Contrasting the effects of prospective attention and retrospective decay in representation learning

Guy Davidson
College of Computational Sciences
Minerva Schools at KGI
San Francisco, CA 94103
guy@minerva.kgi.edu

Angela Radulescu
Department of Psychology
Princeton University
Princeton, NJ 08544
angelar@princeton.edu

Yael Niv
Department of Psychology &
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08544
yael@princeton.edu

Abstract

Previous work has shown that cognitive models incorporating passive decay of the values of unchosen features explained choice data from a human representation learning task better than competing models [1]. More recently, models that assume attention-weighted reinforcement learning were shown to predict the data equally well on average [2]. We investigate whether the two models, which suggest different mechanisms for implementing representation learning, explain the same aspect of the data, or different, complementary aspects. We show that combining the two models improves the overall average fit, suggesting that these two mechanisms explain separate components of variance in participant choices. Employing a trial-by-trial analysis of differences in choice likelihood, we show that each model helps explain different trials depending on the progress a participant has made in learning the task. We find that attention-weighted learning predicts choice substantially better in trials immediately following the point at which the participant has successfully learned the task, while passive decay better accounts for choices in trials further into the future relative to the point of learning. We discuss this finding in the context of a transition at the "point of learning" between explore and exploit modes, which the decay model fails to identify, while the attention-weighted model successfully captures despite not explicitly modeling it.

Keywords: reinforcement learning, representation learning, decay, selective attention, behavioral modeling, Dimensions Task, model comparison, trial-by-trial analysis

Acknowledgements

This project was funded by grant W911NF-14-1-0101 from the Army Research Office to YN. The authors wish to thank the Princeton Neuroscience Institute Summer Internship Program that facilitated this collaboration.

1 Introduction

Previous work on representation learning in humans has investigated a role for selective attention in dynamically shaping task representations [1], [2]. Computational models of this process of carving a task into its constituent states have suggested two different mechanisms for implementing selective attention to different features of environmental stimuli: a feature-level reinforcement learning model with decay of values of features of unchosen options (FRLdecay), and an attention-weighted feature-level reinforcement learning model (awFRL). These two models offer conceptually different accounts of selective attention: decay is retrospective, incrementally forgetting previously learned values if options are not chosen again. In contrast, attentional filtering can be seen as prospective, modulating what is learned now for future use, in line with [3], [4].

Even though these models posit different mechanisms, previous work has shown that, on average, they perform equally well at predicting participants' trial-by-trial choices on a multidimensional bandit task called the "Dimensions Task" (see below). Here, we asked which model better accounts for choices on the Dimensions Task as a function of the participant's stage of learning. If the two models map onto the same cognitive process, we would expect them to be indistinguishable in terms of the likelihood they assign to choices at every stage of the learning process. Conversely, if they capture different aspects of the cognitive process, they should account well for different choices. First, we implement a new model, awFRLdecay, which includes both a decay mechanism and attentional weights, and find it predicts participants' choices better than either FRLdecay or awFRL. This suggests that the two mechanisms capture different components of representation learning. We further investigate this finding using a novel trial-by-trial model comparison analysis which takes into account the participant's "point of learning" when comparing likelihoods. We find that the awFRL model best explains behavior around the time when participants learn the correct task representation, while the FRLdecay model best captures choices further beyond the "point of learning." These findings suggest distinct roles for passive decay of feature weights and selective attention in shaping task representations, and demonstrate the utility of moving beyond average likelihoods when performing model comparison.

2 Task

We reanalyzed data from [2]. Twenty-five human participants were tasked with learning which of nine features was more predictive of reward. Participants played 24 "games" consisting of 25 trials each. On each trial of a game, participants chose between three columns, each comprised of a face, a landmark, and a tool. All features were visible on every trial, but feature combinations within a column varied from trial to trial. The target feature randomly changed between games. Participants had full prior knowledge of the task's generative model. That is, they received instructions in the beginning regarding the reward contingencies, and changes in the target feature (i.e., end of games) were signalled explicitly ("New game starting").

