Lecture 10: Learning to predict

Samuel Gershman

Harvard University

Roadmap

- We will examine reward and punishment prediction (with classical conditioning as a case study) as a paradigmatic predictive learning problem.

Roadmap

- We will examine reward and punishment prediction (with classical conditioning as a case study) as a paradigmatic predictive learning problem.
- Simple reinforcement learning algorithms, relying on gradient-based optimization, can explain many aspects of how animals learn to predict reward/punishment.

Roadmap

- We will examine reward and punishment prediction (with classical conditioning as a case study) as a paradigmatic predictive learning problem.
- Simple reinforcement learning algorithms, relying on gradient-based optimization, can explain many aspects of how animals learn to predict reward/punishment.
- Can be implemented in the basal ganglia, where dopamine provides the critical error signal for learning.

The reinforcement learning problem

- Given state s, predict cumulative future reward over some (possibly infinite) horizon.

The reinforcement learning problem

- Given state s, predict cumulative future reward over some (possibly infinite) horizon.
- We will start with the problem of immediate reward prediction, $\mathbb{E}[r|s]$ and then generalize to long-term reward prediction.

The reinforcement learning problem

- Given state s, predict cumulative future reward over some (possibly infinite) horizon.
- We will start with the problem of immediate reward prediction, $\mathbb{E}[r|s]$ and then generalize to long-term reward prediction.
- We will start with point estimation, and then generalize to distributional prediction.

Classical conditioning

- Animal is exposed to a neutral stimulus (the conditioned stimulus, or CS) followed by an appetitive (good) or aversive (bad) stimulus (the unconditioned stimulus, or US).

Classical conditioning

- Animal is exposed to a neutral stimulus (the conditioned stimulus, or CS) followed by an appetitive (good) or aversive (bad) stimulus (the unconditioned stimulus, or US).
- *Delay conditioning*: onset of the US coincides with the offset of the CS.

Classical conditioning

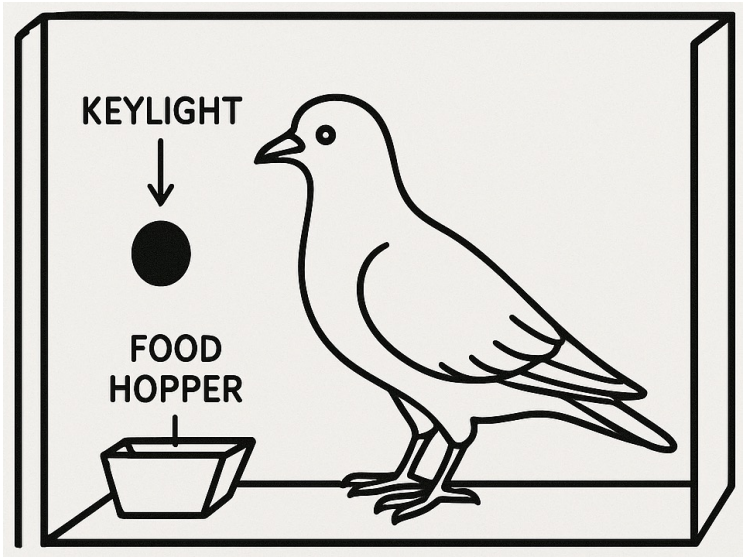
- Animal is exposed to a neutral stimulus (the conditioned stimulus, or CS) followed by an appetitive (good) or aversive (bad) stimulus (the unconditioned stimulus, or US).
- *Delay conditioning*: onset of the US coincides with the offset of the CS.
- Key variable: *conditioned response* (CR) to the CS onset.

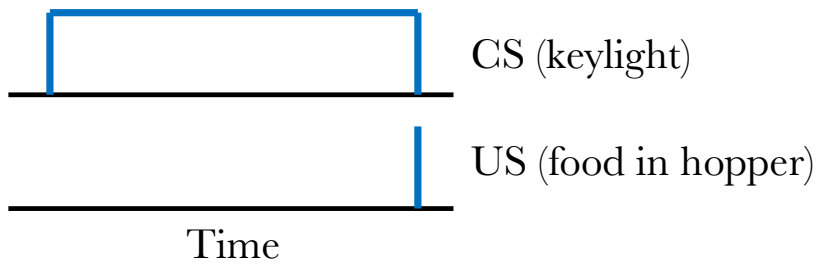
Classical conditioning

- Animal is exposed to a neutral stimulus (the conditioned stimulus, or CS) followed by an appetitive (good) or aversive (bad) stimulus (the unconditioned stimulus, or US).
- *Delay conditioning*: onset of the US coincides with the offset of the CS.
- Key variable: *conditioned response* (CR) to the CS onset.
- CR increases over the course of conditioning. Not due merely to repeated stimulus exposure, because it does not occur if the relative timing of the CS and US is randomized (i.e., there is no stable temporal relationship between the two stimuli); temporal relationship is fundamental.

