

A particle filtering account of selective attention during learning

Angela Radulescu & Yael Niv & Nathaniel D. Daw (angelar,yael,ndaw@princeton.edu)

Department of Psychology and Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA

Abstract

A growing literature has highlighted a role for selective attention in shaping representation learning of relevant task features, yet little is known about how humans learn what to attend to. Here we model the dynamics of selective attention as a memory-augmented particle filter. In a task where participants had to learn from trial and error which of nine features is more predictive of reward, we show that trial-by-trial attention to features measured with eye-tracking is better fit by the particle filter, compared to a reinforcement learning mechanism that had been proposed in the past. This is because inference based on a single particle captures the sparse allocation and rapid switching of attention better than incremental error-driven updates. However, because a single particle maintains insufficient information about past events to switch hypotheses as efficiently as do participants, we show that the data are best fit by the filter augmented with a memory buffer for recent observations. This proposal suggests a new role for memory in enabling tractable, resource-efficient approximations to normative inference.

Keywords: reinforcement learning; decision-making; selective attention; particle filtering

Introduction

Several recent studies have highlighted a role for selective attention in shaping learning under uncertainty (Niv et al., 2015; Marković et al., 2015; Mack et al., 2016; Leong et al., 2017), but left open the question of how attention changes over time. Here we suggest that selective attention during human reinforcement learning arises from sequential sampling of hypotheses about which features of a task are relevant. We formalize this process as a memory-augmented particle filter (Doucet & Johansen, 2009; Bonawitz et al., 2014; Speekenbrink, 2016). Particle filters offer a tractable approximation to rational inference and in the case of only a few particles, resemble sequential hypothesis testing (Wilson & Niv, 2011). The key idea of a particle filter is to represent the target probability distribution using a finite number of point estimates, or particles. The ensemble of particles is dynamic: estimates that are inconsistent with recent evidence are filtered out. Over time, the ensemble comes to better approximate the target distribution. In general, the quality of the approximation increases both with time, and with the number of particles. Here we show that a single-particle model does well in capturing the dynamics of human attention allocation, due to the sparsity of the representation and the model's ability to rapidly switch hy-

potheses about the identity of the reward-predictive feature. But such sparsity is in tension with the main normative appeal of particle filters, which is to treat the ensemble of particles as approximating the exact posterior at each step.

One way to compensate for using fewer particles is through the choice of proposal distribution for re-sampling particles. In general, the closer the match between the proposal distribution and the target posterior, the better the approximation to the posterior will be (Speekenbrink, 2016). In our task, the proposal distribution defines an implicit switching rule for staying with the current hypothesis about which feature is more predictive of reward, or switching to a different one. One suggestion in a similar setting was to use the exact posterior as the proposal distribution, effectively endowing the model with the ability to switch hypotheses in proportion to the true posterior probability (Bonawitz et al., 2014). But this is unrealistic as a process-level model, since it relies on access to the very distribution the particle filter is attempting to approximate.

Here we replace this assumption with a novel memory mechanism that modifies the proposal distribution to incorporate a set of the most recent observations. This modification both solves the efficiency problem associated with single-particle models and highlights a new role for memory in enabling approximate inference. We develop a method for fitting memory-augmented particle filters to trial-by-trial eye-tracking data, and compare the particle filter to a previous reinforcement learning account of selective attention in a multidimensional learning task. We find that the memory-augmented particle filter more closely matches the trial-by-trial dynamics of attention allocation, suggesting a role for memory in guiding attention to task relevant features.

Experimental paradigm

We analyzed data from a multidimensional learning task in which human participants were tasked with learning from trial and error which of nine features was most predictive of reward (Figure 1). On each trial, participants had 2 seconds to select one of 3 columns, each including a face, a house, and a tool. Choosing the column containing the target feature (e.g. Einstein) yielded a reward with 0.75 probability. Choosing any of the other two columns was rewarded with only 0.25 probability. All features were visible on every trial, with feature combinations within columns determined randomly on each trial. We defined each block of 20 trials as a 'game' during which the target feature stayed constant. The target feature randomly changed between games, and this was announced to participants. Participants were instructed about the reward contingencies before the experiment.



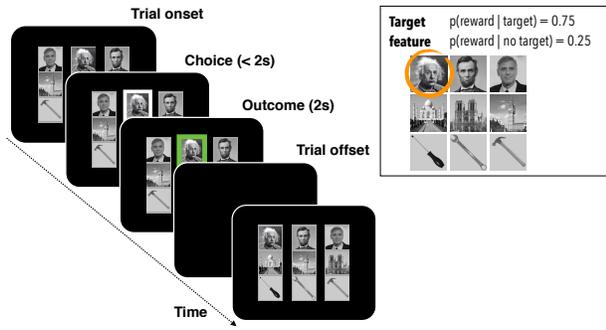


Figure 1: **Task.** Participants had to learn from trial and error which of 9 features was more predictive of reward.

