cortex. *Chemical Senses*, *26*(5), 585–594.

Marder, E. (1998). From biophysics to models of network function. *Annual Review in Neuroscience*, *21*, 25–45.

McGaughy, J., Kaiser, T., & Sarter, M. (1996). Behavioral vigilance following infusions of 192 IgG-saporin into the basal forebrain: Selectivity of the behavioral impairment and relation to cortical AChE-positive fiber density. *Behavioral Neuroscience*, *110*(2), 247–265.

Metherate, R., Asche, J. H., & Weinberger, N. M. (1990). Acetylcholine modifies neuronal acoustic rate-level functions in guinea pig auditory cortex by an action at muscarinic receptors. *Synapse*, *6*, 364–368.

Metherate, R., Asche, J. H., & Weinberger, N. M. (1993). Nucleus basalis stimulation facilitates thalamocortical synaptic transmission in the rat auditory cortex. *Synapse*, 132–143.

Metherate, R., Tremblay, N., & Dykes, R. W. (1987). Acetylcholine permits long-term enhancement of neuronal responsiveness in cat primary somatosensory cortex. *Neuroscience*, *22*(1), 75–81.

Muir, J. L., Dunnett, S. B., Robbins, T. W., & Everitt, B. J. (1992). Attentional functions of the forebrain cholinergic system: Effects of intraventricular hemicholium, physostigmine, basal forebrain lesions

and intra cortical grafts on a multiple-choice serial reaction time task. *Experimental Brain Research*, *89*(3), 611–622.

Muir, J. L., Everitt, B. J., & Robbins, T. W. (1994). AMPA-induced excitotoxic lesions of the basal forebrain: A significant role for the cortical cholinergic system in attentional function. *Journal of Neuroscience*, *14*, 2313–2326.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Croft.

Parasuraman, R., Greenwood, P. M., Haxby, J. V., & Grady, C. L. (1992). Visuospatial attention in dementia of the alzheimer type. *Brain*, *115*, 711–733.

Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.

Perry, E. K., Marshall, E., Thompson, P., McKeith, I. G., Collerton, D., Fairbairn, A. F., Ferrier, I. N., Irving, D., & Perry, R. H. (1993). Monoaminergic activities in Lewy body dementia: Relation to hallucinosis and extrapyramidal features. *Journal of Neural Transmission*, *6*(3), 167–177.

Perry, E. K., & Perry, R. H. (1995). Acetylcholine and hallucinations: Disease-related compared to drug-induced alterations in human consciousness. *Brain and Cognition*, *28*, 240–258.

Pflüger, H. J. (1999). Neuromodulation during motor development and behavior. *Current Opinion in Neurobiology*, *9*, 683–689.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257–286.

Ress, D., Backus, B. T., & Heeger, D. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neuroscience*, *3*(9), 940–945.

Richardson, R. T., & DeLong, M. R. (1991). Electrophysiological studies of the functions of the nucleus basalis in primates. *Advances in Experimental Medical Biology*, *295*, 233–252.

Sarter, M., & Bruno, J. P. (1997). Cognitive functions of cortical acetylcholine: Toward a unifying hypothesis. *Brain Research Reviews*, *23*, 28–46.

Sarter, M., & Bruno, J. P. (1998). Cortical acetylcholine, reality distortion, schizophrenia, and Lewy body dementia: Too much or too little cortical acetylcholine? *Brain and Cognition*, *38*, 297–316.

Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews*, *35*, 146–160.

Schultes, R. E., & Hofmann, A. (1992). Plants of the gods. Rochester, VT: Healing Arts Press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Sillito, A. M., & Kemp, J. A. (1983). Cholinergic modulation of the functional organization of the cat visual cortex. *Brain Research*, *289*, 143–155.

Sillito, A. M., & Murphy, P. C. (1987). The cholinergic modulation of cortical function. In E. G. Jones, & A. Peters (Eds.), *Cerebral cortex*. New York: Plenum Press.

Tremblay, N., Warren, R. A., & Dykes, R. W. (1990). Electrophysiological studies of acetylcholine and the role of the basal forebrain in the somatosensory cortex of the cat. II. Cortical neurons excited by somatic stimuli. *Journal of Neurophysiology*, *64*(4), 1212–1222.

Turchi, J., & Sarter, M. (1997). Cortical acetylcholine and processing capacity: Effects of cortical cholinergic deafferentation on crossmodal divided attention in rats. *Brain Research Cognitive Brain Research*, *6*(2), 147–158.

Turchi, J., & Sarter, M. (2001). Bidirectional modulation of basal forebrain *N*-methyl-D-aspartate receptor function differentially affects visual attention but not visual discrimination performance. *Neuroscience*, *104*(2), 407–417.

Velazquez-Moctezuma, J., Shiromani, P., & Gillin, J. C. (1990). Acetylcholine and acetylcholine receptor subtypes in REM sleep generation. *Progress in Brain Research*, *84*, 407–413.

Weinberger, N. M. (1998). Physiological memory in primary auditory cortex: Characteristics and mechanisms. *Neurobiology of Learning and Memory*, *70*(1–2), 226–251.