Acetylcholine, Norepinephrine, and Spatial Attention

Angela Yu Peter Dayan

Gatsby Computational Neuroscience Unit, UCL, 17 Queen Sq, London WC1N 3AR, UK

1 Introduction

Experimental data on neuromodulators such as norepinephrine (NE), acetylcholine (ACh), and dopamine (DA) show that they are activated quickly and selectively by relevant sensory stimuli, are modulated by internal states, and exert strong effects on a wide variety of neural functions. These results motivate theoretical interpretations in which the neuromodulators play critical and specific roles in neural computations.[17] ACh and NE are of special importance because of their cerebral ubiquity and their evident involvement in controlling the integration of bottom-up and top-down information in the cortex and the course of learning in a variety of tasks. However, although it is known that ACh and NE interact, at least during developmental plasticity,[2] neither experiment nor theory has provided a clear picture of the nature or goals of this interaction. For instance, both neuromodulators are associated with aspects of attention,[10] but there is no apparent rhyme nor reason to the tasks in which pharmacological manipulations of one or the other or, less commonly, both, have an effect.[6, 11, 13]

We have recently developed a theory in which ACh and NE report different aspects of uncertainty, with (cortical) ACh indicating *expected* uncertainty, coming for instance from known variability of the parameters of a task, and NE indicating *unexpected* uncertainty, as when significant aspects of the task are unpredictably changed.[18] This theory is motivated by a wealth of pharmacological, behavioral and physiological data, and should be ideally placed to make predictions about the interaction between the neuromodulators in attentional tasks. Unfortunately, the original version of the theory,[17] which treated ACh by itself, injudiciously jumbled the two different forms of uncertainty, and a later version of the theory,[18] which considered both neuromodulators, was applied to a theoretically convenient, but experimentally awkward task. This inhibited our making predictions that could readily be tested experimentally in standard attentional paradigms.

Posner's quintessential attentional task involves assessing the costs and benefits associated with invalid or valid cueing to a sensory location, at which detection or discrimination must be performed.[9] Many variants of this task have been devised and applied to a wide range of species, including fMRI experiments on humans. Of particular interest for our study is that the effects of manipulating ACh and NE on performance in Posner's task have been investigated, and indeed are qualitatively consistent with our theory. The standard versions of Posner's task involve top-down information of variable quality (the validity of the cues), exactly engaging expected uncertainty. Cue validity is typically left constant throughout an experimental session, and so unexpected uncertainty is not engaged. Bentley (personal communication) has suggested a novel variant of a Posner task which involves both expected and unexpected uncertainty, thus allowing a more complete assessment of the interactions between ACh and NE.

In section 2, we describe the conventional and extended versions of Posner's task, together with a generative model which formally captures the (stochastic) experimental contingencies. In section 3, we describe an approximation to the statistical inverse of this generative model, in which the interaction of ACh and NE plays a critical role. In section 4, we present the results of applying our model to the basic and enriched Posner tasks, and show the expected effects of separate and conjoint manipulations of the neuromodulators. In section 5, we discuss the implications of the model, and relate it more precisely to our previous models.

2 Variations on Posner's Task

In the standard Posner Task, a cue, presented centrally or peirpherally, indicates the likely location of a subsequent visual target stimulus, on which a detection or discrimination task must be performed. Figure 1(a) shows a partic-