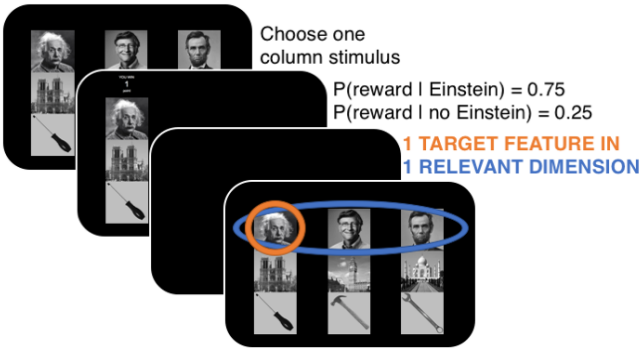

Figure 1: Dimensions Task. On each trial, the participant is presented with three options (columns of features), each including a face, a landmark, and a tool ("dimensions"). After choosing one option, the participant receives feedback, and proceeds to the next trial. One dimension is relevant for determining reward, and within it, one feature rewards with $p = 0.75$ while the other two features reward with $p = 0.25$. The participant does not know *a priori* which dimension is relevant for reward, and which feature is the target feature, and must learn these from trial and error

3 Models Compared

We compared models using leave-one-game-out cross-validation, maximizing the total log-likelihood of the participants' choices. As each participant performed a total of twenty-four games of the task, we held one game out at a time, fit the model on the remaining twenty-three games, and evaluated performance on the held-out game.

3.1 Feature RL with Decay (FRLdecay)

The FRLdecay model introduced in [1] assumes the participant learns a feature weight for each of the nine unique features. We denote these feature weights w and sum to compute a value for each column. Choice likelihood is modeled as a noisy softmax. During learning, the weights of features in the chosen column are updated according to a temporal-difference learning rule. Additionally, the values for features in unchosen columns decay toward zero. The addition of decay implements a retrospective form of selective attention: on any given trial, the model learns about all chosen features equally, later unlearning (decaying) the values of those features that are not chosen again. That is, instead of predicting what should be learned on each trial, the model learns and then forgets values that were learned erroneously.

We denote the ith column on the current trial as S_i, with the feature in each dimension (row) $d \in \{1, 2, 3\}$ accessed as $S_i[d]$, which indexes into the weights vector, $w(S_i[d])$. Column values (analogous to action values, as choices were of

columns) were computed as the sum over the weights of features in the three dimensions. The probability of choosing column c is modeled using a noisy softmax over the values with an inverse temperature parameter β:

$$V(S_i) = \sum_d w(S_i[d]) \tag{1}$$

$$\pi(c) = \frac{e^{\beta V(S_c)}}{\sum_i e^{\beta V(S_i)}} \tag{2}$$

We assume a standard reinforcement learning update rule operating at the feature level. At each time point, we compute the reward prediction error δ from the reward R and the value assigned to the chosen column, $\delta = R - V(S_c)$, and update each chosen feature using a learning rate η. For the unchosen columns, we decay all feature values toward zero with decay rate λ:

$$\text{For all } d: \ w(S_c[d]) = w(S_c[d]) + \eta\delta \tag{3}$$

$$\text{For all } d, i \neq c: \ w(S_i[d]) = (1-\lambda)w(S_i[d]) \tag{4}$$

3.2 Attention weighted feature RL (awFRL)

The attention-weighted feature RL model described in [2] includes empirically-derived dimensional attention weights as a direct measure of selective attention. These weights were computed in two ways: from eye position data, by binning looking time to each dimension within a trial; and from fMRI decoding of information in face-, landmark-, and tool-selective areas of the human cortex. The two measures of attention were then combined to form a single, empirically measured attentional weight for each dimension (d) on each trial, which we denote $\phi[d]$ (see [2] for full details).

The attention weights modified both the value computation and update rule of the FRL model. First, they biased value computation towards features in the attended dimensions. Next, the attention weights also biased the feature weight update, modeling increased attention to a particular dimension as a higher learning rate for features in that that dimension:

$$V(S_i) = \sum_d \phi[d] w(S_i[d]) \tag{5}$$

$$\text{For all } d: \ w(S_c[d]) = w(S_c[d]) + \eta\phi[d]\delta \tag{6}$$

As in the FRLdecay model, choice probabilities followed a noisy softmax distribution on column values.

