Uncertainty and Learning

Peter Dayan dayan@gatsby.ucl.ac.uk
Angela J Yu feraina@gatsby.ucl.ac.uk
Gatsby Computational Neuroscience Unit

September 29, 2002

Abstract

It is a commonplace in statistics that uncertainty about parameters drives learning. Indeed one of the most influential models of behavioural learning has uncertainty at its heart. However, many popular theoretical models of learning focus exclusively on error, and ignore uncertainty. Here we review the links between learning and uncertainty from three perspectives: statistical theories such as the Kalman filter, psychological models in which differential attention is paid to stimuli with an effect on the speed of learning associated with those stimuli, and neurobiological data on the influence of the neuromodulators acetylcholine and norepinephrine on learning and inference.

1 Introduction

The field of classical conditioning probes the ways that animals learn predictive relationships in the world, between initially neutral stimuli such as lights and tones, and reinforcers such as food, water or small electric shocks [1, 2]. This sort of learning is obviously a critical adaptive capacity for animals, and there is an enormous wealth of (largely qualitative) experimental data about the speed and nature of learning, a substantial portion of which has been systematised by a handful of theoretical notions and algorithms. The neural basis of classical conditioning has also been explored [3, 4], providing a rich source of data for exploring the implementation of various of the theories. Further, since learning about the correlational structure of the world is exactly the purview of theoretical statistics, it is possible to build statistical computational theories [5], in which probabilistic learning models form the computational level description of learning, and thus provide rational [6] interpretations of the psychological and neural data. Here we review one set of theories of this sort, showing how they tie together experimental data at multiple levels.

One of the most fertile areas of experimentation in conditioning concerns the phenomenon of blocking [7]. The simplest version of this consists of two stages of learning. In the first stage, one stimulus, a light, say, is shown as being a reliable predictor of the occurrence of a reinforcer. Then, in the second stage, a new stimulus, a tone, is presented with the light, followed by the same reinforcer. Experimentally, the second stimulus, despite being in good temporal and spatial contiguity with the reinforcer, does not undergo any (or perhaps as much) learning about the reinforcer. Blocking was a main spur towards two different theoretical ideas about learning, one involving *error correction,* the other involving *uncertainty.* Most theoretical models of learning concentrate on error correction; in the theories reviewed in this paper, uncertainty has equal importance.

In the error-correcting account of conditioning, learning is driven by infelicitous predictions. In the second stage of blocking, there is no prediction error associated with the tone, and therefore there is nothing to learn about the relationship between it and the reinforcer. Rescorla & Wagner [8] described the best known model of this, which was shown [9] to be equivalent to the well-known delta rule from engineering [10], and has formed the basis of much subsequent theoretical work on accounts of both psychological [11] and neural [12–14] phenomena.

Pearce & Hall [15] suggested an uncertainty-based alternative theoretical explanation for blocking, according to which animals allocate enhanced attention to stimuli like lights and tones if there is uncertainty about the predictions associated with those stimuli. Given enhanced attention to a stimulus, an animal would learn about its relationship with any outcome, whether or not that outcome had already been predicted. In blocking, the animal is never given a reason (*eg* through a prediction error) to be uncertain about the predictions associated with the tone, it therefore never allocates it substantial or enhanced attention, and therefore never accords it any learning. Although this might seem a rather contrived explanation for simple blocking, attentional accounts of other conditioning phenomena (such as downwards unblocking [16]) are more natural than their error-correcting equivalents. Furthermore, Holland and his colleagues (see [17, 18]) have collected evidence about the neural substrate of the enhanced attention (often called enhanced *associability*), showing, at least for appetitive conditioning in rats, that it involves a projection from the central nucleus of the amydala to cholinergic neurons (*ie* neurons whose main neurotransmitter is the neuromodulator acetylcholine, ACh) in the basal forebrain that project to the parietal cortex. Furthermore, lesioning all attentional systems reveals an underlying error-correcting learning model. This implies that the two theoretical notions coexist, and may be jointly responsible for the observed phenomena.

Neuromodulators such as ACh and norepinephrine (NE) are neurotransmitters that are believed to affect the ways that *other* neurotransmitters (such as glutamate and GABA) function, and also the ways that the synapses associated with these other neurotransmitters change their strength or efficacy [19]. Because of this capacity, neuromodulators have been described as being involved in metaplasticity [20]. They have extensive connections to cortical and subcortical regions, and are richly interconnected. ACh is particularly interesting with respect to uncertainty and attention, because of evidence (see [21–23]) about its roles in reporting unfamiliarity to control read-in to hippocampal storage of new memories, and in reporting uncertainty in top down information, with the effect of boosting bottom-up, sensory, information in cortex at the expense of top-down expectations. Norepinephrine is also relevant, because of data as to its association with attention and its involvement in reporting unexpected changes in the external environment, such as the appearance of novel objects and changes in learned stimulus-reward contingencies (see [24–26]).

This paper focuses on ways in which uncertainty controls learning. It is organised around versions of a Kalman filter[27], which is a well-known construct in engineering and statistics. In section 2, we review part of the analysis of [28, 29] which links uncertainty in learning to Kalman filtering [30]. In section 3, we extend this to the consideration of joint uncertainty in the associations of multiple stimuli, to accommodate a conditioning paradigm called backwards unblocking [32–34]. In section 4, we show how the model can be further extended to consider the interaction of acetylcholine and norepinephrine in reporting expected and unex-

pected uncertainty [35].

2 Uncertainty, associability and the Kalman filter

Sutton [30, 36] suggested an account of classical conditioning in which animals are trying to 'reverse-engineer' the experimental setting (see figure 1A). This account makes the assumption that the course of the experiment can be described by a statistical model which *parameterises* key features such as the relationship between stimuli and reinforcers. The parameters ($\mathbf{w}(t)$ in the figure) are determined by the experimenter and may change over time: for instance, almost all experiments involve more than one stage with different relationships in each stage. In this account, learning stimulus-reinforcer relationships is just statistical inference about the parameters, typically in a Bayesian (or approximately Bayesian) context. Although this description of the problem might seem strange, it is actually exactly analogous to statistical models of unsupervised learning that are used as accounts of cortical plasticity (see [37]).