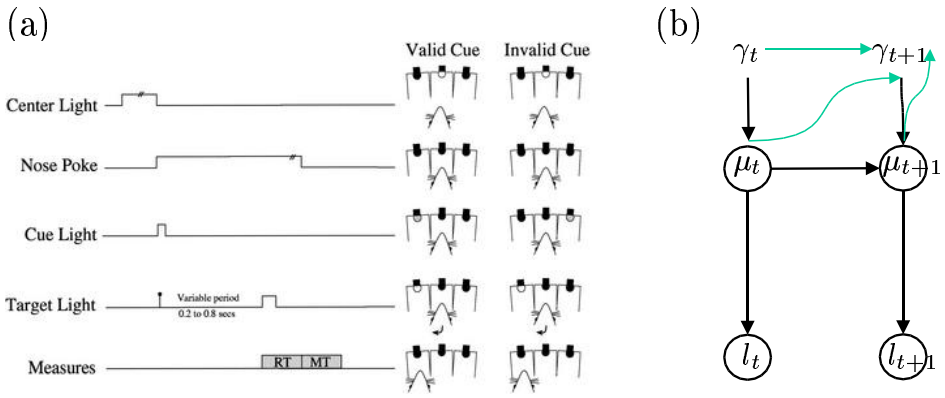

Figure 1: Experimental design. (a) An example of Posner's task adapted to rats, figure from.[15] Trial begins nose poke, terminates with reward retrieval. *MT* movement time. (b) Schematic diagram of the generative model used to generate stimulus patterns in our spatial attention task.

ular example of the task used for rats.[15] Typically, there are cue-induced benefits in reaction time and/or accuracy on correctly cued (valid) trials, and costs on incorrectly cued (invalid) trials. The resulting difference in reaction time (RT(invalid)-RT(valid)) is termed Validity Effect (VE), and is a function of cue validity (probability of stimulus appearing where cued), which is an easily experimentally manipulable form of expected *certainty* in the top-down information. ACh, which reports something like cue *invalidity*, should have effects on VE opposite to those of cue validity, a phenomenon supported by experimental evidence.[3,7,8,14] However, the task contingencies, including cue validity, are kept constant throughout a session, thus allowing unexpected uncertainty no significant role; indeed, NE manipulations have not been observed to affect VE in the standard version of Posner's task.[16]

Instead, we consider a spatial attention task that manipulates both expected and unexpected uncertainty, and whose relationship to Posner's task will become clear shortly. On each trial, the subject must first locate the target stimulus in a visually cluttered scene, and then perform a discrimination task on the target. The location of the target from trial to trial follows a particular pattern, and behavioral measures indicate human subjects can learn such patterns, sometimes implicitly, in the absence of conscious knowledge.[4] The pattern endows the context with the capacity to act as an implicit cue about target locations, and so one can still talk about predictive validity of contextual cues and expected uncertainty, much as in the standard Posner task. However, if the underlying patterns are allowed to change occasionally on a time scale which is on the order of many trials, then that induces a form of *unexpected* uncertainty – as unexpected deviations from the learned pattern accumulate, the subject realizes the need to relearn a new pattern.

The particular form of location patterning we consider is illustrated in Figure 1(b). It consists of the simple stochastic rule that stimuli usually appear in the primary location (valid trials), but can occasionally stray (invalid trials). When the pattern changes, a new location becomes the primary location, associated with a new probability of deviations. More formally, the model for generating stimulus locations can be described as follows: on each trial t, there is a primary location μ_t, taking on the values 1, 2, ... n, where n is the total number of discrete locations where the target stimuli may appear. The probability of the target location l_t being j given $\mu_t = i$ is $\beta_{ij}(\gamma_t) \equiv P(l_t = j|\mu_t = i; \gamma_t)$, where γ_t is the deviation parameter associated with the current primary location. β_{ij} takes on the value $1 - \gamma_t$ for $j = i$, and $\gamma_t/(n-1)$ otherwise. Another part of the model is the transition probability between primary locations, written formally as $\alpha_{ij} \equiv P(\mu_t = j|\mu_{t-1} = i)$. This takes the value α for $j = i$ and $(1-\alpha)/(n-1)$ otherwise. Typically, $\alpha \approx 1$ to reflect the fact that the pattern only changes occasionally. Finally, γ_t is constant throughout all the trials associated with a particular pattern, and is redrawn from a uniform distribution when the primary location changes, i.e. $\gamma_t|\mu_t \neq \mu_{t-1} \sim U[0, \gamma_M]$, where $\gamma_M < 1$ sets the cap value for probability of deviation. Readers familiar with the statistical modelling literature would recognize this generative model as a case of a hidden Markov model with non-stationary output distribution specified by the varying parameter γ.

