Lecture 9: Rescorla-Wagner rule

Jochen Braun

Otto-von-Guericke-Universität Magdeburg,
Cognitive Biology Group

Theoretical Neuroscience II, SS 2020

Credits:
Dayan & Abbott (2001) Chpt. 9; W Schultz (2002) Neuron

Lecture 9: Rescorla-Wagner rule

We define **'conditioning'** *or* **'reinforcement'** *as learning by trial and error, either to expect and experience reinforcement passively ('classical or Pavlovian'), or to expect and procure reinforcement actively ('instrumental or operant'). The* **Rescorla-Wagner** *rule unifies decades of behavioural and computational results. Observed events described by binary* $\boldsymbol{u} \in \{0, 1\}$, *reinforcement by scalar* r. *Learning of event-specific expectation* $\boldsymbol{w}$ *predicts total expected reward, given observations, of* $v = \boldsymbol{w} \cdot \boldsymbol{u}$. *Prediction error* $\delta = r - v$ *drives changes expectations* $\Delta \boldsymbol{w} \simeq \delta \boldsymbol{u}$. *RW rule accounts for* complete, partial, *and* extinction *reinforcement of single events, as well as* complete, partial, exctinction, blocking, *and* overshadowing *reinforcement of multiple events.* **Dopamine neuron** *activity is thought to encode* prediction error *and to provide reinforcement signal to cortex.*

Organization of lecture

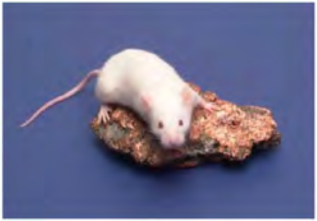
Conditioning: a primitive, reflex-like kind of training, in which specific actions are encouraged (or discouraged) by rewards (or punishments).

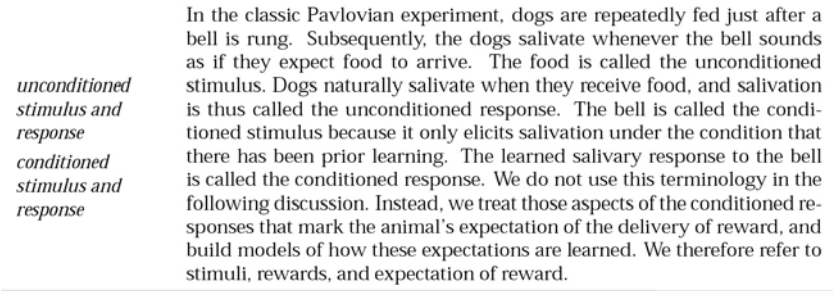
Reinforcement: an area of machine learning where agents learn by trial and error (experience of reward or punishment).

- **1 Basic terms**
- **2 Rescorla-Wagner rule**
- **3 Successes for single stimuli**
- **4 Successes for multiple stimuli**
- **5 Dopamine and reward**

1 Basic terms

- *Reinforcements:* rewards or punishments
- *Classical or Pavlovian conditioning:* Reinforcements are independent of the animal's actions. Animal is conditioned to passively expect reinforcements.
- *Instrumental or operant conditioning:* Reinforcements depend also on the animal's actions. Animals are conditioned to actively procure reinforcements.
- *Reinforcement learning:* Learning to expect or procure reinforcements by trial and error. The only available guidance are the reinforcements awarded or withheld.



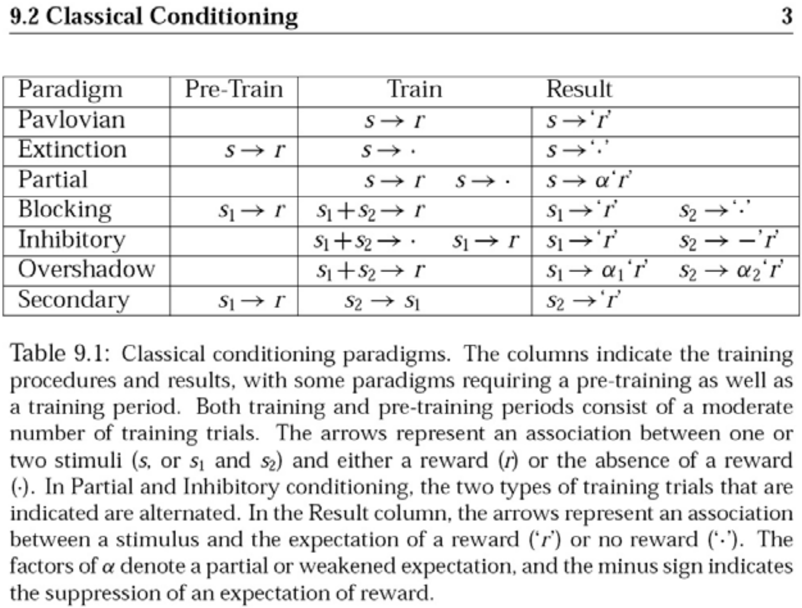


unconditioned stimulus and response

conditioned stimulus and response

In the classic Pavlovian experiment, dogs are repeatedly fed just after a bell is rung. Subsequently, the dogs salivate whenever the bell sounds as if they expect food to arrive. The food is called the unconditioned stimulus. Dogs naturally salivate when they receive food, and salivation is thus called the unconditioned response. The bell is called the conditioned stimulus because it only elicits salivation under the condition that there has been prior learning. The learned salivary response to the bell is called the conditioned response. We do not use this terminology in the following discussion. Instead, we treat those aspects of the conditioned responses that mark the animal's expectation of the delivery of reward, and build models of how these expectations are learned. We therefore refer to stimuli, rewards, and expectation of reward.