3.3 Combined model: attention weighted feature RL with decay (awFRLdecay)

This model combines the above models, including both the attention-weighted value computation and chosen feature value updates (from [2]), as well as the decay of unchosen features towards zero (from [1]). We can formally describe the computations on each trial as:

$$V(S_i) = \sum_d \phi[d] w(S_i[d]) \quad \text{Attention-weighted value computation} \tag{7}$$

$$\pi(c) = \frac{e^{\beta V(S_c)}}{\sum_i e^{\beta V(S_i)}} \quad \text{Choice likelihood} \tag{8}$$

$$\text{For all } d: \ w(S_c[d]) = w(S_c[d]) + \eta\phi[d]\delta \quad \text{Attention-weighted feature weight updates} \tag{9}$$

$$\text{For all } d, i \neq c: \ w(S_i[d]) = (1-\lambda)w(S_i[d]) \quad \text{Unchosen feature decay (not attention-weighted)} \tag{10}$$

4 Results

As previously noted, the FRLdecay model and the awFRL model explained participants' choices equally well on average (Figure 2A). The new, combined, awFRLdecay model improved goodness of fit over both previous models, suggesting that retrospective decay and prospective biasing by attention explain different components of the variability in participants' choices. To test this hypothesis more directly, we next compared these models on a trial-by-trial basis, rather than averaging over all trials within each game (Figure 2B). We found that awFRL explained choices in the first half of each game better, while FRLdecay provided a better fit to choices in the later part of each game.

Motivated by the insight that the meaningful indexing within a game is not the absolute trial index, but the one relative to the participant's successful learning of the task, we repeated the trial-by-trial analysis, aligning data around the last trial in which the participant did not choose the target feature, i.e. the "point of learning"(Figure 2C). We found that the awFRL model predicted choices substantially better immediately following the point of learning, while the FRLdecay