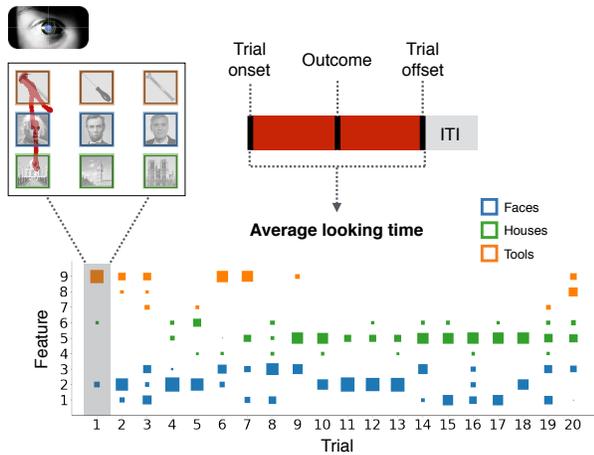


Figure 2: **Sample attention dynamics for one game of the task.** Inset shows the looking patterns corresponding to the first trial of the game. Within each trial, looking time was binned and averaged across the spatially-resolved features.

In previous work, we demonstrated the viability of using eye-tracking to measure trial-by-trial changes in attention to different dimensions of the task (Faces, Houses or Tools; (Leong et al., 2017)). Here we extend this approach by employing high-frequency eye-tracking to derive a trial-by-trial measure of feature-level attention (Figure 2).

Selective attention as particle filtering

We consider an environment in which multidimensional stimuli vary along D dimensions (e.g. Faces). Each dimension can take on F features per dimension (e.g. Einstein). Of the $K = D \times F$ possible features f , target feature f^* yields reward with higher probability than others: $p(r_t | f_t^*) > p(r_t | \neg f_t^*)$, and in particular, $p(r_t | \neg f_t^*) = 1 - p(r_t | f_t^*)$. On each trial, the target feature changes to a random target with probability h .

Under the generative model defined above, we model participants as sequentially approximating the belief state $p(f = f^*)$

using particle filtering (Figure 3). Instead of maintaining and updating the full posterior distribution over f (Niv et al., 2015), we assume that they keep track of a single particle H_t that represents their current hypothesis about the identity of the target feature. Participants have access to observations of the form $O_t = \{C_t, R_t\}$ – the choice and reward experienced on each trial, which they can use to update their hypothesis.

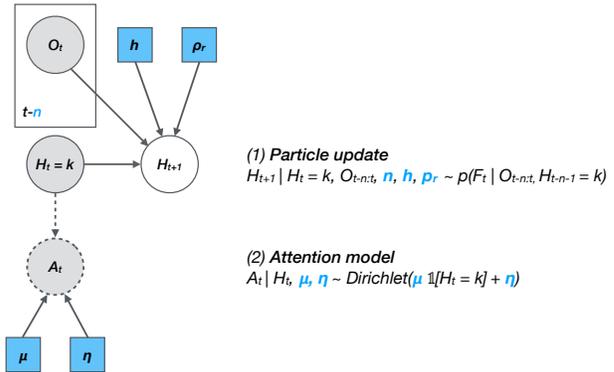


Figure 3: **Graphical model of the particle filter.** Nodes in white represent latent hypotheses. Nodes in gray represent given quantities. Blue squares denote fixed parameters of the model. Solid and dotted lines distinguish between observations experienced by the participant during the task (i.e. choices and rewards), and data that the experimenter gets to observe (A reflects attention to the different features, measured empirically as looking times). $p(F_t)$ is the proposal distribution, which depends on the current observation, the previous n observations from memory, and fixed parameters h and p_r .

Particle filters learn by sampling new particles from a proposal distribution and filtering out particles that are inconsistent with new evidence. Over time, particles settle on hypotheses that in ensemble approximate the true posterior. Previous work has shown that individual behavior in classic associative learning experiments supports a model with a single particle (Daw & Courville, 2008). But particle filters use the ensemble of particles as a stand-in for the complete history of the process (i.e., to approximate the posterior distribution at each step), and a single particle is impoverished in that it does not maintain enough history to evolve correctly over time. We thus consider a class of memory-augmented particle filters in which the proposal distribution at each time point is given by the probability of being in each state, conditional on the current state of the particle and the n most recent observations (Figure 3, top equation). Locally computing this distribution can be accomplished with a simple recursion (Ghahramani, 2001), since the dynamics of the particle correspond to the latent state in a hidden Markov model with the transition matrix defined by h , and the emission probabilities defined by $p(r | f^*)$.