In Sutton's model, parameters associated with different stimuli change at different speeds, because of the different past histories of observations associated with the stimuli. As will be apparent, the more uncertain the animal is about a stimulus, the faster it learns about that stimulus. This is exactly the effect that Pearce & Hall [15] described and formalized in terms of different *associabilities* of the stimuli. Holland, Gallagher, and their colleagues [17, 18, 38, 39] have carried out a series of studies on the neural basis of Pearce & Hall's associabilities, implicating the central nucleus of the amygdala, the cholinergic basal forebrain, the hippocampus, and various of their targets. Other information about the neural basis comes from demonstrations such as accelerated or retarded learning in behavioural tasks induced by pharmacological boosting or suppression of ACh [40].

In Sutton's model, the relevant parameters are the *weights* $\mathbf{w}(t) = \{w_i(t)\}$, one for each stimulus, which represent the association between the stimuli and the reinforcer (or perhaps the motivational value of the reinforcer [41, 42]), which is represented by $r(t)$. Here, t is an index over trials (which we sometimes omit below); more sophisticated models must be built to take account of time within each trial [29, 43]. Since experimental circumstances can change (for instance, many conditioning experiments involve more than one stage, with different relationships in each stage), the weights must be allowed to change too. In the simple model, they undergo slow, random evolution, with

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \boldsymbol{\upsilon}(t) \qquad ie \qquad w_i(t+1) = w_i(t) + \upsilon_i(t) \tag{1}$$

where $\boldsymbol{\upsilon}(t)$ has a Gaussian distribution, with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}$. Equation 1 plays the critical role of ensuring enduring uncertainty in $\mathbf{w}(t)$ because of the continual random change. Figure 1B shows an example of the sort of drift in the weights implied by this equation.

The other key aspect of the model is the way that these parameters are assumed to control the delivery of reinforcement. It is natural that only those stimuli present on a trial can influence the reinforcement; however it is experimentally less clear whether the predictions made by different stimuli should be summed (as if the reinforcements they predict are wholly separate) or

averaged (as if they are competing by making different predictions of the same reinforcement [28, 44, 45]). For simplicity (and along with the standard Rescorla-Wagner rule [8]), we assume an additive model, with

$$r(t) = \mathbf{x}(t) \cdot \mathbf{w}(t) + \eta(t) = \sum_i x_i(t) w_i(t) + \eta(t) \tag{2}$$

Here, the presence of stimuli such as lights and tones are represented by binary stimulus variables $x_i(t)$ so that, for instance, $x_1(t) = 1$ if the light is shown on trial t, and $x_1(t) = 0$ if it is not shown. Also $\eta(t)$ is another noise term following a Gaussian distribution with zero mean and variance τ^2.

Sutton [30] suggested that the animals are trying to perform inference about $\mathbf{w}(t)$, based on combined observations of stimuli $\mathbf{x}(t)$ and reinforcements $r(t)$. In terms that are common in unsupervised learning, the model of equations 1 and 2 form a *generative model,* which amount to assumptions about the statistical properties of the environments. The task for animals is to build the inverse of this generative model, called a *recognition model.* For the simple generative model of equations 1 and 2, the Kalman filter is the exact recognition model. First, note the estimate of $\mathbf{w}(t)$ (which we call $\overline{\mathbf{w}}(t)$) has a Gaussian distribution with mean $\hat{\mathbf{w}}(t)$ and a covariance matrix $\Sigma(t)$. For the case of two stimuli (the light, for which $i = 1$, and the tone, for which $i = 2$), $\Sigma(t)$ has the form

$$\Sigma(t) = \begin{pmatrix} \sigma_1^2(t) & \sigma_{12}(t) \\ \sigma_{12}(t) & \sigma_2^2(t) \end{pmatrix}.$$

$\sigma_i^2(t)$ are the individual uncertainties about the associations of the stimuli, and σ_{12} is the correlation between the two. The correlation term σ_{12} is important for explaining some learning phenomena such as backwards blocking (as in the next section); for the present, we take it to be 0. The Kalman filter [27] is a recursive procedure for updating $\hat{\mathbf{w}}(t)$ and $\Sigma(t)$ on the basis of the information provided on a trial. The update for the mean estimate $\hat{\mathbf{w}}(t) = E[\overline{\mathbf{w}}(t)]$ is

$$\hat{\mathbf{w}}(t+1) = \hat{\mathbf{w}}(t) + \frac{\Sigma(t)\mathbf{x}(t)}{\mathbf{x}(t)^T \Sigma(t) \mathbf{x}(t) + \tau^2} (r(t) - \mathbf{x}(t) \cdot \hat{\mathbf{w}}(t)) \tag{3}$$

thus, if both light and tone are shown

$$\widehat{w}_1(t+1) = \widehat{w}_1(t) + \frac{\sigma_1^2(t)}{\sigma_1^2(t) + \sigma_2^2(t) + \tau^2} (r(t) - \mathbf{x}(t) \cdot \hat{\mathbf{w}}(t)) \tag{4}$$

$$\widehat{w}_2(t+1) = \widehat{w}_2(t) + \frac{\sigma_2^2(t)}{\sigma_1^2(t) + \sigma_2^2(t) + \tau^2} (r(t) - \mathbf{x}(t) \cdot \hat{\mathbf{w}}(t)) \tag{5}$$

Although these update equations might look a little complicated, they are actually comprised of a set of readily interpretable pieces. First, $r(t) - \mathbf{x}(t) \cdot \hat{\mathbf{w}}(t)$ is the prediction error on trial t (also known as the *innovation*). This is the difference between the actual reinforcement $r(t)$ and the reinforcement that was predicted based on the best available prior information about $\mathbf{w}(t)$. As mentioned, prediction errors are key to the Rescorla-Wagner learning rule [8]. Temporally more sophisticated versions of such prediction errors are the basis for accounts

of the phasic activation of the dopamine [12, 13] and potentially the serotonin [31] systems during conditioning. Second, $\sigma_1^2(t)/(\sigma_1^2(t)+\sigma_2^2(t)+\tau^2)$ amounts to a competitive allocation of the error signal to the light, where the competition is based on the uncertainties.

This competitive allocation of learning is directly related to the attentional definition of associabilities in conditioning. That is, the conventional Rescorla-Wagner rule [8] would have $\widehat{w}_1(t+1) = \widehat{w}_1(t) + \alpha(r(t) - \mathbf{x}(t)\cdot\widehat{\mathbf{w}}(t))$, where α is a single learning rate (note that learning rates are also known as associabilities). In equation 4, the learning rate is determined competitively according to the uncertainties associated with the stimuli. The larger the uncertainty about the light, *ie* the larger σ_1^2, the faster learning to the light, just as in Pearce & Hall's [15] suggestion. In many conditioning theories, the associability of a stimulus is supposed to be determined by an attentional process controlling access to a limiting resource such as a restricted capacity learning processor. Here, we see that attentional competition is actually the statistically optimal solution to the inference problem of deciding how much responsibility for a prediction error to lay at the door of each stimulus. Finally, the larger the expected variability in the reinforcer delivery (τ^2), the slower the learning. This is because each trial provides less information about the relationship between stimulus and reinforcer.