Чучело собаки. Подарок Института экспериментальной медицины г. Ленинграда. 1975 год



Paradigm	Pre-Train	Train		Result	
Pavlovian		$s \rightarrow r$		$s \rightarrow$ ‘r’	
Extinction	$s \rightarrow r$	$s \rightarrow \cdot$		$s \rightarrow$ ‘$\cdot$’	
Partial		$s \rightarrow r$	$s \rightarrow \cdot$	$s \rightarrow \alpha$‘r’	
Blocking	$s_1 \rightarrow r$	$s_1 + s_2 \rightarrow r$		$s_1 \rightarrow$ ‘r’	$s_2 \rightarrow$ ‘$\cdot$’
Inhibitory		$s_1 + s_2 \rightarrow \cdot$	$s_1 \rightarrow r$	$s_1 \rightarrow$ ‘r’	$s_2 \rightarrow -$‘r’
Overshadow		$s_1 + s_2 \rightarrow r$		$s_1 \rightarrow \alpha_1$‘r’	$s_2 \rightarrow \alpha_2$‘r’
Secondary	$s_1 \rightarrow r$	$s_2 \rightarrow s_1$		$s_2 \rightarrow$ ‘r’	

Table 9.1: Classical conditioning paradigms. The columns indicate the training procedures and results, with some paradigms requiring a pre-training as well as a training period. Both training and pre-training periods consist of a moderate number of training trials. The arrows represent an association between one or two stimuli (s, or s_1 and s_2) and either a reward (r) or the absence of a reward ($\cdot$). In Partial and Inhibitory conditioning, the two types of training trials that are indicated are alternated. In the Result column, the arrows represent an association between a stimulus and the expectation of a reward (‘r’) or no reward (‘$\cdot$’). The factors of α denote a partial or weakened expectation, and the minus sign indicates the suppression of an expectation of reward.

2 Rescorla-Wagner rule

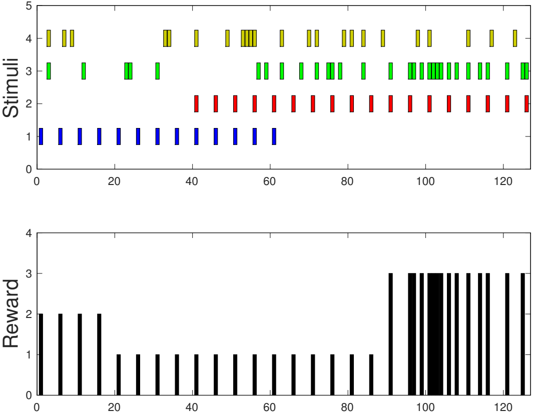
We model how animals learn to expect a reward in terms of the "Rescorla-Wagner rule". This rule captures many (but not all) aspects of the vast experimental literature on classical conditioning.

Following D&A, we use terms such as "stimuli", "rewards", and "expectation of rewards", rather than "conditioned stimuli", "unconditioned stimuli", and "conditioned response".

We introduce two **external** state variables ...

$$\textit{Stimulus (vector)} \qquad \boldsymbol{u}(t) = [u_1, u_2, u_3, u_4] \in \begin{cases} 1 & \text{if present} \\ 0 & \text{if absent} \end{cases}$$

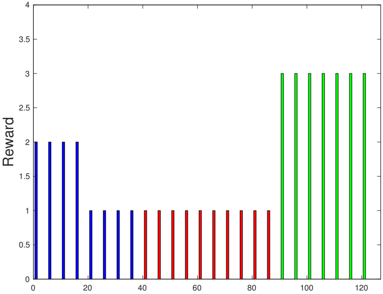
$$\textit{Reward (scalar)} \qquad r(t)$$



124 trials: up to four stimuli $\boldsymbol{u}(t)$ (top) and reward $r(t)$ (bottom).

... and postulate a **causal heuristic** between **u** and r:

- Assumption I: stimuli and rewards are causally related.
- Assumption II: the proximate cause of rewards are consistently associated stimuli.
- Assumption III: association is predictive and therefore justifies expectations.

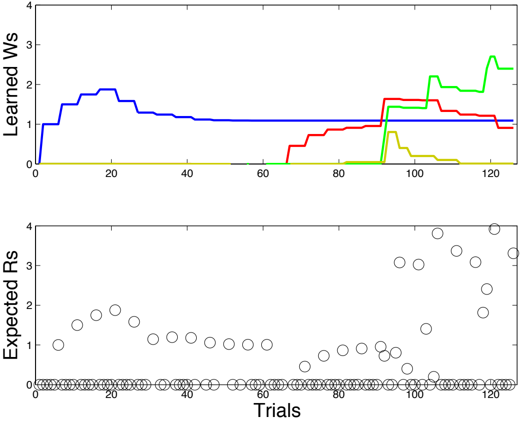


Objective: which stimuli (colors) cause which amount of reward?

Quantitative formulation

Postulate **internal** state variables **w(t)** for reward expectation:

$$\begin{array}{ll} \textit{Specific expectation (top)} & \boldsymbol{w}(t) = [w_1, w_2, w_3, w_4] \\ \textit{Total expectation (bottom)} & v = \boldsymbol{w}(t) \cdot \boldsymbol{u}(t) = w_1 u_1 + w_2 u_2 + \dots \end{array}$$



Specfic reward expectations formed by **Rescorla-Wagner rule**

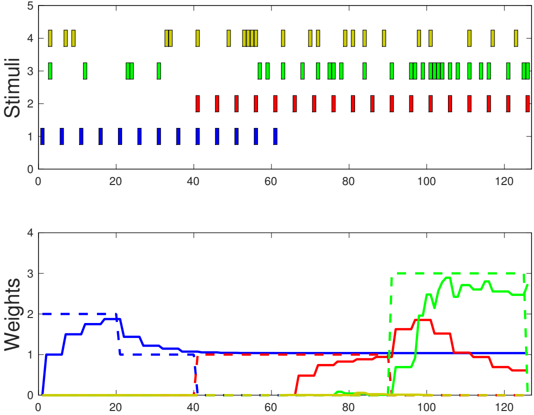
$$\boldsymbol{w} \rightarrow \boldsymbol{w} + \epsilon\,\delta\,\boldsymbol{u}, \qquad \delta = r - v$$

where ϵ is **learning rate** and δ is **prediction error**. For example, $\boldsymbol{u} = [0, 0, 0, 1]$:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_4 \end{pmatrix} \rightarrow \begin{pmatrix} w_1 \\ \vdots \\ w_4 \end{pmatrix} + \epsilon\,\delta \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_4 + \epsilon\,\delta \end{pmatrix}$$

Any prediction error δ (positive or negative) is attributed to *any and all* stimuli present, increasing or decreasing specific expectations. Thus, any prediction error is assumed to have a reason! There is no room for accidents!