Pigeon autoshaping

Canonical example of classical conditioning. CS is keylight which predicts the delivery of food (the US) into the hopper. With repeated pairings, the pigeon begins to peck at the keylight (the CR). Note that food delivery is independent of pecking.





Some notation

- Uppercase letters (A, B, C, etc.) denote CSs; compound CSs will be denoted by concatenations (e.g., AB denotes the compound presentation of A and B).

Some notation

- Uppercase letters (A, B, C, etc.) denote CSs; compound CSs will be denoted by concatenations (e.g., AB denotes the compound presentation of A and B).
- A reinforced CS will be denoted by A+, and an unreinforced CS will be denoted A-.

Some notation

- Uppercase letters (A, B, C, etc.) denote CSs; compound CSs will be denoted by concatenations (e.g., AB denotes the compound presentation of A and B).
- A reinforced CS will be denoted by A+, and an unreinforced CS will be denoted A-.
- A test stimulus (typically presented without reinforcement) is denoted by B? and the resulting test CR is the variable of interest.

Predictive view of classical conditioning

- Hypothesis: CR reflects a prediction about upcoming reward.

Predictive view of classical conditioning

- Hypothesis: CR reflects a prediction about upcoming reward.
- CR rate is greater when the CS-US delay is shorter (except for very short delays, at which point anticipation may not be useful) and when the reinforcement rate (the CS-conditional US rate) is greater.

Predictive view of classical conditioning

- Hypothesis: CR reflects a prediction about upcoming reward.
- CR rate is greater when the CS-US delay is shorter (except for very short delays, at which point anticipation may not be useful) and when the reinforcement rate (the CS-conditional US rate) is greater.
- Suggests that the CR is closely tied to the expected rate of reinforcement in the near future following the appearance of the CS.

Prediction errors drive learning

- Unexpected US following the CS tends to increase the CR, whereas the omission of an expected US following the CS tends to decrease the CR.

Prediction errors drive learning

- Unexpected US following the CS tends to increase the CR, whereas the omission of an expected US following the CS tends to decrease the CR.
- CR is weaker if CS (A) is paired with another previously reinforced CS (B), compared to reinforcing the CS alone.

Prediction errors drive learning

- Unexpected US following the CS tends to increase the CR, whereas the omission of an expected US following the CS tends to decrease the CR.
- CR is weaker if CS (A) is paired with another previously reinforced CS (B), compared to reinforcing the CS alone.
- In both cases, the CS was reinforced the same number of times, but in the former case the previously reinforced CS B "blocks" the new CS A because it already adequately predicts the US—no prediction error to drive learning.

Prediction errors drive learning

- **Overexpectation**: two CSs are reinforced separately (A+/B+), then reinforced in compound (AB+), and finally the response to each one is tested individually (A? B?).

Prediction errors drive learning

- **Overexpectation**: two CSs are reinforced separately (A+/B+), then reinforced in compound (AB+), and finally the response to each one is tested individually (A? B?).
- This produces a *reduction* in the CR compared to a protocol in which the compound reinforcement phase is omitted.

Prediction errors drive learning

- **Overexpectation**: two CSs are reinforced separately (A+/B+), then reinforced in compound (AB+), and finally the response to each one is tested individually (A? B?).
- This produces a *reduction* in the CR compared to a protocol in which the compound reinforcement phase is omitted.
- In other words, conditioned responding is lower despite the animal receiving more reinforcements! Why?

Prediction errors drive learning

- **Overexpectation**: two CSs are reinforced separately (A+/B+), then reinforced in compound (AB+), and finally the response to each one is tested individually (A? B?).
- This produces a *reduction* in the CR compared to a protocol in which the compound reinforcement phase is omitted.
- In other words, conditioned responding is lower despite the animal receiving more reinforcements! Why?
- Animals come to predict a fixed amount of reward for each individual CS, and then predict twice as much when the two CSs are presented together (under a summation assumption). The receipt of the same amount produces a negative prediction error, driving reduction in predictions for each CS.