The Kalman filter also specifies a recursive equation for updating the covariance matrix of uncertainties $\Sigma(t)$, according to which uncertainty grows because of the possible changes to $\mathbf{w}(t)$ that are mandated by equation 1, but shrinks because of the observations in each trial. The initial value $\Sigma(1)$ should be set to reflect the degree of the animal's initial ignorance about the relationship (something which itself might be adaptive, if, for instance, it experiences one family of relationships over many repeated blocks). One particular characteristic of the simple linear and Gaussian form is that the changes to $\Sigma(t)$ do not depend on the actual errors that are made, but only on the sequence of stimuli presented. We consider a more reasonable model in section 4.

Figure 1C;D show two examples of inference using the Kalman filter: one case of blocking (C), and one of upwards unblocking (D), in which blocking does not occur because the reinforcement is increased when the second stimulus, the tone, is added. The solid lines show the course of learning to the light; the dashed lines show that to the tone. In unblocking, that the tone has higher uncertainty than the light implies that it is allocated most of the prediction error.

The Kalman filter recognition model thus offers a statistical computational account of the conditioning data. Behavioural neuroscience data [17, 18] on the anatomical substrate of associabilities suggest something about the implementation of the uncertainties, notably that they are reported cholinergically. There is not a great deal of evidence about the synaptic representation of the uncertainties $\sigma_1^2(t), \sigma_2^2(t)$; and certainly nothing to support the suggestion [28] of the involvement of the nucleus accumbens in the competition between different stimuli to take responsibility for any prediction error.

These behavioural neuroscience data also contain one main surprise for the model, in suggesting that the enhancement of associabilities to stimuli whose consequences are uncertain (*ie* have high σ_i^2) depends on *different* neural structures from that of the decrease of associabili-

ties to stimuli whose consequences are certain. There is no apparent correspondence for this in the model, unless two different types of uncertainty actually underlie the two conditioning paradigms. One possibility is that the experiments in which associability decreases induce drastic representational changes [46], whereas experiments involving incremental associability only involve parametric changes within a representation.

3 Joint uncertainty and backwards blocking

In the simple version of the Kalman filter model presented in the previous section, we made the approximation that the diagonal terms of the covariance matrix are zero (for the case of just two stimuli, $\sigma_{12}(t) = 0$). This means that there is no form of joint uncertainty in the relationships of the light and the tone. In this section we consider a key learning paradigm that forces this assumption to be removed.

Since the estimate $\overline{\mathbf{w}}(t)$ of $\mathbf{w}(t)$ has a Gaussian distribution, joint uncertainty does not affect the mean prediction. However, it can affect learning. One important example of this is the case of backwards blocking. This is a variant of the blocking paradigm we discussed above, in which the two stages are reversed. That is, in the first stage, the light and tone are shown together in association with the reinforcer; in second stage, the light by itself is shown with the reinforcer. The result after the second stage is similar to regular blocking, namely that the association between the tone and reinforcer is significantly weaker than that of the light, and also significantly weaker than it was at the end of the first stage of learning. What is unsettling about backwards blocking is that it seems as if the association of the tone decreases during the second stage of learning, even though it is not presented. Although somewhat harder to demonstrate experimentally than conventional blocking, both humans (*eg* [32]) and animals (*eg* [33]) exhibit the phenomenon.

The Kalman filter provides a parsimonious explanation for backwards blocking [34]. During the first stage of learning, both light and tone are provided, followed by a unit amount of reinforcer (say $r(t) = 1$). Then, according to equation 2, the trials only provide information about the sum $\overline{w}_1 + \overline{w}_2$, and not anything about $\overline{w}_1$ and $\overline{w}_2$ independently. Therefore, although the mean values $\widehat{w}_1 = \widehat{w}_2 = 0.5$ share the value of the reinforcer equally, the uncertainties are *anti-correlated,* that is $\sigma_{12} < 0$. Another way of putting this is that if $\overline{w}_1$ is actually a bit bigger than its mean 0.5, then since information from the first stage of trials constrains the sum $\overline{w}_1 + \overline{w}_2$, $\overline{w}_2$ has to be a bit smaller than its mean 0.5. Figure 2A-C represent this graphically. These two-dimensional plots show the joint distributions of the estimates of w_1 and w_2 at the start (A) and end (B) of the first stage of learning. Figure 2B shows that $\overline{w}_1 + \overline{w}_2$ is arranged near the value 1, with the slope of the ellipse indicating the anticorrelation.

The second stage of learning, in which the light alone is shown in conjunction with the same $r(t) = 1$, duly provides information that $\overline{w}_1$ is larger than 0.5, and thus the mean value $\widehat{w}_1$ gets larger and the mean value $\widehat{w}_2$ gets smaller. Figure 2C shows the joint distribution at the end of this stage; the mean values $\widehat{w}_1$ and $\widehat{w}_2$ have clearly moved in the direction indicated.

This behaviour of the means arises from the full learning equation 3, because the change to $\widehat{w}_2$ is proportional to $(r - \widehat{w}_1)\sigma_{12}$, which is negative.

The Kalman filter thus shows how joint uncertainty between multiple different potentially predictive stimuli can account for backwards blocking. Given the apparent involvement of the cholinergic basal forebrain in representing or implementing the diagonal terms of the covariance matrix σ_i^2, it is natural to wonder about their involvement in the off-diagonal elements such as σ_{12}, which can be either positive or negative. There is no direct neural evidence about this. Dayan & Kakade [34] suggested a simple network model that remaps the stimulus representation $\mathbf{x}(t)$ into a quantity that is approximately proportional to $\Sigma(t)\mathbf{x}(t)$, as required for equation 3, and showed how it indeed exhibits backwards blocking.