Stimuli – putative causes of rewards.



Conditional rewards – expected (solid) and true (dashed).

Note that expectation for blue persists, without evidence!

Points to note

- We hypothesize that animals (and humans) apply a causal heuristic to experiences of reward or punishment.
- Stimulus events associated with reinforcing events are considered as putative causes.
- For each stimulus, a *specific reward expectation* is learned and used to predict future rewards.
- This offers a biological basis for (and explanation of) superstition!

Further points

The RW rule of reinforcement learning postulates:

- For each stimulus, a **specific reward expectation** is learned.
- The sum of these contributions forms the **total reward expectation**.
- The difference between actual and expected total reward is the **prediction error** and drives incremental learning.
- The RW captures many forms of biological learning (including superstition!).

3 Successes for single stimuli

To concretize these ideas, we would like to consider detailed examples. But first we derive a steady-state condition for RW learning. Recall the RW steps:

- A binary value u represents the presence or absence of the stimulus.
- A w is learned, representing the reward contribution associated with the stimulus.
- On any given trial, the total expected reward v and prediction error δ are

$$v = w\,u \qquad\qquad \delta = r - v$$

- Learning is governed by the RW rule

$$w \rightarrow w + \epsilon\,\delta\,u$$

A steady-state is reached when the **average** prediction error vanishes (average $\langle\rangle$ over many trials):

$$0 = \langle \delta u \rangle = \langle (r - v)\, u \rangle = \langle r\, u \rangle - \langle v\, u \rangle = \langle r\, u \rangle - \langle w_{ss}\, u^2 \rangle$$

Solving for the steady-state weight:

$$\langle w_{ss}\, u^2 \rangle = w_{ss}\, \langle u^2 \rangle = \langle r\, u \rangle$$
$$w_{ss} = \frac{\langle r\, u \rangle}{\langle u^2 \rangle}$$

For binary stimuli $u = u^2 \in \{0, 1\}$, $\langle u \rangle = \langle u^2 \rangle$ is the marginal stimulus probability, while $\langle r\, u \rangle$ is the joint expectation of stimulus and reward:

$$w_{ss} = \frac{\langle r\, u \rangle}{\langle u \rangle} \equiv \langle r | u \rangle$$

Thus, w_{ss} is the conditional expectation of reward, given stimulus.

A. Complete reinforcement

Stimulus probability $1/4$, consistently associated with reward 1:

$$u, r \in \{0, 1\} \qquad \langle u \rangle = \langle u^2 \rangle = \langle r \rangle = \langle r\, u \rangle = 1/4$$

RW rule:

$$w_{(i+1)} = w_i + \epsilon\, \delta_i\, u_i, \qquad \delta_i = r_i - v_i, \qquad v_i = w_i\, u_i$$

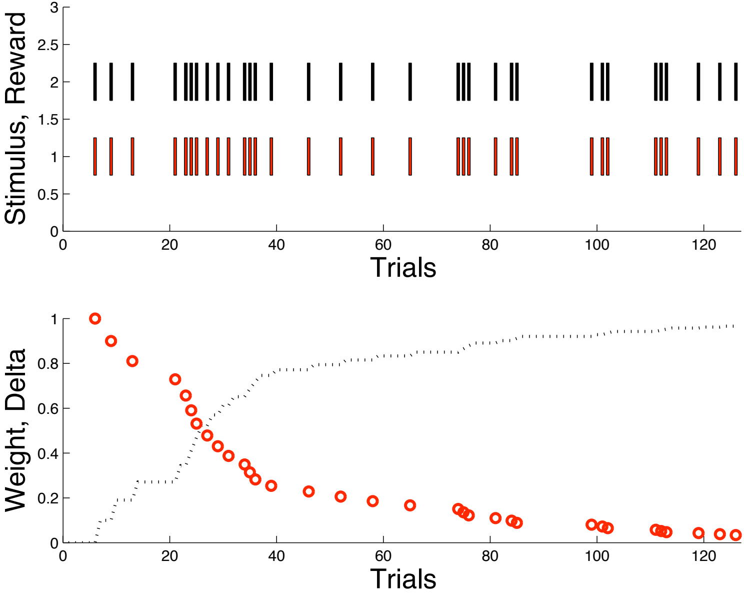
Asymptotic weight:

$$0 = \langle \delta\, u \rangle = \langle (r - v)\, u \rangle = \langle r\, u - w\, u^2 \rangle = \langle r\, u \rangle - w \langle u^2 \rangle$$

$$w \to \frac{\langle r\, u \rangle}{\langle u^2 \rangle} = 1$$

(conditional probability of reward, given stimulus)

Complete reinforcement: $w \to 1$



B. Partial reinforcement

Stimulus probability ${}^1\!/_4$, reward association ${}^3\!/_4$ (conditional probability of reward 1, given stimulus):

$$u, r \in \{0, 1\}\,, \qquad \langle u \rangle = \langle u^2 \rangle = {}^1\!/_4$$