The free parameter n can be interpreted as the memory capacity of the model. This relaxes the assumption that the proposal process has access to the full posterior (Bonawitz et al., 2014), and approximates it relying on a limited memory buffer (which we associate with working or episodic memory) and on the particle as a summary of the distribution prior to that. Note that a similar proposal distribution could be achieved with less computation by only consulting the memory to propose where to switch following a reward omission (Bonawitz et al., 2014).

Finally, fitting the model to empirical data requires a link function between the dynamics of the latent hypothesis and the trial-by-trial measure of attention. Since our measure is proportion looking time, we use the Dirichlet distribution, a generalization of the Beta distribution over N-dimensional compositional vectors (i.e. vectors of proportions that sum to 1) (Figure 3, bottom equation). This likelihood function places μ probability mass on the current hypothesis, and assumes a fixed noise level $\epsilon > 1$.

Alternative model: feature reinforcement learning with decay

In previous work, we introduced Feature Reinforcement Learning with decay (FRLdecay) as a candidate mechanism for learning what to attend to in this task. The full model is described in (Niv et al., 2015), but in brief, FRLdecay assumes the participant learns a feature weight W_f for each of the nine features. The predicted value for the chosen stimulus is the sum of its feature weights. After each observation, the weights of the chosen features are updated according to the difference between the obtained reward and the predicted value (a ‘prediction error’) multiplied by learning rate η . Weights of unchosen features decay toward zero in proportion to decay rate d . We previously modeled dimensional attention as a softmax over maximum feature weights in each dimension (Leong et al., 2017). Here we model attention to each feature as a softmax over the vector of learned feature weights W , where the inverse temperature β dictates how focused attention is on features with larger weights:

$$\phi_t(k) = \frac{e^{\beta W_t(k)}}{\sum_i e^{\beta W_t(i)}} \quad (1)$$

As with the particle filter, we assign likelihood to the attention data by using a Dirichlet distribution with parameters determined by the predicted feature-level attention.

$$p(A_t) \sim \text{Dirichlet}(\mu \times \phi_t + \epsilon) \quad (2)$$

This likelihood function assigns probability mass out of a fixed μ in proportion to the attention to each feature, and assumes a fixed noise level $\epsilon > 1$.

Fitting procedure

Maximum likelihood estimation (MLE) requires computing the likelihood of a data sequence $\mathcal{D}_{1:T}$ under a set of parameters θ . While MLE is standard in reinforcement learning (Wilson & Collins, 2019), evaluating the likelihood of data under models with stochastic latent states is typically intractable because the state space of possible trajectories grows exponentially with the number of trials (c.f. (Findling et al., 2018)). In our case, since we cannot directly observe what hypothesis the participant was considering at each time point, we need to marginalize our uncertainty over $H_{1:T}$. This again can be accomplished efficiently for the current model using the forward algorithm for inference in hidden Markov models.

To optimize the parameters of the particle filter, we maximized the likelihood via grid search across h and p_r bounded between 0.001 – 1 in increments of 0.02, given values of $n \in \{0, 1, 2, 3, 4, 5, 6\}$. For the FRLdecay model, we computed the log likelihood across η and η_k bounded between 0.001 – 1 in increments of 0.02, given values of $\beta \in \{5, 10, 20, 30, 40, 50\}$. For both models, μ and ϵ were fixed to 4 and 1.1, respectively. Maximizing across the different likelihoods thus yielded estimates that are adequately matched with respect to the number of parameters.

Results and discussion

We first investigated whether the particle filter can reliably recover the structure of the task (Figure 4). We simulated the choice behavior of particle filter agents with different memory capacities on the same stimulus sequence as human participants were exposed to. For illustration, we fixed h at 0.001 and p_r at 0.99, to compensate for the assumption that h does not exactly match the generative dynamics of the task (i.e. there are no unsignaled changes in the target feature). We generated the agent’s choices using a greedy choice rule (i.e. the model always chooses the stimulus containing the current hypothesis). The performance of the model increased with memory capacity, and approached human performance for the 5-back condition both in terms of speed of learning (Figure 4 top) and accuracy on the last 6 trials of a game (Figure 4 bottom).