The Rescorla-Wagner model

- Compound configuration represented by a vector x, where $x_d = 1$ denotes the presence of CS d and $x_d = 0$ denotes its absence.

The Rescorla-Wagner model

- Compound configuration represented by a vector x, where $x_d = 1$ denotes the presence of CS d and $x_d = 0$ denotes its absence.
- US is denoted by r.

The Rescorla-Wagner model

- Compound configuration represented by a vector x, where $x_d = 1$ denotes the presence of CS d and $x_d = 0$ denotes its absence.
- US is denoted by r.
- US prediction $\hat{r}$ is a linear function of the CS vector:

$$\hat{r} = \sum_d w_d x_d$$

where w_d is an associative strength (weight) between CS d and the US, typically initialized to 0.

The Rescorla-Wagner model

- Weights are updated based on the prediction error $r - \hat{r}$:

$$\Delta w_d = \eta x_d (r - \hat{r})$$

where $\eta \in [0, 1]$ is a learning rate.

The Rescorla-Wagner model

- Weights are updated based on the prediction error $r - \hat{r}$:

$$\Delta w_d = \eta x_d (r - \hat{r})$$

  where $\eta \in [0, 1]$ is a learning rate.
- Weight is increased when the prediction error is positive (more reward was received than predicted) and decreased when the prediction error is negative (less reward was received than predicted).

The Rescorla-Wagner model

- Weights are updated based on the prediction error $r - \hat{r}$:

$$\Delta w_d = \eta x_d (r - \hat{r})$$

  where $\eta \in [0, 1]$ is a learning rate.
- Weight is increased when the prediction error is positive (more reward was received than predicted) and decreased when the prediction error is negative (less reward was received than predicted).
- CS must be present in order for its weight to change.

The Rescorla-Wagner model

- Despite its simplicity, the RW model can explain a wide range of classical conditioning phenomena.

The Rescorla-Wagner model

- Despite its simplicity, the RW model can explain a wide range of classical conditioning phenomena.
- **Blocking**: if CS A has already been paired with the US, then $w_A \approx 1$. During the compound conditioning phase $\hat{r} = w_A + w_B \approx 1 + 0$, and thus the prediction error $r - \hat{r}$ is approximately 0, preventing learning of a non-zero weight for CS B.

The Rescorla-Wagner model

- Despite its simplicity, the RW model can explain a wide range of classical conditioning phenomena.
- **Blocking**: if CS A has already been paired with the US, then $w_A \approx 1$. During the compound conditioning phase $\hat{r} = w_A + w_B \approx 1 + 0$, and thus the prediction error $r - \hat{r}$ is approximately 0, preventing learning of a non-zero weight for CS B.
- **Overexpectation**: after separate reinforcement of A and B, each weight ≈ 1. In compound, the US prediction is $\hat{r} = w_A + w_B \approx 1 + 1 = 2$. When $r = 1$ is received during the compound conditioning phase, the prediction error is $r - \hat{r} \approx 1 - 2 = -1$. This leads to a decrement of both w_A and w_B.

The Rescorla-Wagner model

Key principles:

- Learning driven by prediction errors.
- Additive combination of weights.
- Credit assignment based on CS presence.

The Rescorla-Wagner model and gradient descent

- Loss function based on predictive accuracy: $L(\hat{r}, r) = (r - \hat{r})^2$.

The Rescorla-Wagner model and gradient descent

- Loss function based on predictive accuracy: $L(\hat{r}, r) = (r - \hat{r})^2$.
- Taking the gradient of the loss with respect to the weights yields $\nabla_w L \propto x(r - \hat{r})$.

The Rescorla-Wagner model and gradient descent

- Loss function based on predictive accuracy: $L(\hat{r}, r) = (r - \hat{r})^2$.
- Taking the gradient of the loss with respect to the weights yields $\nabla_w L \propto x(r - \hat{r})$.
- Thus, the RW update can be derived as gradient descent on the squared error loss.

Learning in the absence of stimuli

- Animals don't learn in the absence of prediction errors... or do they?

Learning in the absence of stimuli

- Animals don't learn in the absence of prediction errors... or do they?
- **Latent inhibition**: prior to CS-US pairings, the CS is presented repeatedly by itself. These CS-alone trials retard subsequent acquisition of the CR.

Learning in the absence of stimuli

- Animals don't learn in the absence of prediction errors... or do they?
- **Latent inhibition**: prior to CS-US pairings, the CS is presented repeatedly by itself. These CS-alone trials retard subsequent acquisition of the CR.
- If the weight is initialized to 0, so the US prediction is initially 0 when the CS-alone trials occur, then the prediction error will also be 0. The Rescorla-Wagner model predicts no learning in this case, and yet the animal is clearly learning something.