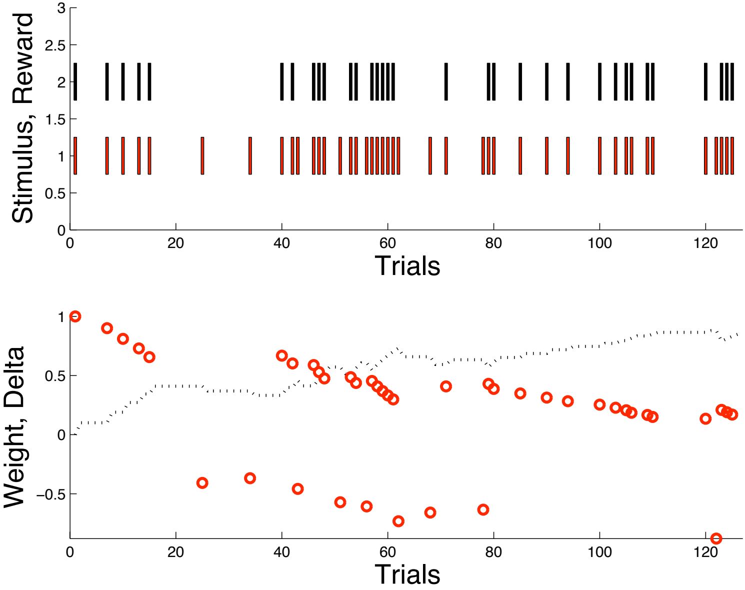
$$\langle r \rangle = \langle r\, u \rangle = {}^1\!/_4 \times {}^3\!/_4 = {}^3\!/_{16}$$

Asymptotic weight:

$$w \rightarrow \frac{\langle r\, u \rangle}{\langle u^2 \rangle} = {}^3\!/_4$$

Next slide: RW reinforcement is unable to expect variable rewards, resulting in prediction errors (+ or - red circles) on every trial.

Partial reinforcement: $w \rightarrow {}^{3}\!/_{4}$



3. Extinction

Stimulus probability $1/4$, reward association changes from $3/4$ to 0:

$$u, r \in \{0, 1\}, \qquad \langle u \rangle = \langle u^2 \rangle = 1/4$$

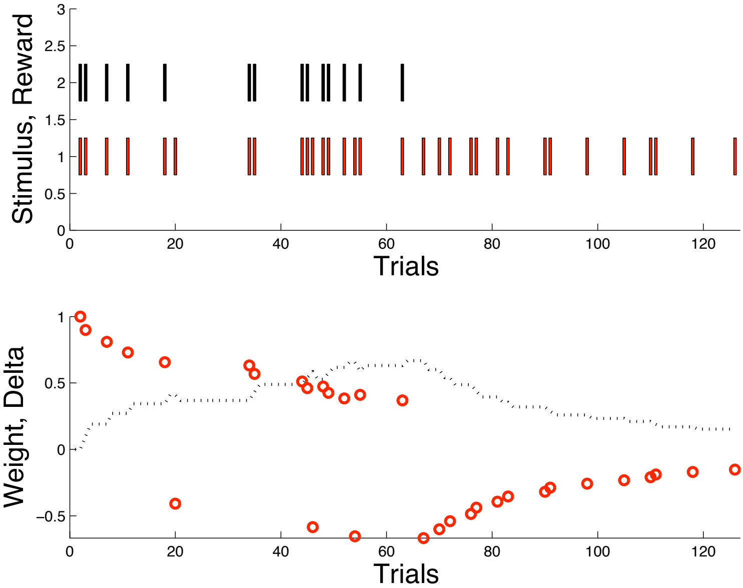
$$\langle r \rangle = \langle r\, u \rangle = 1/4 \times 3/4 \to 0$$

Asymptotic weight:

$$w_\infty = \frac{\langle r\, u \rangle}{\langle u^2 \rangle} = 3/4 \to 0$$

Next slide: repeated disappointment (prediction errors, - red circles) eventually eliminate positive expectation.

Extinction: $w \to {}^3\!/_4 \to 0$



Points to note

The RW rule correctly predicts the outcome of conditioning with a single stimulus:

- In **complete reinforcement**, stimuli are always associated with rewards. The steady-state weight reflects the average reward size.
- In **partial reinforcement**, stimuli are sometime associated with rewards. The steady-state weight reflects average conditional reward size (given presence of a stimulus).
- In **extinction**, expected rewards fail to materialize. The weights gradually decrease to reflect the new situation.

4 Successes for multiple stimuli

Next we want to consider examples with multiple stimuli. But first we derive the steady-state condition for this situation. Recall RW rule:

- A binary-valued vector $\boldsymbol{u}_i$ represents the presence or absence of multiple stimuli at time i.
- Weights $\boldsymbol{w}_i$ are learned separately for each stimulus.
- The expected reward v_i is the dot product

$$v_i = \boldsymbol{w}_i \cdot \boldsymbol{u}_i = \sum_k w_{ik}\, u_{ik}$$

- The RW rule modifies to

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \epsilon\, \delta_i\, \boldsymbol{u}_i, \qquad \delta_i = r_i - v_i$$

ctd

Steady-state is reached when average prediction error vanishes:

$$0 = \langle \delta \boldsymbol{u} \rangle = \langle r \boldsymbol{u} - v \boldsymbol{u} \rangle = \langle r \, \boldsymbol{u} \rangle - \langle \boldsymbol{w}_{ss} \cdot \boldsymbol{u} \, \boldsymbol{u} \rangle$$

$$\langle \boldsymbol{u} \, \boldsymbol{u} \rangle \cdot \boldsymbol{w}_{ss} = \langle r \, \boldsymbol{u} \rangle$$

$$\boldsymbol{w}_{ss} = \langle \boldsymbol{u} \, \boldsymbol{u} \rangle^{-1} \cdot \langle r \, \boldsymbol{u} \rangle = \frac{\langle r \, \boldsymbol{u} \rangle}{\langle \boldsymbol{u} \, \boldsymbol{u} \rangle}$$

Asymptotic weights:

$$\boldsymbol{w}_{ss} = \boldsymbol{Q}^{-1} \cdot \langle r \, \boldsymbol{u} \rangle, \qquad \qquad \boldsymbol{Q} = \langle \boldsymbol{u} \, \boldsymbol{u} \rangle$$

where $\boldsymbol{Q}^{-1}$ is the inverse of the autocorrelation matrix $\boldsymbol{Q}$ (which exists when elements of $\boldsymbol{u}$ are sufficiently decorrrelated).