4 Unexpected uncertainty and norepinephrine

One way to think about the role suggested for ACh in the Kalman filter model is that it reports expected uncertainty [35], since the uncertainty Σ about $\overline{w}_1$ translates exactly into uncertainty about the delivery of the reinforcer r when the light is presented, and so the expected amount of error. We have already discussed how expected uncertainty has a major effect in determining the amount of learning. Intuitively (and formally, at least in a slightly more complex model [22, 23]) it should also have a significant effect on inference, if different sources of information (notably bottom-up information from sensation and top-down information based on prior expectations from general and specific knowledge) are differentially uncertain. ACh is also known to have an effect like this on cortical inference, boosting bottom-up connections (*eg* of thalamo-cortical fibres into primary somatosensory cortex [47] via a nicotinic mechanism) and suppressing top-down and recurrent information (via a muscarinic mechanism [48–50]).

An obvious counterpart to expected uncertainty is unexpected uncertainty, arising in cases in which the error is much greater than it should be based on Σ, for instance as the experimenter changes from one stage of an experiment to the next. As we mentioned above, in the simple Kalman filter model, changes to Σ do not depend on the observed error, and so unexpected uncertainty has no apparent direct impact. However, in the case of experimental change, it should have a major effect on both inference and learning: suppressing potentially outdated top-down expectations from influencing inference, and encouraging faster learning about a potentially new environment [35]. Indeed, there is evidence that a different neuromodulatory system, involving cortical NE, plays an important role in reporting various types of changes in the external world that can be interpreted as involving unexpected uncertainty (see [51]). Together, then, expected and unexpected uncertainty should interact appropriately to regulate learning and inference.

One advantage of a Bayesian model such as the Kalman filter is that all the assumptions are laid bare for refutation. In our case [23], a main problem is that the relationship between stimuli and reinforcer is modelled as changing slowly via diffusion (equation 1), whereas in reality there are times, notably the switches between different stages of the trials, at which the relationship can change drastically. These dramatic changes are formally signalled by

prediction errors $\left(r(t) - \mathbf{x}(t) \cdot \hat{\mathbf{w}}(t)\right)^2$ that are much greater than expected on the basis of $\Sigma(t)$. The consequence of a dramatic change is that $\Sigma(t)$ should be reset to something like its initial value.

One way to formalize this is to change the evolution of equation 1 to

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \boldsymbol{\upsilon}(t) + c(t)\boldsymbol{\phi}(t) \tag{6}$$

where $c(t) \in \{0, 1\}$ is a binary variable signalling a dramatic change, and $\boldsymbol{\phi}(t)$ is another Gaussian noise term, but with a large variance (compared to $\sigma^2\mathbf{I}$). Setting $c(t) = 1$ is like declaring that the relationship $\mathbf{w}(t)$ between stimulus and reinforcer may have changed dramatically between times t and $t+1$. The probability p_c that $c(t) = 1$ is set to be very small, to capture the relative rarity of dramatic shifts in $\mathbf{w}$. Figure 3A shows an example in which there is just a scalar $\mathbf{w}(t)$ that changes according to the dual noise processes of equation 6.

Unfortunately, statistically exact estimation of $\mathbf{w}$ in this new system is significantly complicated by the addition of the binary noise element. This is because of the need to assign at every time step a probability to whether or not $\mathbf{w}(t)$ has shifted ($c(t) = 1$ or 0). This causes the hidden space to explode exponentially. In statistical terms, the distribution of the estimate $\overline{\mathbf{w}}$ is not a single Gaussian as in the regular Kalman filter, but a mixture of Gaussians, with a number of components that grows exponentialy with the number of observations. This makes exact inference impractical.

We have proposed a simple and tractable approximate learning algorithm, in which ACh and NE signal expected and unexpected uncertainty, respectively [35]. As before, we assume the animal keeps a running estimate of the current weight in the form of a mean $\hat{\mathbf{w}}(t)$ and a variance $\Sigma(t)$ (both quantities are scalar in this case). Here, the variance is again supposed to be reported through ACh. Based on these quantities, a measure of normalized prediction error can be obtained from quantities related to those in equation 3

$$\beta(t) = (r(t) - x(t) \cdot \hat{\mathbf{w}}(t))^2 / (\mathbf{x}(t)^T \Sigma(t) \mathbf{x}(t) + \tau^2)$$

We propose that NE signalling is based on $\beta(t)$, which is effectively a Z-score. If $\beta(t)$ exceeds a threshold γ, then we assume $\hat{c}(t) = 1$, and reset $\Sigma(t)$ accordingly large to take into account the large variance of $\boldsymbol{\phi}$. If $\beta(t) < \gamma$, then we assume $\hat{c}(t) = 0$ and update $\hat{\mathbf{w}}$ and Σ as in the basic Kalman filter of section 2. In general, we would predict a higher level of ACh when the NE system is active.

Figure 3B shows the performance of this approximate scheme on the example sequence of Figure 3A, with the corresponding evolution of ACh and NE signals in Figure 3C. $\hat{\mathbf{w}}$ closely tracks the actual $\mathbf{w}$, despite the considerable noise in r. Whenever $\beta(t) > \gamma$, ACh level shoots up to allow faster learning, and drops back down again as in the basic Kalman filter. In general, the performance of the approximate scheme depends on the choice of γ, as can be seen in Figure 3D. Optimal value of γ is slightly larger than 3 for this particular setting of parameters. When γ is very small, top-down uncertainty is constantly high relative to noise in the observations, making estimates heavily dependent on r and resulting in errors similar to a scheme where no internal representation of past observations are kept at all (top line). A very large

γ, on the other hand, corresponds to a very conservative view toward dramatic changes in $\mathbf{w}$, resulting in very slow adaptation when dramatic changes do occur, hence the large errors for large values of γ.

Our example illustrates that a dual-modulatory system, involving both ACh and NE, can competently learn about hidden relationships in the world that both drift slowly at a fine timescale and change more dramatically on occasions. Interference with the normal signalling of ACh and NE, as one might expect, would cause malfunctioning in this system. We have shown that such malfunctioning [35], induced by various disabling manipulations of ACh and NE in simulations, correspond to similar impairments in animals under pharmacological manipulation in comparable experimental paradigms. For example, simulated NE-depletion causes severe disruptions in the learning of drastic shifts in $\mathbf{w}$ but not the slow drifts [35], just as NE-lesioned animals are impaired in learning changes in reinforcement contingencies [51, 56] but not in performing previously learned discrimnation taks [57]. Simulated cholinergic depletion in our model, on the other hand, leads to a particular inability to integrate differences in top-down and bottom-up information, causing abrupt switching between the two. This tendency is particularly severe when the new $\mathbf{w}(t)$ is similar to the previous one, which can be thought of as a form of interference. Hippocampal cholinergic deafferentation in animals also bring about a stronger susceptibility to interference compared with controls [21].