3 State Inference and Parameter Estimation

"Learning the pattern" in our task requires the subject to maintain an online estimate of the state μ_t and the parameter γ_t from only the observations $\mathbf{l}_t \equiv \{l_1, ..., l_t\}$ (we ignore the fixed parameter α and assume the equilibrium condition when the subject has already learned its value). This is a standard example of Bayesian state inference and parameter estimation. We first derive the exact algorithm and then suggest an approximate model, in which ACh and NE control inference and learning. In the exact model, we compute the joint posterior:

$$\begin{aligned} p(\mu_t = i, \gamma_t | \mathbf{l}_{t-1}, l_t = j) &= p(\mu_t = i, \gamma_t, l_t = j | \mathbf{l}_{t-1})/Z_t \quad \text{where } Z_t \text{ is a normalizing term unique for each } t \\ &\propto \sum_k \int p(\mu_{t-1} = k, \gamma_{t-1} | \mathbf{l}_{t-1}) P(\mu_t = i | \mu_{t-1} = k) p(\gamma_t | \mu_t = i, \mu_{t-1} = k, \gamma_{t-1}) \\ &\quad P(l_t = j | \mu_t = i, \gamma_t) d\gamma_{t-1} \\ &= \beta_{ij}(\gamma_t) \sum_k \alpha_{ki} \int p(\mu_{t-1} = k, \gamma_{t-1} | \mathbf{l}_{t-1}) p(\gamma_t | \mu_t = i, \mu_{t-1} = k, \gamma_{t-1}) d\gamma_{t-1} \\ &= \beta_{ij}(\gamma_t)(\alpha p(\mu_{t-1} = i, \gamma_{t-1} = \gamma_t | \mathbf{l}_{t-1}) + (1 - P(\mu_{t-1} = i | \mathbf{l}_{t-1})) \gamma_M \frac{1-\alpha}{n-1}) \end{aligned} \quad (1)$$

This last expression would seem to provide a convenient iterative algorithm for computing the posterior distribution, except the integration of the posterior over γ_t, a large mixture of beta-distributions growing with nt, is representationally costly and computationally intractable (this operation is required twice in every trial, once in the computation of the marginal posterior of μ_{t-1} in the second term of equation 1, and once in the normalization term Z_t).

In order to learn the underlying patterns in the task proficiently, the brain must implement a tractable *approximate* learning algorithm with reasonable performance. We propose one such approximate algorithm, which also utilizes ACh and NE signalling with the appropriate semantics. The idea is to only keep track of the most likely primary location $\mu_t^* = j$, the estimated posterior $\lambda_t^* \equiv P^*(\mu_t^* | \mathbf{l}_t)$ (assuming $P^*(\mu_t = i \neq j | \mathbf{l}_t) = (1 - \lambda_t^*)/(n-1)$), the estimated most likely deviation parameter γ_t^* corresponding to the current μ_t^*, and h_t^*, the estimated number of trials associated with the current pattern. For the case $l_{t+1} = k = j$, it is easy to show $\mu_{t+1}^* = \mu_t^*$, and

$$\lambda_{t+1}^* \approx \frac{(1 - \gamma_t^*)\lambda_t^*}{(1 - \gamma_t^*)\lambda_t^* + (\gamma_M/2)(1 - \lambda_t^*)/(n-1)} \quad (2)$$