A. Complete reinforcement

Independent stimulus probabilities $^1/_4$ and $^1/_{10}$, reward associations 1 and $^1/_4$ (i.e., u_2 does not predict any reward over and above u_1):

$$\langle \boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 \\ ^1/_{10} \end{pmatrix} \qquad \langle r | \boldsymbol{u} \rangle = \begin{pmatrix} 1 \\ ^1/_4 \end{pmatrix} \qquad \langle r\, \boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 \\ ^1/_{40} \end{pmatrix}$$

$$\langle \boldsymbol{u}\boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 & ^1/_{40} \\ ^1/_{40} & ^1/_{10} \end{pmatrix} \qquad \langle r\, \boldsymbol{u}\boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 & ^1/_{40} \\ ^1/_{40} & ^1/_{40} \end{pmatrix}$$

Asymptotic weights:

$$\boldsymbol{w}_{ss} = \langle \boldsymbol{u}\boldsymbol{u} \rangle^{-1} \cdot \langle r\, \boldsymbol{u} \rangle = {}^{40}/_{39} \begin{pmatrix} 4 & -1 \\ -1 & 10 \end{pmatrix} \cdot \begin{pmatrix} ^1/_4 \\ ^1/_{40} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

In general, if stimuli are statistically independent, RW learning reveals the correct reward association (ignoring spurious assoc.)!

Details:

Independent stimulus probabilities $^1\!/_4$ and $^1\!/_{10}$, reward associations 1 and $^1\!/_4$:

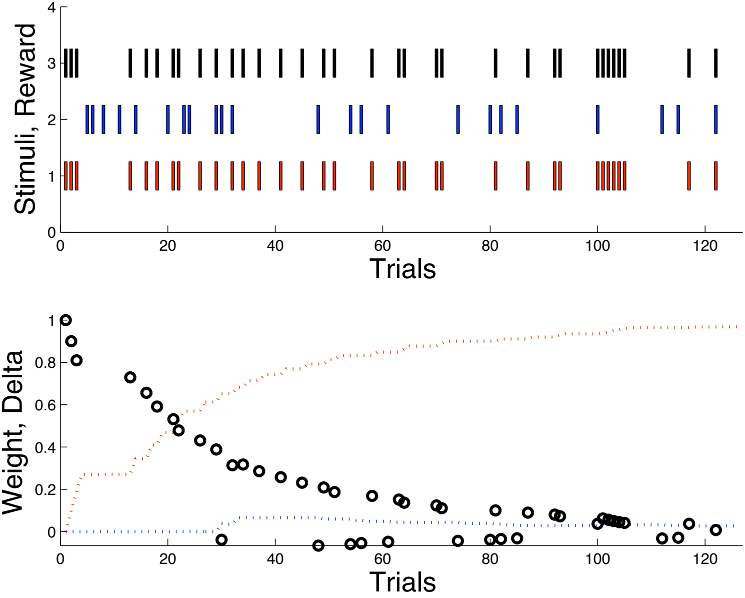
$$P_{1X} = {}^1\!/_4, \quad P_{X1} = {}^1\!/_{10}$$

$$P_{11} = {}^1\!/_{40}, \quad P_{10} = {}^9\!/_{40}, \quad P_{01} = {}^3\!/_{40}, \quad P_{00} = {}^{27}\!/_{40}$$

$$\langle r \rangle = 1 \cdot P_{11} + 1 \cdot P_{10} + 0 \cdot P_{01} + 0 \cdot P_{00} = {}^1\!/_4$$

$$\langle r\, \boldsymbol{u} \rangle = \begin{pmatrix} 1 \cdot P_{1X} \\ {}^1\!/_4 \cdot P_{X1} \end{pmatrix} = \begin{pmatrix} {}^1\!/_4 \\ {}^1\!/_{40} \end{pmatrix}$$

Complete reinforcement: $w_{1,2} \to 1, 0$



B. Partial reinforcement

Independent stimulus probabilities $^1/_4$ and $^1/_{10}$, reward associations $^3/_4$ and $^3/_{16}$ (i.e., u_2 does not predict reward over and above u_1):

$$\langle \boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 \\ ^1/_{10} \end{pmatrix} \qquad \langle r | \boldsymbol{u} \rangle = \begin{pmatrix} ^3/_4 \\ ^3/_{16} \end{pmatrix} \qquad \langle r\, \boldsymbol{u} \rangle = \begin{pmatrix} ^3/_{16} \\ ^3/_{160} \end{pmatrix}$$

$$\langle r\, \boldsymbol{u}\boldsymbol{u} \rangle = \begin{pmatrix} ^3/_{16} & ^3/_{160} \\ ^3/_{160} & ^3/_{160} \end{pmatrix} \qquad \langle \boldsymbol{u}\boldsymbol{u} \rangle = \begin{pmatrix} ^1/_4 & ^1/_{40} \\ ^1/_{40} & ^1/_{10} \end{pmatrix}$$

Asymptotic weights:

$$\boldsymbol{w}_{ss} = \langle \boldsymbol{u}\boldsymbol{u} \rangle^{-1} \cdot \langle r\, \boldsymbol{u} \rangle = {}^{40}/_{39} \begin{pmatrix} 4 & -1 \\ -1 & 10 \end{pmatrix} \cdot \begin{pmatrix} ^3/_{16} \\ ^3/_{160} \end{pmatrix} = \begin{pmatrix} ^3/_4 \\ 0 \end{pmatrix}$$

Again RW learning reveals the correct reward associations!