We then compared the performance of the particle filter and FRLdecay models in predicting trial-by-trial fluctuations in selective attention (Figure 5). We found that the particle filter outperforms FRLdecay for every participant, suggesting that shifts in attention are more consistent with hypothesis testing than with gradual error-driven learning (Figure 5A). We also found significant variability in the estimated memory capacity of the particle filter (Figure 5B).

Taken together, these results support the idea that approximate inference over task-relevant features guides selective attention during trial and error learning. We propose a new mechanism by which memory of recent experiences in-

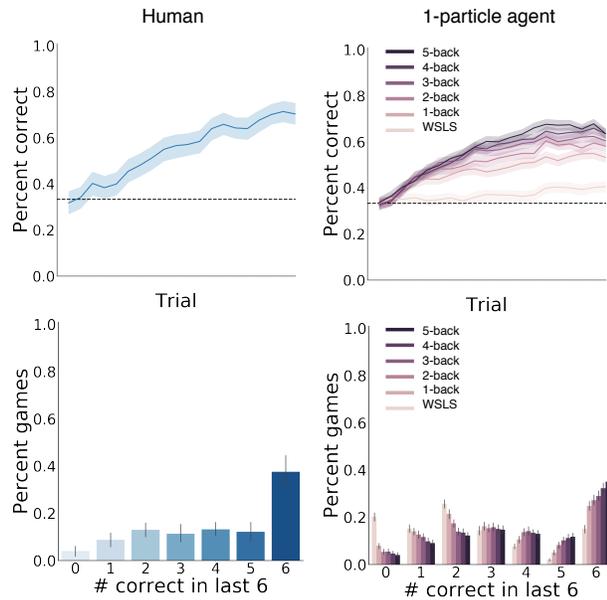


Figure 4: Model performance. Top left: learning curves for 21 participants. Performance was assessed as the proportion of trials in which the participant chose the stimulus containing the target feature. Top right: learning curves for a simulated particle filter agent with different memory capacities. The Win-Stay-Lost-Shift (WLSL) agent only has access to the current sensory observation in determining the next hypothesis. Bottom left: histogram of the average number of correct choices participants made in the last 6 trials of each game. Games in which they made the correct choice in 6 of the last 6 trials can be considered “learned”. Bottom right: average number of correct choices in the last 6 trials of a game by the particle filter model, as a function of memory capacity.

forms this inference, enabling efficient switching to hypotheses that are most consistent with recent evidence. Future work will address whether this model also explains trial-by-trial choices.

Acknowledgments

This work was supported in part by grant W911NF-14-1-0101 from the Army Research Office and by grant R01DA042065 from the National Institute on Drug Abuse. We are thankful to Ian Ballard for helpful discussions.

References

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference. *Cognitive psychology*, 74, 35–65.

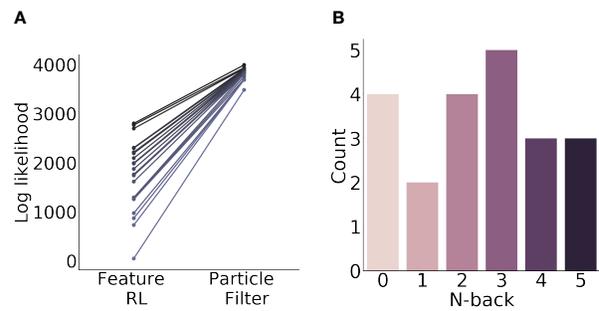


Figure 5: Model comparison. A: Feature reinforcement learning vs. particle filter. Each line represents an individual subject. B: Histogram of inferred MLE memory capacity (n) values. 17 out of 21 participants were better fit by a model which in addition to the current observation, has access to memory of the recent trials.

Daw, N. D., & Courville, A. (2008). The pigeon as particle filter. *Advances in neural information processing systems*, 20, 369–376.

Doucet, A., & Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704), 3.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2018). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *BioRxiv*, 439885.

Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. In *Hidden markov models: applications in computer vision* (pp. 9–41). World Scientific.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, 93(2), 451–463.

Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proceedings of the National Academy of Sciences*, 113(46), 13203–13208.

Marković, D., Gläscher, J., Bossaerts, P., O’Doherty, J., & Kiebel, S. J. (2015). Modeling the evolution of beliefs using an attentional focus mechanism. *PLoS computational biology*, 11(10), e1004558.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, 35(21), 8145–8157.

Speekenbrink, M. (2016). A tutorial on particle filters. *Journal of Mathematical Psychology*, 73, 140–152.

Wilson, R. C., & Collins, A. (2019). Ten simple rules for the computational modeling of behavioral data.

Wilson, R. C., & Niv, Y. (2011). Inferring relevance in a changing world. *Frontiers in Human Neuroscience*, 5.