where the numerator gives $P(\mu_{t+1} = j, l_{t+1} = j | \mathbf{l}_t)$, the denominator gives $P(l_{t+1} = j | \mathbf{l}_t)$, and the approximation results from assuming $\alpha \approx 1$. Note λ_{t+1}^* always increases by a nonzero amount from λ_t^*. h_{t+1}^* is updated by incrementing h_t^* by 1, and γ_{t+1}^* can be approximated as a simple running average: $\gamma_{t+1}^* = \gamma_t^* + ((1 - \delta_{jk}) - \gamma_t^*)/h_{t+1}^* = \gamma_t^* - \gamma_t^*/h_{t+1}^*$, where δ_{jk} is the Kronecker delta. If $k \neq j$, the situation is more complicated, as there is a need to differentiate between the possibility of μ having changed from j to k, and the possibility of l_{t+1} simply being an outlier (we ignore the relatively small possibility of μ_{t+1} being neither j nor k). The first scenario corresponds to $P(\mu_{t+1} = k, l_{t+1} = k | \mathbf{l}_t) \approx (1 - \gamma_M/2)(1 - \lambda_t^*)/(n-1)$ and the second corresponds to $P(\mu_{t+1} = j, l_{t+1} = k | \mathbf{l}_t) \approx (1 - \gamma_t^*)\lambda_t^*/(n-1)$. It is obvious that μ_{t+1}^* should be assigned k if and only if the second probability exceeds the first, in which case

$$\lambda_{t+1}^* \approx \frac{\lambda_t^* \gamma_t^*}{\lambda_t^* \gamma_t^* + (1 - \gamma_M/2)(1 - \lambda_t^*)} \quad (3)$$

h_{t+1}^* is still incremented as before, but now γ_{t+1}^* increases: $\gamma_{t+1}^* = \gamma_t^* + ((1 - \delta_{jk}) - \gamma_t^*)/h_{t+1}^* = \gamma_t^* + (1 - \gamma_t^*)/h_{t+1}^*$. If the probability of pattern change dominates, then $\mu_{t+1}^* = k$ and it is easy to show that all the other quantities should be reset as follows: $\lambda_{t+1}^* = 1 - \gamma_M$, $\gamma_{t+1}^* = \gamma_M$, $h_{t+1}^* = 1$.

The computations of this algorithm are simple – we propose they are implemented in the cortex. We further suggest that ACh and NE report γ_t^* and $1 - \lambda_t^*$, respectively. γ_t^* is the expected deviation in a particular context, and therefore appropriate for ACh's role as expected uncertainty. $1 - \lambda_t^*$ is the "doubt" that the current model of location pattern is correct at all. It can be interpreted as a form of unexpected uncertainty since $1 - \lambda_t^*$ is small if many trials of a

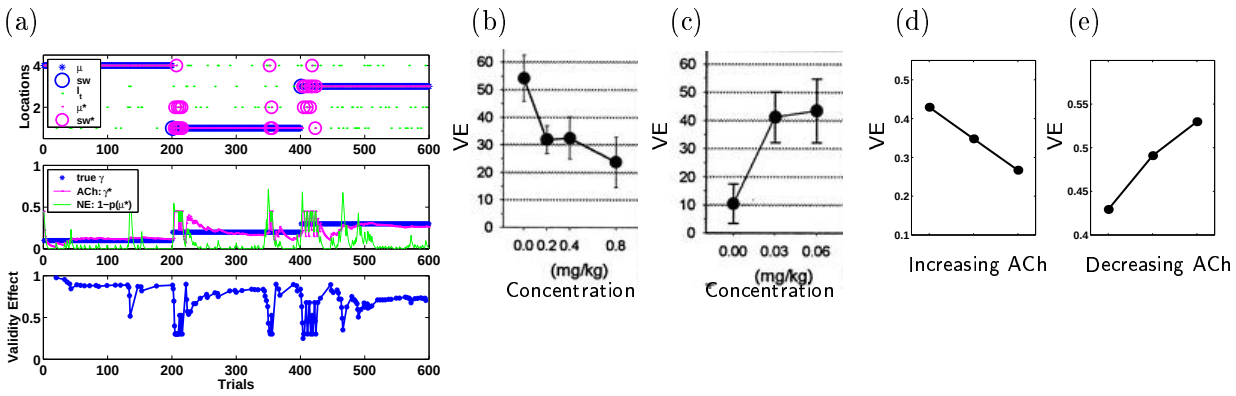

Figure 2: Approximate learning and Posner's task. (a) Sample run of the approximate learning algorithm. SW = switches. Simulation parameters: $n = 4, \alpha = 0.99, \gamma_M = 0.5$. *SW* switches. * estimated quantities. (b) VE interacts with nicotine level dose-dependently, and (c) with scopolamine dose-dependently.[8] (d) and (e) show similar interactions in simulation. 50% validity in both experiment and simulation.

particular pattern have been observed, and only increases to a significant magnitude if many more deviations have been observed than expected. The only other two quantities that need to be represented are μ_t^* and h_t^*, and we suggest that this is part of the role for the prefrontal cortex. That the frontal cortex has dense recurrent connections with both the cholinergic and noradrenergic nuclei makes it the ideal site for the integration of and computation based on these various sources of signals.