Learning in the absence of stimuli

- **Backward blocking**: switch the order of A+ and AB+ training, so that AB+ comes first. After AB+ training, the animal produces a CR in response to B, but this response is reduced after A+ training.

Learning in the absence of stimuli

- **Backward blocking**: switch the order of A+ and AB+ training, so that AB+ comes first. After AB+ training, the animal produces a CR in response to B, but this response is reduced after A+ training.
- This can't happen in the RW model, because $x_B = 0$ during A+ training, and therefore Δw_B must equal 0.

Retrospective revaluation

- In the Kamin blocking protocol, presenting A by itself after compound training (A+/AB+/A-) has the effect of "unblocking" B [Blaisdell et al 1999].

Retrospective revaluation

- In the Kamin blocking protocol, presenting A by itself after compound training (A+/AB+/A-) has the effect of "unblocking" B [Blaisdell et al 1999].
- Presenting A by itself after the compound training phase of the overexpectation protocol (A+/B+/AB+/A-) has the effect of rescuing the CR to B [Blaisdell et al 2001].

A probabilistic view

- Gradient descent on the squared error loss is a point estimation procedure; it ignores uncertainty about the weights.

A probabilistic view

- Gradient descent on the squared error loss is a point estimation procedure; it ignores uncertainty about the weights.
- We can derive a different normative analysis by computing a full posterior over the weights given the stimulus history.

Kalman filtering

- Suppose that the US is a noisy linear combination of CS features:

$$r = \sum_d w_d x_d + \epsilon$$

where $\epsilon \sim \mathcal{N}(0, \sigma_r^2)$ and $w_d \sim \mathcal{N}(0, \sigma_w^2)$.

Kalman filtering

- Suppose that the US is a noisy linear combination of CS features:

$$r = \sum_d w_d x_d + \epsilon$$

where $\epsilon \sim \mathcal{N}(0, \sigma_r^2)$ and $w_d \sim \mathcal{N}(0, \sigma_w^2)$.

- Posterior is Gaussian, with mean $\hat{w}$ and covariance matrix Σ, updated according to:

$$\Delta\hat{w} = \eta(r - \hat{r}), \qquad \Delta\Sigma = -\eta x^\top \Sigma$$

where $\eta = \frac{\Sigma x}{x^\top \Sigma x + \sigma_r^2}$ is now vector-valued with CS-specific learning rates.

Relationship to Rescorla-Wagner

- If the covariance is the identity matrix, $\Sigma = I$, then the two models are equivalent apart from a time-varying learning rate.

Relationship to Rescorla-Wagner

- If the covariance is the identity matrix, $\Sigma = I$, then the two models are equivalent apart from a time-varying learning rate.
- In general, the covariance will not be diagonal, and this has important implications for classical conditioning.

Single CS case

- Σ is a scalar. Each time the CS is presented, Σ decreases.

Single CS case

- Σ is a scalar. Each time the CS is presented, Σ decreases.
- This has the effect of decreasing the learning rate. Intuitively, the animal becomes more confident as it collects more data, which makes it more resistant to learning from future observations.

Single CS case

- Σ is a scalar. Each time the CS is presented, Σ decreases.
- This has the effect of decreasing the learning rate. Intuitively, the animal becomes more confident as it collects more data, which makes it more resistant to learning from future observations.
- Latent inhibition (slower learning after CS pre-exposure) is a natural consequence of these dynamics.

Study question

Why does the probabilistic Kalman filter model naturally account for latent inhibition, while the Rescorla-Wagner model does not?

Compound conditioning

- Off-diagonals of the covariance matrix become negative.

Compound conditioning

- Off-diagonals of the covariance matrix become negative.
- Zero-sum nature of the linear model: associative strengths must add to 1, so increasing the weight for one CS requires that the weight for the other be decreased.

Compound conditioning

- Off-diagonals of the covariance matrix become negative.
- Zero-sum nature of the linear model: associative strengths must add to 1, so increasing the weight for one CS requires that the weight for the other be decreased.
- Produces many retrospective revaluation effects, because the learning rates for absent stimuli will be negative. In the backward blocking protocol (AB+/A+), for example,strengthening w_A during the second phase weakens w_B due to $\eta_B < 0$ during A+ training.