5 Discussion

We have considered three aspects of the way that uncertainty should control aspects of learning in conditioning and other tasks. Expected uncertainty comes as the variance of parameter values such as those governing the relationship between a stimulus and a reinforcer. Such variances are automatic consequences of probabilistic inference models in which initial ignorance and continuing change battle against observations from experiments. The consequence of expected uncertainty is generally comparatively *faster* learning – given two stimuli, one whose relationships are well established from past experience (*ie* have low variance) and another whose are not; clearly the latter should dominate in having its parameters be changed to take account of any prediction errors. This theory complements those of the way that two of the other major neuromodulators (dopamine [12, 13] and serotonin [31]) interact to specify the prediction error component for learning.

We also considered joint uncertainty in the form of correlations and anti-correlations in the unknown values of multiple parameters. This neatly accounts for backwards blocking, an experimental paradigm of some theoretical importance, as it shows apparent revision in the estimated association of a stimulus even when the stimulus is not present.

Finally, we considered the interaction between expected and unexpected uncertainty, the latter coming from radical changes in parameter values. These induce large changes in the uncertainty of the parameters, and so, unlike the earlier, simpler models, causally couple error to uncertainty. That is, unexpected prediction errors in a trial lead to increased expected prediction errors in subsequent trials.

These suggestions arise from a Kalman filter, which is a fairly standard statistical inference model. However, they are not merely of theoretical interest, but are tied directly to psychological data on the associabilities of stimuli, *ie* the readiness of stimuli to enter into learned associations, and other sophisticated learning phenomena such as backwards blocking. They are also tied directly to neural data on the involvement of cholinergic and noradrenergic systems in modulating learning in cases of expected and unexpected surprise.

Of course, the model is significantly simplified in all respects too. Even in our studies, it was only consideration of radical change in experimental circumstance that led to the slightly more complex model of section 4, which makes explicit the link to NE. Further, the effects of ACh and NE on target tissue (for instance in modulating experience-dependent plasticity, depending on the behavioural relevance [19]), and also on each other, are many and confusing; our simple account cannot be the last word. It is also important to address neurophysiological data from Aston-Jones and colleagues on differential tonic and phasic activities of NE neurons depending on factors such as attentional demand and, presumably, level of vigilance [24, 25]. Finally, there is a wealth of behavioural experiments on the effects of forms of uncertainty; it has not been possible to consider them in anything other than a very broad-brush form. Of particular concern is that the account in section 4 is based on just one single model, whose parameters are more or less uncertain. One might expect an animal to maintain a whole quiver of models, and have uncertainty both within each model, as to its current parameters, and between models, as to which applies.

Various aspects of this account deserve comment. We have concentrated on the effects of uncertainty on learning. Of equal importance are the effects on inference. This has been studied in two related frameworks. One is the substantial work of Hasselmo and his colleagues [21] in the context of recurrent associative memory models of cortex and hippocampus. The suggestion is that ACh and other neuromodulators report on unfamiliarity of input patterns to affect learning and inference. It is obvious that learning should be occasioned by unfamiliarity; it is less obvious that inference should be affected too. Hasselmo argued that associative interactions are damaging in the face of unfamiliarity, since the information available through recurrence is known to be useless. Thus, neuromodulators should turn down the strength of recurrent input compared with bottom-up input, at the same time as plasticising it. Unfamiliarity is a close cousin of uncertainty, and our theory of cortical ACh [22, 23] borrowed this same key idea, but implemented it in a hierarchical model of cortical information processing, which is more general than an associative memory. In this model, as in the electrophysiological data, ACh comparatively boosts bottom-up input to the cortex (via nicotinic receptors [47]) and suppresses recurrent and top-down input (via muscarinic receptors [48]). There is quite some evidence for this hypothesis, both direct and indirect, which is reviewed in [23]. The inferential role of NE is less clear. In the model of section 4, it would be possible for NE to have a role only indirectly via its effects on ACh. Contrary to this, there is evidence that NE and ACh have very similar effects on cortical processing in promoting stimulus-evoked activities at the expense of top-down and recurrent processing [53, 54], as well as some less well-understood mutual interactions. The computation and use of uncertainty here is quite closely related to certain of the signals in Carpenter & Grossberg's adaptive resonance theory [55].

A different way for variance to affect inference arises in competitive models of prediction. We

mentioned above that the additive model of the Rescorla-Wagner rule, in which the predictions made by different stimuli are summed, is merely one possibility. In other circumstances, and indeed to build a quantitative model of a wealth of conditioning data [29, 43], it is important to consider how the predictions might compete. One natural idea is that they compete according to a measure of their reliabilities; this can also be formalised in probabilistic terms [28, 44].

One important issue about which there is little experimental data is the instantiation of the computation leading to the putative ACh and NE uncertainty signals. This computation happens quickly; for instance, as with dopamine cells [4], NE cells change their firing rates within 100ms of the presentation of a stimulus, in addition to showing some rather longer lasting tonic effects [51, 52]. The amygdala and basal forebrain pathway investigated by Holland *et al* [17, 18] does not have an obvious noradrenergic counterpart. An equally pressing concern is the degree of stimulus-specificity in the neuromodulatory control over learning. Although there could be some selectivity in the nature of the connections, particularly for ACh, it is not clear how the delivery of ACh could be specific enough to boost learning to one stimulus but suppress it to another which is simultaneously presented. Unfortunately, the critical experiment to test whether this occurs in an ACh-sensitive manner, has yet to be performed.

Finally, it is interesting to consider experimental directions for testing and elaborating the theory. Issues of specificity and joint uncertainty are most acute; it would be most interesting to repeat the experiments on the role of the cholinergic system in enhanced uncertainty but including the possibility for stimuli to compete with each other in the final test phase of learning. The first issue is whether the different stimuli do indeed learn at different rates, and whether this is controlled by ACh. If so, then it would be interesting to use more selective lesions to try to ascertain the substrate of the specificity. Another issue is the one we mentioned about the difference in the neural substrate of incremental and decremental attention. One natural explanation under our model is that two different forms of uncertainty are masquerading as being the same; it should then be possible to design an experiment under which the neural system identified as responsible for increasing attention is demonstrably important also for decreasing attention.