This approximate learning algorithm implies that ACh and NE will interact in interesting opponent and synergistic ways. From equation 2, we can see that the decrement in NE, resulting from observing an expected stimulus, varies inversely with ACh. But equation 3 shows that the increase in NE, as a consequence of observing an outlier location, varies proportionally with ACh. These properties correspond to intuitive notions of how expected and unexpected uncertainties should interact.

To link the model with behavior, we consider the behavioral measure validity effect (VE) mentioned in section 2. One natural formalization of VE is to make it proportional to the confidence in the prediction of the next location given all previous observations, or $\text{VE}_{t+1} \propto P(l_{t+1} = \mu_t^* | \mathbf{l}_t) \propto (1 - \gamma_t^*)\lambda_t^*$. Note that VE depends on two factors, the estimated probability of deviation, and the confidence in the current internal model.

Figure 2(a) shows a sample run for a particular setting of the task with 4 stimulus locations. The approximate algorithm does a good job of learning the underlying, varying location patterns from the noisy data: both the estimate of the primary location (panel 1) and that of the deviation parameter (panel 2) quickly converge to the true values. Direct predictions about ACh and NE activities follow from the model traces in panel 2: ACh converges to the true γ, NE rises at pattern changes and also other chance accumulations of unexpected observations. These predictions could be verified via either electrophysiology or dialysis, as the task can be adapted to monkeys or rats. Panel 3 shows the modeled VE for all the "outlier" trials over the course of the simulated session, and illustrates properties that can be verified in behavioral experiments, including in human psychophysics studies. In particular, note the strong dip at the transition between two location patterns, and the inverse relationship between VE and the deviation parameter γ (or the proportional relationship between VE and cue validity).

As a sanity check, we also simulate a special case of our spatial attention task that closely corresponds to the standard Posner task, and compare simulations with experimental data. Numerous studies have demonstrated that cholinergic pharmacological manipulations interact robustly with VE.[3,7,8,14] In figure 2(b-e), we compare simulated cholinergic psychopharmacology to one such study[8] (same behavioral paradigm as[15]), which administered cholinergic agonist nicotine and antagonist scopolamine at various dosages. The correspondence between experiment and theory is good: the monotonic relationships between VE and ACh are clear in both (though the graphs should not be compared liter-

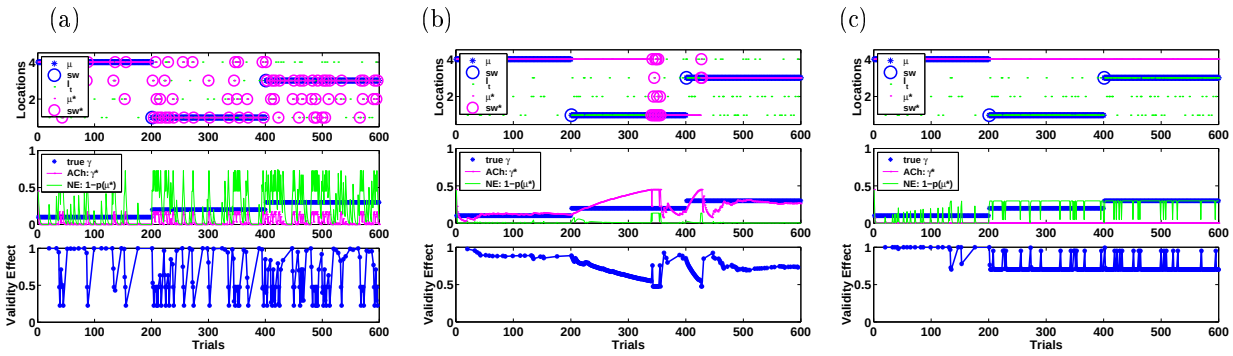

Figure 3: Simulated psychopharmacology. (a) 70% depletion of ACh. (b) 70% depletion of NE. (c) 70% combined depletion of ACh and NE.

ally).