Long-range prediction

- We've been assuming that the computational problem facing the animal is predicting the next US, but this neglects the fact that animals care about events farther in the future.

Long-range prediction

- We've been assuming that the computational problem facing the animal is predicting the next US, but this neglects the fact that animals care about events farther in the future.
- **Second-order conditioning**: A+ followed by B $\rightarrow$ A. Animals acquire a CR to B even though it is never paired with the US.

Long-range prediction

- We've been assuming that the computational problem facing the animal is predicting the next US, but this neglects the fact that animals care about events farther in the future.
- **Second-order conditioning**: A+ followed by B $\rightarrow$ A. Animals acquire a CR to B even though it is never paired with the US.
- Apparently the animals treat A as a proxy for future reinforcement.

Long-range prediction

- Goal is now to predict *expected discounted return*, or *value*:

$$V_t = \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots]$$

where we have introduced the time index t. The parameter $\gamma \in [0, 1)$ is the *temporal discount factor*, which reflects the animal's preference for obtaining rewards sooner rather than later.

Long-range prediction

- Goal is now to predict *expected discounted return*, or *value*:

$$V_t = \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots]$$

  where we have introduced the time index t. The parameter $\gamma \in [0, 1)$ is the *temporal discount factor*, which reflects the animal's preference for obtaining rewards sooner rather than later.
- The expectation averages over stochasticity in the reward sequence.

Long-range prediction

- Goal is now to predict *expected discounted return*, or *value*:

$$V_t = \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots]$$

  where we have introduced the time index t. The parameter $\gamma \in [0, 1)$ is the *temporal discount factor*, which reflects the animal's preference for obtaining rewards sooner rather than later.
- The expectation averages over stochasticity in the reward sequence.
- How can an animal estimate an expectation of this infinite series?

Long-range prediction

- Assume that the rewards are conditionally independent given an underlying state s: the expected reward in state s is given by $R(s)$.

Long-range prediction

- Assume that the rewards are conditionally independent given an underlying state s: the expected reward in state s is given by $R(s)$.
- State evolves according to a Markov process with transition probability $T(s'|s)$.

Long-range prediction

- Assume that the rewards are conditionally independent given an underlying state s: the expected reward in state s is given by $R(s)$.
- State evolves according to a Markov process with transition probability $T(s'|s)$.
- Value can be expressed in a recursive form (the *Bellman equation*):

$$V(s) = R(s) + \gamma \sum_{s'} T(s'|s) V(s')$$

where we have redefined the value as a function of the state.

Learning values

- The Bellman equation stipulates that the *temporal difference* (TD) prediction error $\delta = r + \gamma \hat{V}(s') - \hat{V}(s)$ is on average 0 when the approximation is exact, $\hat{V} = V$.

Learning values

- The Bellman equation stipulates that the *temporal difference* (TD) prediction error $\delta = r + \gamma \hat{V}(s') - \hat{V}(s)$ is on average 0 when the approximation is exact, $\hat{V} = V$.
- When $\delta > 0$, the value function has been underestimated and $\hat{V}(s)$ should be increased; when $\delta < 0$, the value function has been overestimated and $\hat{V}(s)$ should be decreased.

Learning values

- The Bellman equation stipulates that the *temporal difference* (TD) prediction error $\delta = r + \gamma \hat{V}(s') - \hat{V}(s)$ is on average 0 when the approximation is exact, $\hat{V} = V$.
- When $\delta > 0$, the value function has been underestimated and $\hat{V}(s)$ should be increased; when $\delta < 0$, the value function has been overestimated and $\hat{V}(s)$ should be decreased.
- TD learning updates $\hat{V}$ according to $\Delta \hat{V}(s) \propto \delta$.

Learning values

- The Bellman equation stipulates that the *temporal difference* (TD) prediction error $\delta = r + \gamma \hat{V}(s') - \hat{V}(s)$ is on average 0 when the approximation is exact, $\hat{V} = V$.
- When $\delta > 0$, the value function has been underestimated and $\hat{V}(s)$ should be increased; when $\delta < 0$, the value function has been overestimated and $\hat{V}(s)$ should be decreased.
- TD learning updates $\hat{V}$ according to $\Delta \hat{V}(s) \propto \delta$.
- We can understand this update as gradient descent on a TD loss function $L(\hat{V}(s), \hat{V}^*(s)) = [\hat{V}^*(s) - \hat{V}(s)]^2 = \delta^2$, where the target $\hat{V}^*(s) = R(s) + \gamma \hat{V}(s')$ is "bootstrapped"—it relies on the estimator for the next state.