Details:

Independent stimulus probabilities $^1/_4$ and $^1/_{10}$, reward associations $^3/_4$ and $^3/_{16}$:

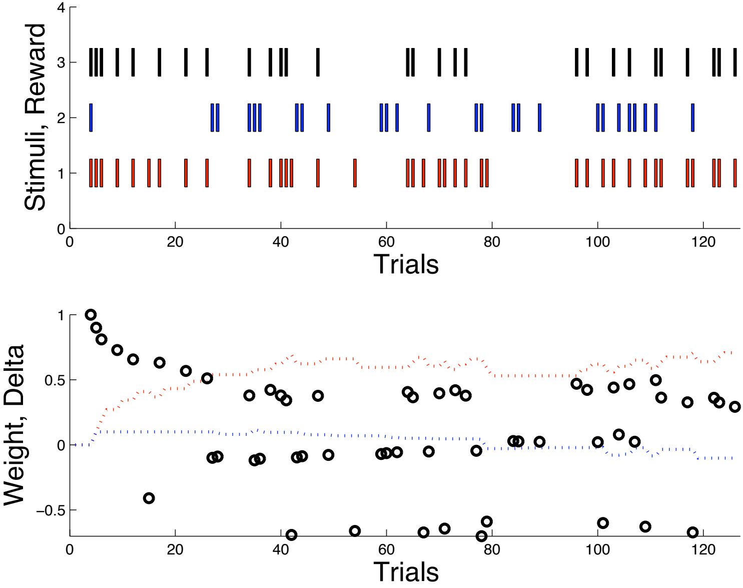
$$P_{1X} = {}^1/_4, \quad P_{X1} = {}^1/_{10}$$

$$P_{11} = {}^1/_{40}, \quad P_{10} = {}^9/_{40}, \quad P_{01} = {}^3/_{40}, \quad P_{00} = {}^{27}/_{40}$$

$$\langle r \rangle = {}^3/_4 \cdot P_{11} + {}^3/_4 \cdot P_{10} + 0 \cdot P_{01} + 0 \cdot P_{00} = {}^3/_{16}$$

$$\langle r\, \boldsymbol{u} \rangle = \begin{pmatrix} {}^3/_4 \cdot P_{1X} \\ {}^3/_{16} \cdot P_{X1} \end{pmatrix} = \begin{pmatrix} {}^3/_{16} \\ {}^3/_{160} \end{pmatrix}$$

Partial reinforcement: $w_{1,2} \rightarrow 3/4, 0$



C. Blocking

First step "conditioning": Stimulus probabilities $^1/_4$ and 0, reward associations $^3/_4$ and 0, respectively:

$$\boldsymbol{w}_{ss} = \begin{pmatrix} ^3/_4 \\ 0 \end{pmatrix}$$

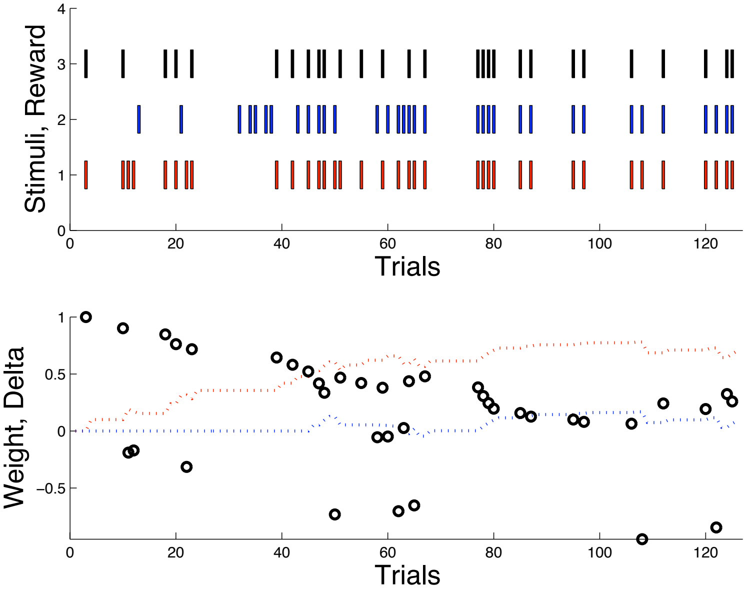
Second step "pairing": *Dependent* stimulus probabilities $^1/_4$ and $^1/_4$, reward association $^3/_4$ (for each stimulus):

$$\langle r\,\boldsymbol{u}\rangle = \begin{pmatrix} ^3/_{16} & ^3/_{16} \\ ^3/_{16} & ^3/_{16} \end{pmatrix} \qquad \langle \boldsymbol{u}\,\boldsymbol{u}\rangle = \begin{pmatrix} ^1/_4 & ^1/_4 \\ ^1/_4 & ^1/_4 \end{pmatrix} \qquad \langle \boldsymbol{u}\,\boldsymbol{u}\rangle^{-1} = n.d.$$

$$\boldsymbol{w}_{ss} = n.d.$$

Note: RW learning cannot distinguish between correlated stimuli!

Blocking: $w_{1,2} \to 3/4, 0 \to 3/4, 0$



D. Inhibitory conditioning

Independent stimulus probabilities $^1\!/_2$ and $^1\!/_2$, reward association 1 with $\boldsymbol{u} = [1, 0]$, and 0 with all other stimulus combinations (i.e., u_2 predicts failure of reward!):