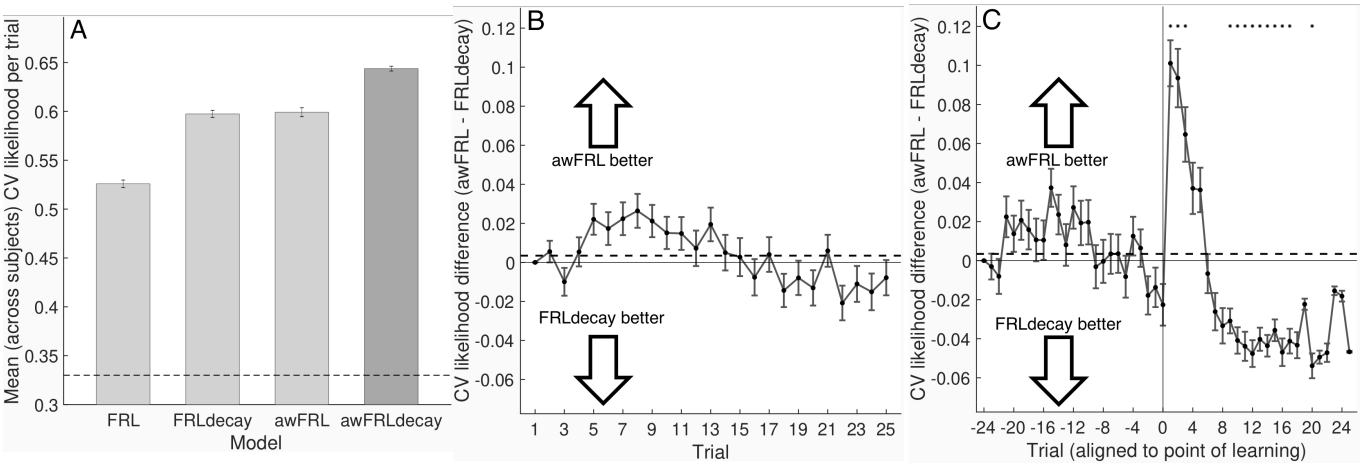

Figure 2: Trial-by-trial model comparison. A: Model comparison showing the mean cross-validated (CV) likelihood of participant's choices per trial. The FRLdecay (best fitting in [1]) and awFRL (best fitting in [2]) models both explain the data equally well, and better than a baseline FRL model (equations (1)-(3)). The combined model surpassed both previous models (paired-sample t tests, $t(24) = 8.998$, $p < .001$; $t(24) = 12.374$, $p < .001$, for the FRLdecay and awFRL models, respectively). **B**: Each bin shows the the average likelihood difference as a function of trial, subtracting the FRLdecay predicted likelihood from the awFRL predicted likelihood. **C**: The same trial-by-trial likelihood comparison, aligned to the last trial in each learned game in which the participant did not choose the target feature (the "point of learning"). Error bars: standard error of the mean (SEM). *: $p < 0.001$.

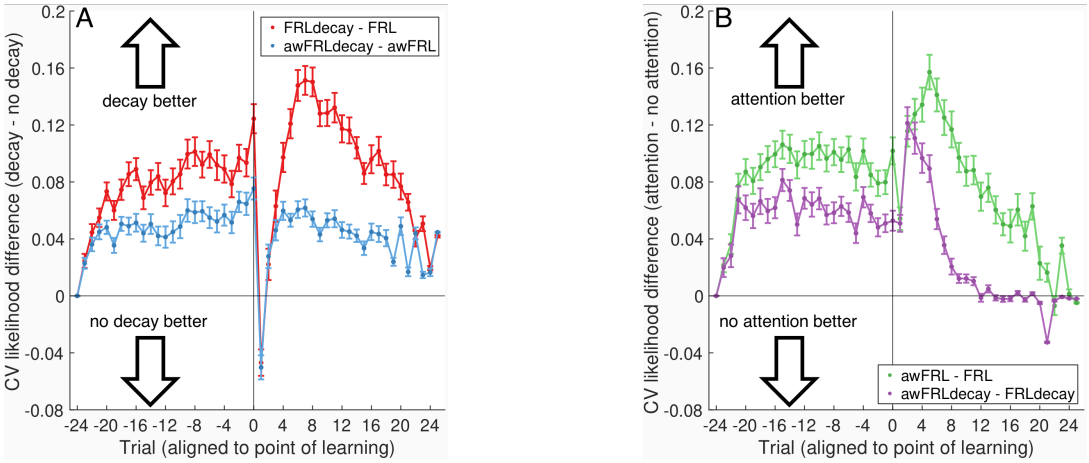

Figure 3: Isolating the effects of introducing decay and attention. A: The difference in choice likelihoods between FRLdecay and FRL (red) and awFRLdecay and awFRL (blue) aligned to the point of learning suggests that models that include retrospective decay lag in capturing the participant's point of learning. **B**: The difference in choice likelihoods between awFRL and FRL (green) and awFRLdecay and FRLdecay (purple) suggests that models that include measured prospective attention weights perform best shortly following the point of learning.

model performed better on the remainder of the game. Interestingly, both models fit the data to an equal extent before the point of learning, that is, during the representation learning phase (see discussion).

To dissociate whether differences in likelihood around the point of learning are due to awFRL being a better predictor of participants' choices, or FRLdecay being a worse predictor, Figure 3 shows the same trial-by-trial analysis as before, separating the effect of adding decay (A) and attention (B) to the models. Adding decay improved goodness of fit throughout, except for the trial immediately following the "point of learning," where decay models predicted choice substantially worse than models omitting decay. Conversely, models including attention weights predicted choices considerably better for a few trials after the "point of learning". Both model components improved goodness of fit before the point of learning.

5 Discussion

We compared the performance of two competing reinforcement learning models on a human representation learning task. The models, introduced in previous work, posit conceptually different mechanisms, one suggesting retrospective

"forgetting" of values learned in error, and the other employing prospective attention-gated learning of only those values that are expected to be useful. Nevertheless, the two models predict choice data equally well on average. The improved goodness of fit when combining both processes suggests that these models may account for different components of the variance in the choice data. We uncovered differences in predictive likelihood between the models using insight into the structure of the task: once participants successfully learn the correct task representation, they terminate a game with a streak of correct choices, transitioning from exploring potential representations to exploiting the correct one. By aligning data to the "point of learning," we recovered diverging predictive capacities of the FRLdecay and awFRL models. The retrospective FRLdecay model lags in capturing behavior at the explore-exploit transition, while the prospective awFRL attention model excels in predicting choices immediately following the transition. Together, these findings provide a more complete account of the complementary role of these mechanisms in enabling representation learning in multidimensional reinforcement learning tasks.