Linear function approximation

- We need to go beyond the value look-up table to model classical conditioning. Represent each state with a feature vector $x = f(s)$ for some encoding function f, with $\hat{V}(s) = \sum_d w_d x_d$.

Linear function approximation

- We need to go beyond the value look-up table to model classical conditioning. Represent each state with a feature vector $x = f(s)$ for some encoding function f, with $\hat{V}(s) = \sum_d w_d x_d$.
- Gradient descent on the TD loss leads to the following update: $\Delta w_d \propto x_d \delta$.

Linear function approximation

- We need to go beyond the value look-up table to model classical conditioning. Represent each state with a feature vector $x = f(s)$ for some encoding function f, with $\hat{V}(s) = \sum_d w_d x_d$.
- Gradient descent on the TD loss leads to the following update: $\Delta w_d \propto x_d \delta$.
- This is very similar to RW. In fact, the two learning rules are identical in the "myopic limit" ($\gamma = 0$), where the animal only cares about predicting immediate reward.

Linear function approximation

- We need to go beyond the value look-up table to model classical conditioning. Represent each state with a feature vector $x = f(s)$ for some encoding function f, with $\hat{V}(s) = \sum_d w_d x_d$.
- Gradient descent on the TD loss leads to the following update: $\Delta w_d \propto x_d \delta$.
- This is very similar to RW. In fact, the two learning rules are identical in the "myopic limit" ($\gamma = 0$), where the animal only cares about predicting immediate reward.
- TD learning goes beyond RW by estimating long-range predictions. This allows TD learning to capture phenomena like second-order conditioning, because the future reward term $\gamma \hat{V}(s')$ is greater than 0 during B$\rightarrow$A training even though the immediate reward (r) is 0.

Study question

In what sense can temporal difference learning be viewed as a generalization of the Rescorla-Wagner model? How does the discount factor γ expand the predictive horizon of learning?

The neural architecture of reinforcement learning

- Value estimate: nucleus accumbens (ventral striatum).

The neural architecture of reinforcement learning

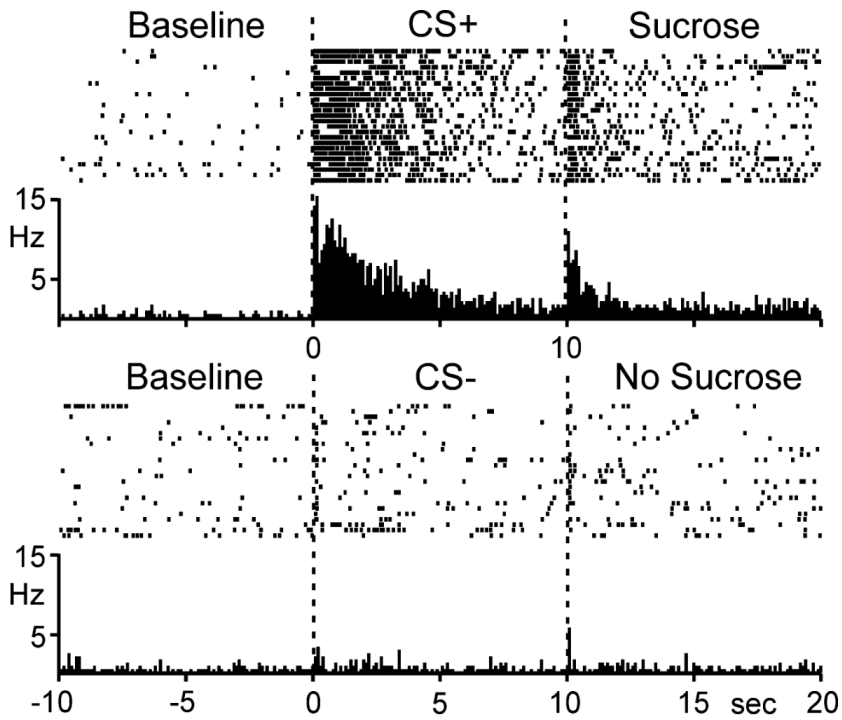
- Value estimate: nucleus accumbens (ventral striatum).
- State features: cortical inputs to the striatum.

The neural architecture of reinforcement learning

- Value estimate: nucleus accumbens (ventral striatum).
- State features: cortical inputs to the striatum.
- TD error (δ): phasic dopamine.

Ventral striatum representation of value

Responses of a single neuron in the nucleus accumbens following classical conditioning.