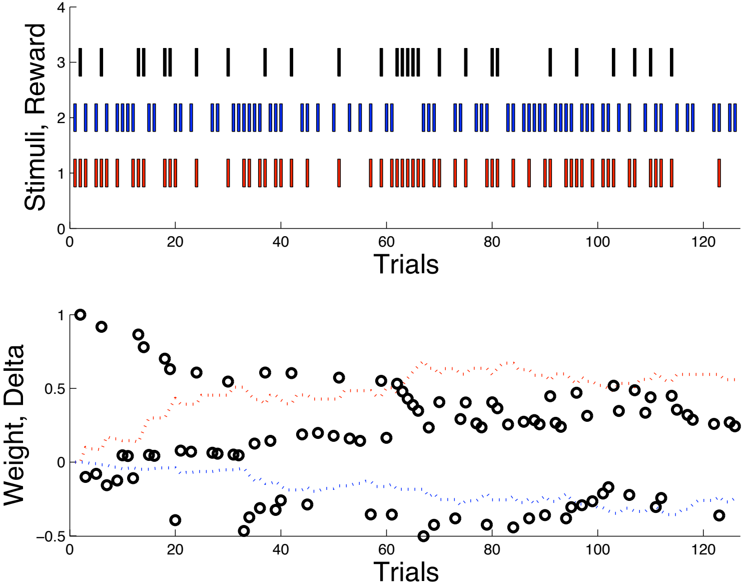
$$\langle r\,\boldsymbol{u}\rangle = \begin{pmatrix} ^1\!/_4 \\ 0 \end{pmatrix} \qquad \langle r\,\boldsymbol{u}\,\boldsymbol{u}\rangle = \begin{pmatrix} ^1\!/_4 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\langle \boldsymbol{u}\,\boldsymbol{u}\rangle = \begin{pmatrix} ^1\!/_2 & ^1\!/_4 \\ ^1\!/_4 & ^1\!/_2 \end{pmatrix}, \quad \langle \boldsymbol{u}\,\boldsymbol{u}\rangle^{-1} = \begin{pmatrix} ^8\!/_3 & -^4\!/_3 \\ -^4\!/_3 & ^8\!/_3 \end{pmatrix}$$

Asymptotic weight:

$$\boldsymbol{w}_{ss} = \begin{pmatrix} ^8\!/_3 & -^4\!/_3 \\ -^4\!/_3 & ^8\!/_3 \end{pmatrix} \cdot \begin{pmatrix} ^1\!/_4 \\ 0 \end{pmatrix} = \begin{pmatrix} ^2\!/_3 \\ -\,^1\!/_3 \end{pmatrix}$$

Inhibitory conditioning: $w_{1,2} \rightarrow {}^{2}\!/_{3}, -{}^{1}\!/_{3}$



E. Overshadowing

Independent stimulus probabilities $^1\!/_4$, independent reward association of $^3\!/_4$, but no double reward for $\boldsymbol{u} = [1, 1]$:

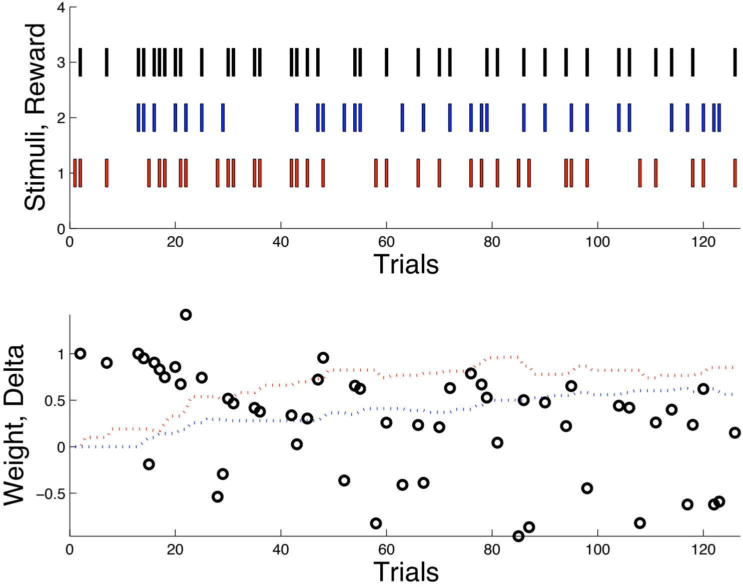
$$\langle r\,\boldsymbol{u}\rangle = \begin{bmatrix} ^3\!/_{16} \\ ^3\!/_{16} \end{bmatrix} \qquad \langle r\,\boldsymbol{u}\,\boldsymbol{u}\rangle = \begin{bmatrix} ^3\!/_{16} & ^3\!/_{64} \\ ^3\!/_{64} & ^3\!/_{16} \end{bmatrix}$$

$$\langle \boldsymbol{u}\,\boldsymbol{u}\rangle = \begin{bmatrix} ^1\!/_4 & ^1\!/_{16} \\ ^1\!/_{16} & ^1\!/_4 \end{bmatrix} \qquad \langle \boldsymbol{u}\,\boldsymbol{u}\rangle^{-1} = \begin{bmatrix} ^{64}\!/_{15} & ^{-16}\!/_{15} \\ ^{-16}\!/_{15} & ^{64}\!/_{15} \end{bmatrix}$$

Asymptotic weights:

$$\boldsymbol{w}_{ss} = \begin{bmatrix} ^{64}\!/_{15} & ^{-16}\!/_{15} \\ ^{-16}\!/_{15} & ^{64}\!/_{15} \end{bmatrix} \cdot \begin{bmatrix} ^3\!/_{16} \\ ^3\!/_{16} \end{bmatrix} = \begin{bmatrix} ^3\!/_5 \\ ^3\!/_5 \end{bmatrix}$$

Overshadowing: $w_{1,2} \rightarrow {}^{3}\!/_{5}$ instead of ${}^{3}\!/_{4}$



Points to note

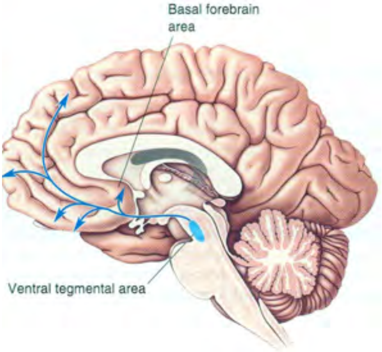
The RW rule correctly predicts the outcome of conditioning with multiple stimuli:

- In **complete** or **partial** reinforcement, the steady-state weights reflect average conditional reward, given the presence of a stimulus.
- In **blocking**, a previously learned reward expectation for an old stimulus blocks the formation of an association for a new stimulus which is paired with the old one.
- In **inhibitory conditioning**, negative reward expectations form for any stimulus consistently associated with the failure of an expected reward.
- In **overshadowing**, multiple reward-associated stimuli form smaller reward expectations than a single such stimulus (as reinforcement disappoints for multiple stimuli).