Acknowledgements

We are very grateful to Sham Kakade with whom many of these ideas have been developed, and to Andrea Chiba, Chris Córdova, Nathaniel Daw, Tony Dickinson, Zoubin Ghahramani, Geoff Hinton, Theresa Long, Read Montague and David Shanks for extensive discussion over an extended period. Funding was from the Gatsby Charitable Foundation and the National Science Foundation.

References

[1] A Dickinson, Contemporary Animal Learning Theory, Cambridge: Cambridge University Press, 1980.

[2] N J Mackintosh, Conditioning and Associative Learning, Oxford: Oxford University Press, 1983.

[3] M A Gluck, E S Reifsnider, & R F Thompson, Adaptive signal processing and the cerebellum: Models of classical conditioning and VOR adaptation. In MA Gluck, & DE Rumelhart, eds., Neuroscience and Connectionist Theory. Developments in Connectionist Theory, 131–185. Hillsdale, NJ: Erlbaum, 1990.

[4] W Schultz, Predictive reward signal of dopamine neurons. Journal of Neurophysiology, vol 80, pp 1-27, 1998.

[5] D Marr, Vision. New York, NY: WH Freeman, 1982.

[6] J R Anderson. The Adaptive Character of Thought, Hillsdale, NJ: Erlbaum, 1990.

[7] L J Kamin, Predictability, surprise, attention, and conditioning. In: Punishment and Aversive Behavior, edited by Campbell BA, and Church RM. New York: Appleton-Century-Crofts, pp 242-259, 1969.

[8] R A Rescorla & A R Wagner, A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In AH Black & WF Prokasy, eds., Classical Conditioning II: Current Research and Theory, pp 64-69, New York: Appleton-Century-Crofts, 1972.

[9] R S Sutton & A G Barto, Toward a modern theory of adaptive networks: Expectation and prediction. Psychological Review, vol 88, pp 135-170, 1981.

[10] B Widrow & M E Hoff, Adaptive switching circuits. WESCON Convention Report vol 4, pp 96-104, 1960.

[11] R S Sutton & A G Barto, Time-derivative models of Pavlovian conditioning. In M Gabriel, & JW Moore, eds., Learning and Computational Neuroscience, pp 497-537. Cambridge, MA: MIT Press, 1990.

[12] P R Montague, P Dayan, & T J Sejnowski, A framework for mesencephalic dopamine systems based on predictive Hebbian learning. Journal of Neuroscience, vol 16, pp 1936-1947, 1996.

[13] Schultz, W, Dayan, P & Montague, PR, A neural substrate of prediction and reward. Science, vol 275, pp 1593-1599, 1997

[14] W Schultz & A Dickinson, Neuronal coding of prediction errors. Annual Review of Neuroscience, vol 23, pp 473-500, 2000.

[15] J M Pearce & G Hall, A model for Pavlovian learning: Variation in the effectiveness of conditioned but not unconditioned stimuli, Psychological Review, vol 87, pp 532-552, 1980..

[16] P C Holland, Journal of Experimental Psychology: Animal Behavior Processes, vol 14, pp 261-279, 1988.

[17] P C Holland, Brain mechanisms for changes in processing of conditioned stimuli in Pavlovian conditioning: Implications for behavior theory. Animal Learning & Behavior, vol 25, pp 373-399, 1997.

[18] P C Holland & M Gallagher, Amygdala circuitry in attentional and representational processes, Trends in Cognitive Sciences, vol 3, pp 65-73, 1999.

[19] Q Gu, Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. Neuroscience, vol 111, pp 815-835, 2002.

[20] K Doya, Metalearning and neuromodulation, Neural Networks, vol 15, pp 495-506, 2002.

[21] M E Hasselmo. Neuromodulation and cortical function: Modeling the physiological basis of behavior. Behavioural Brain Research, vol 67, pp 1-27, 1995.

[22] P Dayan & A J Yu, ACh, uncertainty, and cortical inference. In TG Dietterich, S Becker & Z Ghahramani, editors, NIPS 2001, 2002.

[23] A J Yu & P Dayan, P, Acetylcholine in cortical inference. Neural Networks, vol 15, pp 719-7302, 2002.

[24] J Rajkowski, P Kubiak,& G Aston-Jones, Locus Coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. Brain Res Bull, vol 35, pp 607-16, 1994.

[25] G Aston-Jones, J Rajkowski, & J Cohen, Role of locus coeruleus in attention and behavioral flexibility, Biol Psychiatry, vol 46, pp 1309-1320, 1999.

[26] S J Sara, Learning by neurons: role of attention, reinforcement and behavior, Comptes Rendus de l'Academie des Sciences Serie III-Sciences de la Vie-Life Sciences, vol 321, pp 193-198, 1998.

[27] B D Anderson & J B Moore. Optimal Filtering., Prentice-Hall, Englewood Cliffs, NJ, 1979.

[28] P Dayan, S Kakade, & P R Montague, Learning and selective attention. Nature Neuroscience, vol 3, pp 1218-1223, 2000.

[29] S Kakade & P Dayan, Acquisition and extinction in autoshaping. Psychological Review, vol 109, pp 533-544, 2002.

[30] R S Sutton, Gain adaptation beats least squares? In Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems, 1992.

[31] N D Daw, S Kakade, & P Dayan, Opponent interactions between serotonin and dopamine. Neural Networks, vol 15, pp 603-616, 2002.

[32] D R Shanks, Forward and backward blocking in human contingency judgement. Quarterly Journal of Experimental Psychology: Comparative & Physiological Psychology, vol 37, pp 1-21, 1985.

[33] R R Miller & H Matute, Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgement, Journal of Experimental Psychology: General, vol 125, pp370-386, 1996.

[34] P Dayan & S Kakade, Explaining away in weight space. In TK Leen, TG Dietterich & V Tresp, editors, NIPS 2000, pp 451-457, 2002.

[35] A Yu & P Dayan, Expected and unexpected uncertainty: ACh and NE in the neocortex, Accepted by NIPS, 2003.

[36] M Gluck, P Glauthier, & R S Sutton, Adaptation of cue-specific learning rates in network models of human category learning, Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, pp 540-545, Erlbaum, 1992.

[37] G E Hinton & T J Sejnowski, Unsupervised Learning: Foundations of Neural Computation, MIT Press, Cambridge, MA, 1999.

[38] D J Bucci, P C Holland, & M Gallagher, Removal of cholinergic input to rat posterior parietal cortex disrupts incremental processing of conditioned stimuli, Journal of Neuroscience, vol 18, no 19, pp 8038-8046, October 1998.