[Day et al 2006]

Predictive Hebbian learning

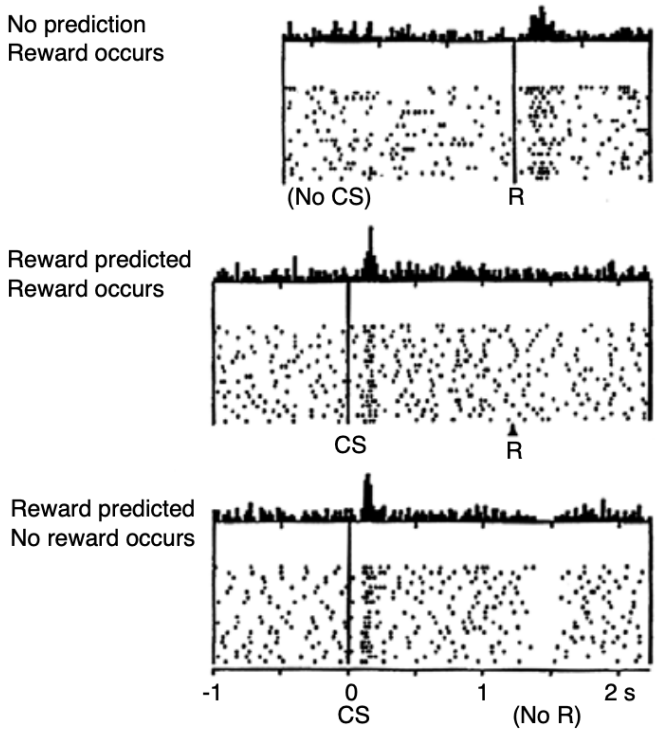
- If cortical inputs to the striatum encode the feature vector $x = f(s)$, and striatal neurons encode $\hat{V}(s)$, then (under the linear function approximation assumption) corticostriatal synapses correspond to the weights w.

Predictive Hebbian learning

- If cortical inputs to the striatum encode the feature vector $x = f(s)$, and striatal neurons encode $\hat{V}(s)$, then (under the linear function approximation assumption) corticostriatal synapses correspond to the weights w.
- We can understand the TD update $\Delta w_d \propto x_d \delta$ as a form of "predictive Hebbian learning" [Montague et al 1996], where weights are updated in proportion to the coincidence of presynaptic activity (x_d) with a prediction error signal (δ).

Dopamine signaling of prediction errors

Activity of dopamine neurons during classical conditioning. "R" indicates reward delivery.



[Schultz et al 1997]

Stimulus representation

- A standard assumption, known as the *complete serial compound* (CSC), breaks each stimulus down into a contiguous set of binary temporal features. Only one feature is active ($x_d = 1$) during each time interval relative to the stimulus onset.

Stimulus representation

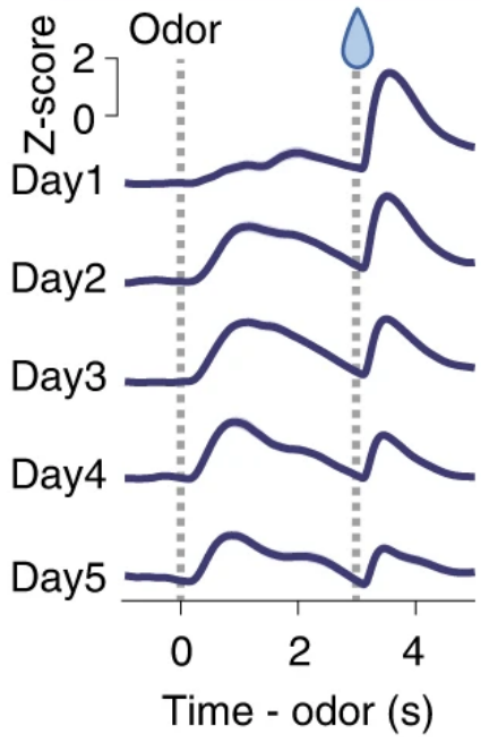
- A standard assumption, known as the *complete serial compound* (CSC), breaks each stimulus down into a contiguous set of binary temporal features. Only one feature is active ($x_d = 1$) during each time interval relative to the stimulus onset.
- This allows the function approximator to learn value estimates for each interval.

Stimulus representation

[Ludvig et al 2012]

Backward propagation

CSC representation correctly predicts that dopamine signals should propagate backwards over time from the US to the CS, as the weight for each feature is progressively updated.