4 Simulated Psychopharmacology

Given the critical interactions between ACh and NE in the proposed spatial attention task, experimental manipulation of ACh and/or NE levels, either through psychopharmacology or lesions of the neuromodulatory nuclei, should result in rich patterns of impairments. Figure 3 shows the results of depleting (a) ACh and (b) NE independently, and (c) jointly on the same example run as before. Suppressing ACh signalling leads to the underestimation of the amount of variation within a given context. Consequently, the significance of deviations from the primary location is exaggerated and causes the NE system to over-react with hyper "distractability." The resulting NE compensation for ACh depletion is incomplete, and is reflected in the decreased and more variable VE of panel 3. NE depeletion corresponds to excessive confidence in the internal model and results in perserverative behavior when there *has* been a pattern change. As a consequence of this reluctance to adapt to new environments, ACh level, or expected uncertainty, rises to take into account of all the accumulating evidence of deviation from the current model. Eventually, the ACh level may hit some threshold and alert the system to a representational change, but this process is rather delayed compared to the case of normal NE functioning. VE levels reflect these deficits: the dramatic drop at the onset of a model switch disappears and is replaced by a slow, gradual downward slide, corresponding to an increasingly more active ACh signal. The most severe impairments come from joint depletion of ACh and NE, which completely prevents the system from learning the underlying patterns. This is reflected in a consistently high and unvarying VE level, which implies that any switch in the underlying location pattern is ignored.

5 Discussion

We have presented a model of inference in a new spatial attention task (Bentley, personal communication). In the task, the context (specified by μ_t) changes slowly, and specifies a probabilistic scheme for each trial. ACh reports the expected variability associated with each context; NE reports the degree of doubt about the correctness of the current estimate of the context. After repeated trials of a particular context, the ACh level converges to a value reflecting the true variability in the context, while the NE level decreases to baseline. Moreover, NE and ACh interact in a precisely specified and partly opponent, partly synergestic manner. We showed the consequences of manipulating them jointly and severally. Suppressing ACh, whilst leaving noradrenergic signalling intact, causes the true variability of each context to be underestimated, thus making every chance deviation from the current context seem more significant than it actually is. This results in a more active NE system, thus making the model exhibit symptoms of distractability,

akin to impairments in Alzheimer's disease patients.[1] Suppressing NE, whilst leaving ACh intact, prevents the system from reacting appropriately when the context does in fact change, leading to perseveration. This in turn leads to an enhanced level of ACh, since the deviations from the perseverating expectations are large. The resulting compensation is only partial. Finally, suppressing both neuromodulators has a devastating impact on performance on this task. Note that the iterative approximate algorithm we have proposed is closely related to the EM algorithm,[12] in which an approximate posterior distribution, derived from the NE signal, and parameter estimation, derived from the ACh signal, are improved alternately while holding the other constant.

One major caveat of this work, and indeed our previous suggestions, is that, although it seems as if expected and unexpected uncertainty are very distinct quantities, it is in fact hard to draw a precise line between them. Uncertainty is only unexpected if it cannot be predicted from a model; but it is always hard to be sure that there is not a more sophisticated model which, fed with the same information, would make more proficient predictions. Such models effectively capture uncertainties about the uncertainties, thereby allowing them to be predicted. Indeed, Bayesian statistical modeling[5] often uses hierarchies of uncertainties about parameters in just such a way. Since even human subjects are not always aware of the degree of sophistication of the models they might implicitly be employing, it is hard to be exactly sure what is expected and what is not. Experimental measures such as reaction times might provide independent evidence, elucidating the levels of inference employed.