[39] J-S Han, P C Holland,& M Gallagher, Disconnection of the amygdala central nucleus and substantia innominata/nucleus basalis disrupts increments in conditioned stimulus processing in rats. Behavioral Neuroscience, vol 113, pp 143-151, 1999.

[40] D D Rasmusson DD, The role of acetylcholine in cortical synaptic plasticity. Behav Brain Res, vol 115, pp 205-18, November 2000.

[41] A Dickinson & B M Balleine, The role of learning i n motivation. In CR Gallistel (Ed) Learning, Motivation & Emotion, Volume 3 of Steven's Handbook of Experimental Psychology, Third Edition, New York: John Wiley & Sons, 2002.

[42] P Dayan & B M Balleine, BW, Reward, motivation and reinforcement learning. Submitted to Neuron, 2002.

[43] C R Gallistel & J Gibbon, Time, Rate, and Conditioning. Psychological Review, vol 107, pp 289-344, 2000.

[44] P Dayan & T Long, Statistical models of conditioning. In MI Jordan, M Kearns & SA Solla, editors, Advances in Neural Information Processing Systems, 10, Cambridge, MA: MIT Press, 117-123, 1998.

[45] J K Kruschke, Toward a unified model of attention in associative learning, Journal of Mathematical Psychology, vol 45, pp 812-863, 2001.

[46] M Gluck & C Myers, Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning, Cambridge, MA: MIT Press, 2000.

[47] Z Gil, B M Conners, & Y Amitai, Differential regulation of neocortical synapses by neuromodulators and activity. Neuron, vol 19, pp 679-86, 1997.

[48] F Kimura, M Fukuada, & T Tsumoto, 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, vol 11, pp 3597-3609, 1999.

[49] C Y Hsieh, S J Cruikshank, & R Metherate, Differential modulation of auditory thalamocortical and intracortical synaptic transmission by cholinergic agonist, Brain Research, vol 880, pp 51-64, 2000.

[50] M E Hasselmo & M Cekic, Suppression of synaptic transmission may allow combination of associative feedback and self-organizing feedforward connections in the neocortex, Behav Brain Res, vol 79, pp 153-61, 1996.

[51] S J Sara, A Vankov, & A Herve, Locus coeruleus-evoked responses in behaving rats: a clue to the role of noradrenaline in memory. Brain Res Bull, vol 35, pp 457-65, 1994.

[52] A Vankov, A Herve-Minvielle, & S J Sara, Response to novelty and its rapid habituation in locus coeruleus neurons of freely exploring rat. Eur J Neurosci, vol 109, pp 903-911, 1995.

[53] M E Hasselmo, C Linster, M Patil, D Ma, & M Cekic, Noradrenergic suppression of synaptic transmission may influence cortical signal-to-noise ratio, J. Neurophysiol, vol 77, pp 3326-3339, 1997.

[54] M Kobayashi, K Imamura, T Suai, N Onoda, M Yamamoto, S Komai, & Y Watanabe, Selective suppression of horizontal propagation in rat visual cortex by norepinephrine, European Journal of Neuroscience, vol 12, pp 264-272, 2000.

[55] G A Carpenter & S Grossberg, editors, Pattern Recognition by Self-Organizing Neural Networks, Cambridge, MA: MIT Press, 1991.

[56] T W Robbins, Cortical noradrenaline, attention and arousal, Psychological Medicine, vol 14, pp 13-21, 1984.

[57] J McGaughy & M Sarter, Lack of effects of lesions of the dorsal noradrenergic bundle on behavioral vigilance, Behav Neurosci, vol 111, no 3, pp 646-52, 1997.

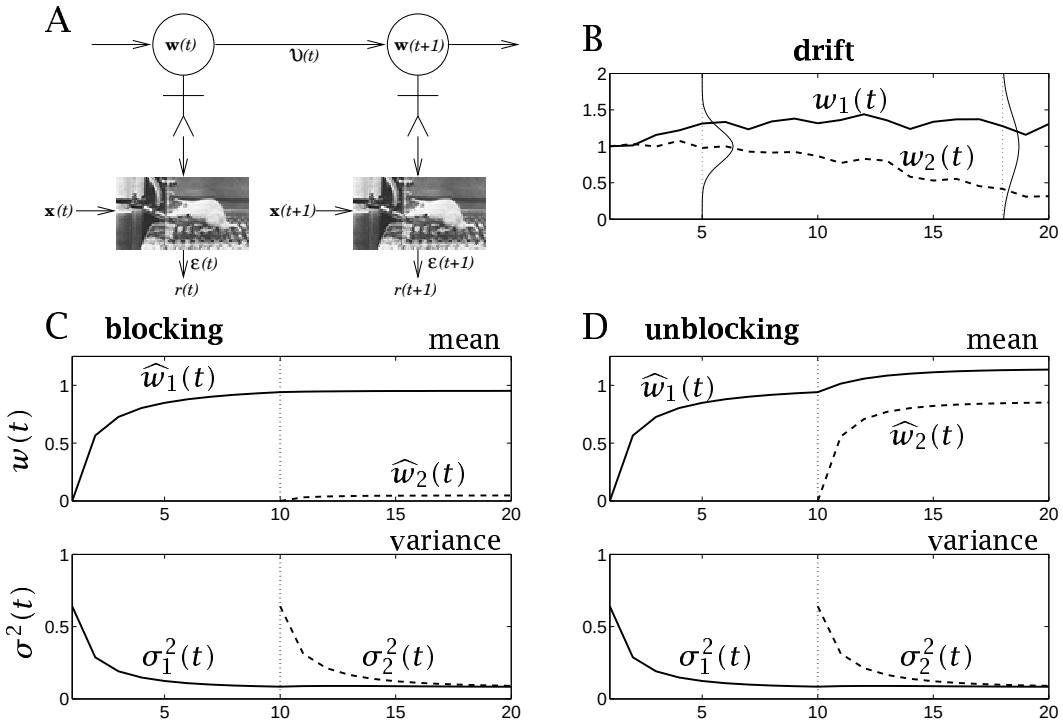

Figure 1: Acquisition and the Kalman filter. A) Conceptual scheme. The experimenter (stick figures) sets the parameters $\mathbf{w}(t)$ governing the relationship between stimulus $\mathbf{x}(t)$ and reward $r(t)$ (the pictures of the Skinner box) subject to noise $\epsilon(t)$. Parameters can change over time according to a diffusion process $\boldsymbol{\upsilon}(t)$. The animal has to track the changing relationship based on observations. B) Example of the change to the weights over time steps. Here, the variance of $\upsilon_i(t)$ is 0.1. The thin lines show the distribution of weights at times $t = 5$ and $t = 18$ showing the effect of growing uncertainty. C) Blocking. Here learning to the tone (dashed line) in the second half of the timesteps is blocked by the prior learning to the light (solid line). The upper plot shows the means of the estimated weights $\widehat{\mathbf{w}}(t)$; the lower plot shows the variances of the weights. Since the tone is unexpected until the second stage, its variance is set to the same initial value as the light on trial 10 when it is introduced (vertical dotted line). D) Upwards unblocking. The same as (C) except that the reward is doubled when the tone is introduced. This unblocks learning to the tone. Since there is more uncertainty associated with the tone than the light at the time that it is introduced, it takes responsibility for the bulk of the prediction error, and so ends up with a mean association $\widehat{w}_2$ almost as large as that of the light.