[Amo et al 2022]

Problems for the CSC

- A large number of such features is needed to approximate the value function well, and they don't afford any temporal generalization.

Problems for the CSC

- A large number of such features is needed to approximate the value function well, and they don't afford any temporal generalization.
- For example, if an animal is trained with one CS-US interval and then tested on a slightly longer interval, the weights learned for the CSC representation will abruptly drop after the expected US time. In contrast, animals show a gradual decline in conditioned responding when tested on longer intervals.

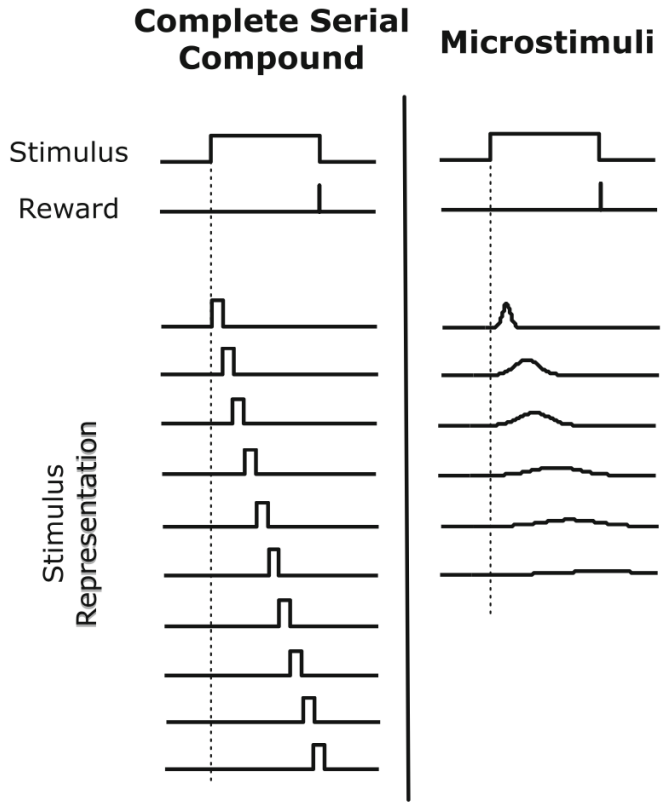
Microstimulus representation

- An alternative *microstimulus* representation was proposed by Ludvig et al [2008].

Microstimulus representation

- An alternative *microstimulus* representation was proposed by Ludvig et al [2008].
- Each microstimulus corresponds to a radial tuning function with a particular preferred time interval.

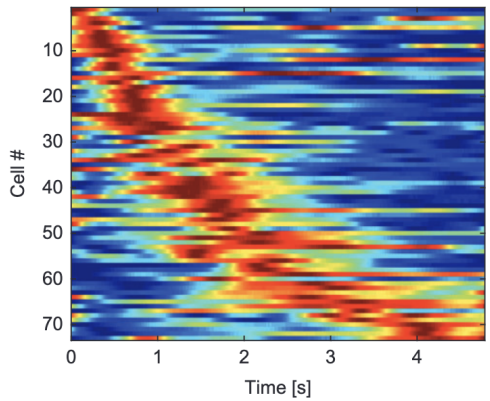
Stimulus representation



[Ludvig et al 2012]

Time cells in the medial prefrontal cortex

Each row corresponds to a single neuron, sorted by preferred time interval. The color shows firing rate.



[Tiganj et al 2017]

Summary

- Common algorithmic motif: Learning from prediction errors.

Summary

- Common algorithmic motif: Learning from prediction errors.
- Credit assignment is based on the active representation of stimuli (both those that are present and those that are linked to the present stimuli).

Summary

- Common algorithmic motif: Learning from prediction errors.
- Credit assignment is based on the active representation of stimuli (both those that are present and those that are linked to the present stimuli).
- Learning is sensitive to uncertainty.

Summary

- Common algorithmic motif: Learning from prediction errors.
- Credit assignment is based on the active representation of stimuli (both those that are present and those that are linked to the present stimuli).
- Learning is sensitive to uncertainty.
- Prediction target for learning is long-range (not just immediate upcoming reward).

Summary

- Common algorithmic motif: Learning from prediction errors.
- Credit assignment is based on the active representation of stimuli (both those that are present and those that are linked to the present stimuli).
- Learning is sensitive to uncertainty.
- Prediction target for learning is long-range (not just immediate upcoming reward).
- These principles can be realized neurally using simple mechanisms.