This model represents an advance on our previous suggestions. Our first model[17] only treated ACh, and used a task with just a single underlying form of uncertainty. We suggested that ACh reports this uncertainty to control cortical inference (there being no requirement for plasticity); however, there, the particular form of uncertainty is similar to the unexpected uncertainty, reported by NE, in the present model. This is a good illustration of the point raised above about the lack of a completely clean distinction between expected and unexpected uncertainty. The present model is closer to our second model,[18] though there, ACh reports the uncertainty associated with the estimate of a parameter, whereas here, ACh reports the value of a parameter which itself conveys information about uncertainty in the task. In more complicated tasks, both aspects may be important, and so should influence the level of ACh. The degree to which different projections of ACh can report different information to different cortical and sub-cortical locations is as yet experimentally unclear. Similarly, the way that prediction error controls NE levels subtly differs between this and the previous model; again, more experimental evidence is required.

Our model makes qualitative, and even quantitatively precise, predictions about the task of figure 1(b). The task entails the construction and maintenance of hierarchical representations, and behavioral measures such as reaction time may provide revealing insights into the various components during the course of learning. Further, since the task is not substantially more difficult for monkeys or rats than standard versions, it should be possible to measure ACh and NE activity, either electrophysiologically or by dialysis, for comparison with traces such as those in figures 2 and 3. Finally, we have discussed the predictions made by the model about the effects of suppressing (or boosting) the levels of ACh and/or NE. In particular, we expect that suppressing NE and ACh at the same time should have an especially catastrophic effect on behavior.

References

[1] A D Baddeley, H A Baddeley, R S Bucks, and G K Wilcock. Attentional control in Alzheimer's disease. *Brain*, 124:1492–508, 2001.
[2] M F Bear and W Singer. Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature*, 320(6058):172–6, 1986.
[3] A A Chiba, P J Bushnell, W M Oshiro, and M Gallagher. *NeuroReport*, 10:3119–23, 1999.
[4] S A Huettel, P B Mack, and G McCarthy. *Nature Neuroscience*, 5(5):485–90, 2002.
[5] M I Jordan, Z Ghahramani, TS Jaakkola, and L K Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233.
[6] J McGaughy, M Sandstrom, S Ruland, J P Bruno, and M Sarter. *Behav Neurosci*, 111:646–52, 1997.
[7] P M Parasuraman, R Greenwood, J V Haxby, and C L Grady. Visuospatial attention in dementia of the Alzheimer type. *Brain*, 115:711–33, 1992.
[8] J M Phillips, K McAlonan, W G K Robb, and V Brown. *Psychopharmacology*, 150:112–6, 2000.
[9] M I Posner. Orienting of attention. *Q J Exp Psychol*, 32:3–25, 1980.
[10] T W Robbins and B J Everitt. *The Cognitive Neurosciences*, chapter Arousal systems and attention. The MIT Press, 1995.
[11] S J Sara and M Segal. Plasticity of sensory responses of LC neurons in the behaving rat: implications for cognition. *Progr in Brain Research*, 88:571–85, 1991.
[12] R H Shumway and D S Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *J. Time Series Analysis*, 3:253–64, 1982.
[13] J Turchi and M Sarter. *Neuroscience*, 104:407–17, 2001.
[14] M L Voytko, D S Olton, R T Richardson, L K Gorman, J R Tobin, and Price D L. *J Neurosci*, 14:167–86, 1994.
[15] N M Ward and V J Brown. Covert orienting of attention in the rat and the role of striatial dopamine. *J Neurosci*, 16:3082–8, 1996.
[16] E A Witte and R T Marrocco. *Psychopharmacology*, 132:315–23, 1997.
[17] A J Yu and P Dayan. Acetylcholine in cortical inference. *Neural Networks*, 15:719–30, 2002.
[18] A J Yu and P Dayan. Expected and unexpected uncertainty: ACh & NE in the neocortex. *Advances in Neural Processing Systems 15*, 2003.