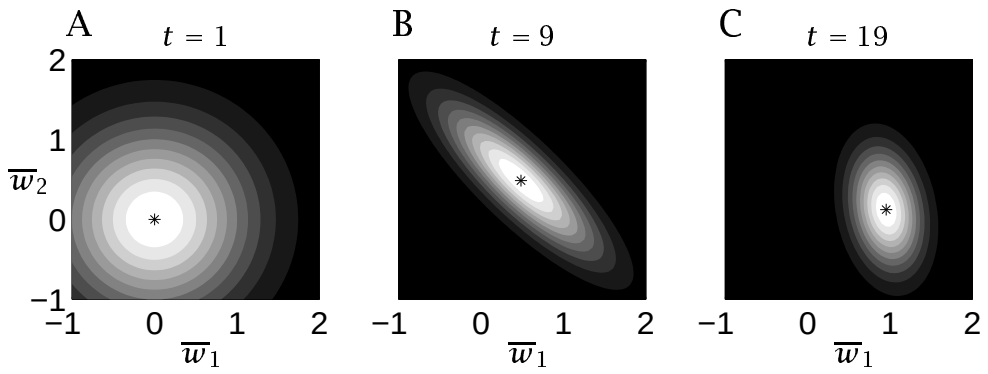

Figure 2: Backwards blocking. The three figures show contour plots of the joint distribution of $\overline{\mathbf{w}}(t)$ (with the mean value $\hat{\mathbf{w}}(t)$ indicated by the asterisk) at three stages during a backwards blocking experiment, the start (A); at the end of the first stage (B); and at the end of the second stage (C). (B) shows the anticorrelation between $\overline{w}_1$ (for the light) and $\overline{w}_2$ (for the tone) induced by the first stage of learning ($\sigma_{12} = -0.34$) which explains the basic phenomenon. (C) shows that during the second stage of learning, the mean values $\widehat{w}_1 \to 1, \widehat{w}_2 \to 0$; and the uncertainty about $\overline{w}_2$ grows (hence the vertical orientation of the ellipse). After [34].

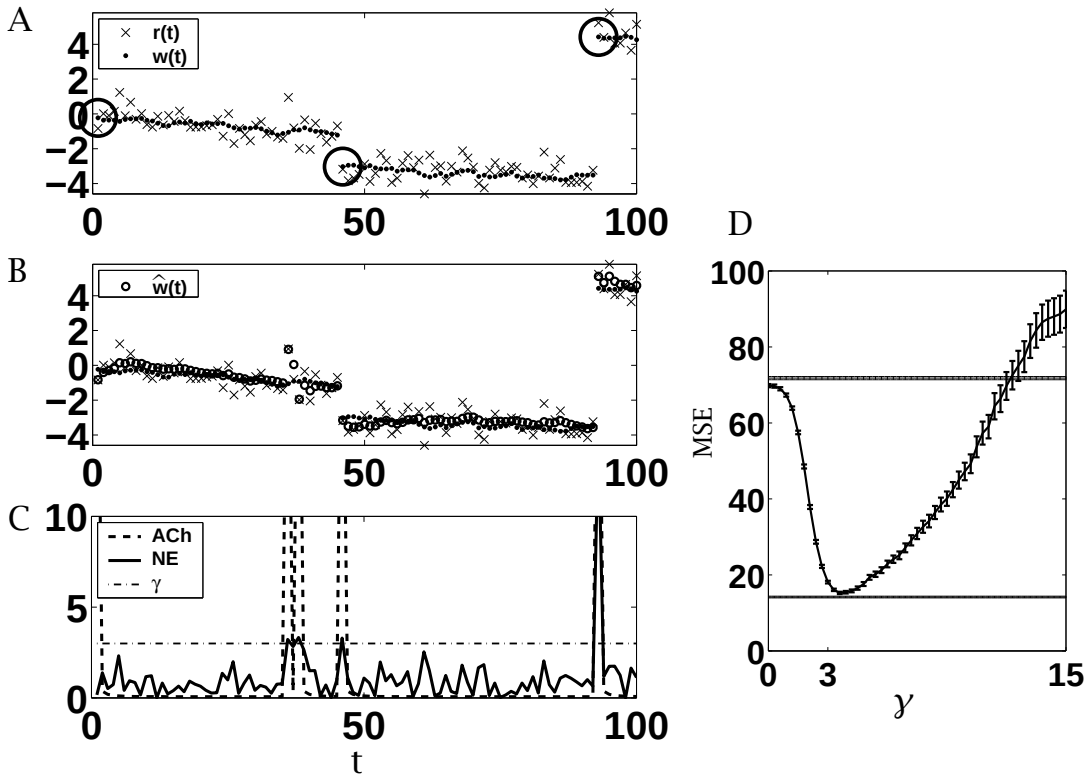

Figure 3: Cholinergic and noradrinergic learning in an adaptive factor analysis model. (A) Sample sequence of $\mathbf{w}(t)$ ('.'), exhibiting both slow drift and occasional drastic shifts, and a corresponding sequence of noisy $r(t)$ ('x'). (B) The mean estimate $\hat{\mathbf{w}}$ from the ACh-NE approximate learning scheme closely tracks the actual $\mathbf{w}$. (C) Corresponding levels of ACh (dashed) and NE (solid) for the same sequence. (D) MSE of the approximate scheme as a function of the threshold γ. Optimal γ for this particular setting of parameters is 3.30, for which performance approaches that of the exact algorithm (lower line). When γ is too large or too small, performance is as bad as, or worse than, if $\mathbf{w}(t)$ is directly estimated from $r(t)$ without taking previous observation into account (top line).