This transition from exploration to exploitation, and the models' success (or lack thereof) capturing it, allows a more nuanced examination of the two models. The inclusion of decay, which essentially creates a learning-relevant choice kernel [5], [6], as it maintains learning only for repeatedly chosen features, seems to improve overall model fit but fails to account for the immediate transition at the "point of learning." Perhaps this lag is due to its retrospective nature: first learn about everything, then decay anything that is revealed by participants' choices to have been learned (by the mode) in error. Models incorporating prospective attention fare better at accounting for behavior around this change-point, but rapidly lose their advantage as participants switch from exploration to exploitation. This effect may be due to the attention measures used diminishing in effectiveness, participants deploying less selective attention once they have learned the correct representation, or both. This comparison points to two other potential shortcomings of these models. Neither model explicitly accounts for the change-point between exploration and exploitation, even though it appears to be a distinct marker of the cognitive strategy employed by participants solving this task. Recent evidence suggests that confidence might govern this transition [7], suggesting that a model incorporating separate exploration and exploitation strategies may fare better at predicting the behavioral data. Of course, the transition may not be as abrupt as we have portrayed it to be, which such a model could attempt to capture. Here, reaction times for different choices may provide useful data that has been previously underexplored in the context of this task.

Another interesting finding is that the two models, which we have cast as conceptually different, explain the choice data equally well before the point of learning, that is, during the actual representation learning stage. Three possible explanations for this finding come to mind: 1) prospective and retrospective attention may contribute equally to the representation learning process itself, 2) the models may be disguising as one another in the learning phase, and 3) the differential contributions of the two processes are perhaps not separable with our task. In some sense, we cannot know if both models account for the cognitive strategy participants take, or possibly neither model captures it well.

Finally, from the perspective of model comparison methodology, our results reaffirm the value of diving beyond average model comparison metrics, such as mean cross-validated likelihood and BIC, and employing more granular comparisons. Examining specific cases in which competing models make different predictions enables more nuanced investigation than the overall goodness of fit of each model. Had we known in advance that such a stark behavioral change-point exists in the data, we could have simulated data using this "qualitative signature," which we expect a good model for the task to be able to recover, and evaluated the models accordingly [8].

6 References

[1] Y. Niv, R. Daniel, A. Geana, S. J. Gershman, Y. C. Leong, A. Radulescu, and R. C. Wilson, "Reinforcement learning in multidimensional environments relies on attention mechanisms," *The Journal of Neuroscience*, vol. 35, no. 21, pp. 8145–8157, May 27, 2015.

[2] Y. C. Leong, A. Radulescu, R. Daniel, V. DeWoskin, and Y. Niv, "Dynamic interaction between reinforcement learning and attention in multidimensional environments," *Neuron*, vol. 93, no. 2, pp. 451–463, Jan. 2017.

[3] N. J. Mackintosh, "A theory of attention: Variations in the associability of stimuli with reinforcement," *Psychological Review*, vol. 82, no. 4, pp. 276–298, 1975.

[4] J. Gottlieb, "Attention, learning, and the value of information," *Neuron*, vol. 76, no. 2, pp. 281–295, Oct. 18, 2012.

[5] B. Lau and P. W. Glimcher, "Dynamic response-by-response models of matching behavior in rhesus monkeys," *Journal of the Experimental Analysis of Behavior*, vol. 84, no. 3, pp. 555–579, Nov. 2005.

[6] R. Akaishi, K. Umeda, A. Nagase, and K. Sakai, "Autonomous mechanism of internal choice estimate underlies decision inertia," *Neuron*, vol. 81, pp. 195–206, Jan. 8, 2014.

[7] A. Boldt, C. Blundell, and B. De Martino, "Confidence modulates exploration and exploitation in value-based learning," 2017.

[8] R. C. Wilson and A. Collins, "Ten simple rules for the computational modeling of behavioral data," 2019.