5 Dopamine and reward

The neural substrate for the prediction error δ is thought to involve dopaminergic activity in the midbrain (ventral tegmental area VTA and substantia nigra).

Its axons project topographically to striatum (caudate nucleus and putamen), ventral striatum (including nucleus accumbens), and most areas of neocortex, including prefrontal cortex. Conditioning is thought to modify cortico-striatal loops.

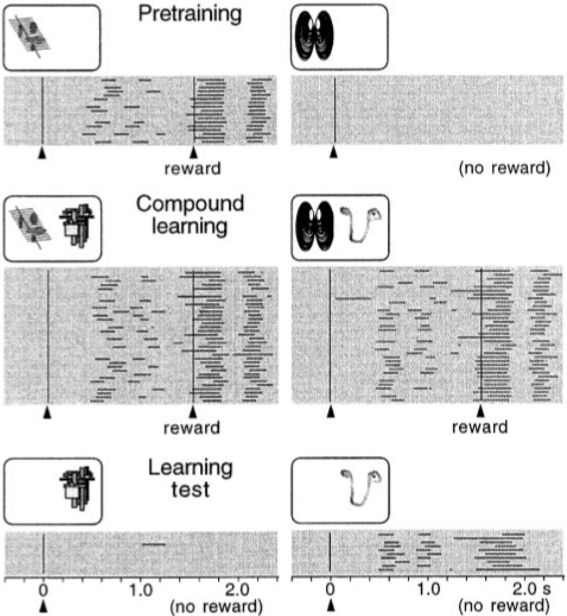


Phasic activation midbrain dopamine neurons (DNs)

- 75% show phasic activation to primary rewarding stimuli (different foods, liquids).
- 70% show phasic activation to onset of reward-predicting (conditioned) visual or auditory stimuli

Next slide: rats lick a water spout when they expect liquid (and also when they drink it). Thus, licking reveals ‘reward expectation’ on the part of the rat. In this three-phase experiment, rewards were paired with different visual patterns during a ‘pretraining’, a ‘compound learning’, and a ‘test’ phase.

Licking behavior during compound learning (blocking)



DN activity during compound learning (blocking)

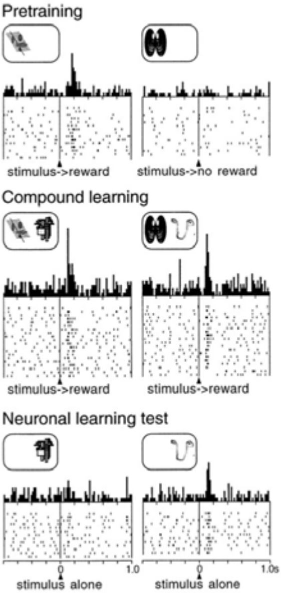
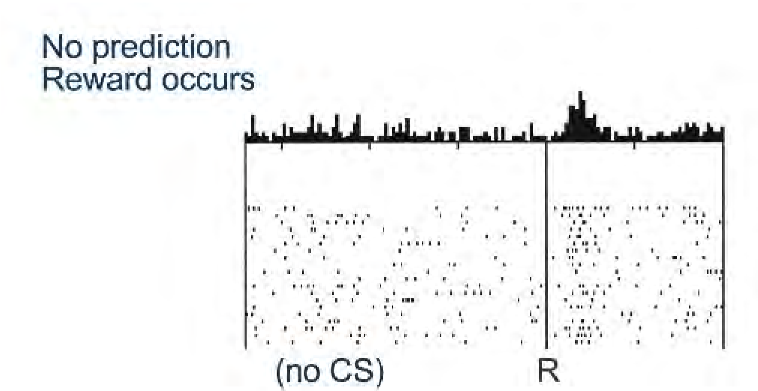


Figure 3. Learning of Dopamine Responses to Conditioned Stimuli in the Blocking Paradigm Depends on Prediction Error Rather than Stimulus-Reward Pairing Alone

(Top) During pretraining, differential activation follows reward-predicting stimulus but not unrewarded control stimulus (right). (Middle) After compound learning, activation to reward-predicting compound is maintained (no prediction error), and activation to control stimulus is learned (right) (positive prediction error). (Bottom) Test trials reveal absent (blocked) neuronal response to the added stimulus but learned response to the control stimulus. Dots denote neuronal impulses, referenced in time to the stimuli (arrows). Histograms contain the sums of raster dots. Reprinted with permission from *Nature* (Waelti et al., 2001). Copyright (2001) Macmillan Publishers Ltd.

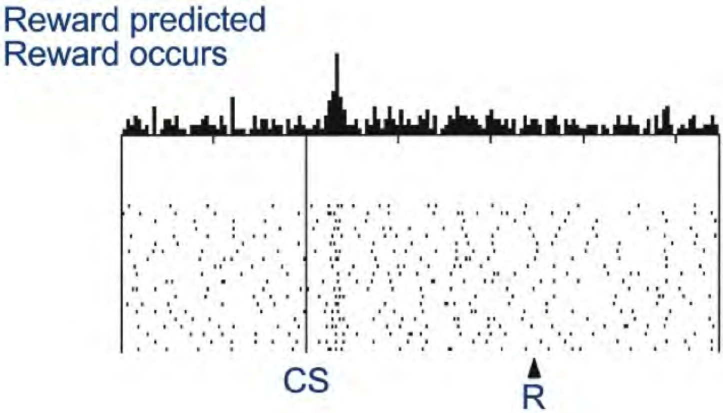
Dopamine neurons report prediction errors (Schultz, 1997)

DN activity with(out) conditioned stimulus (CS) and reward (R).



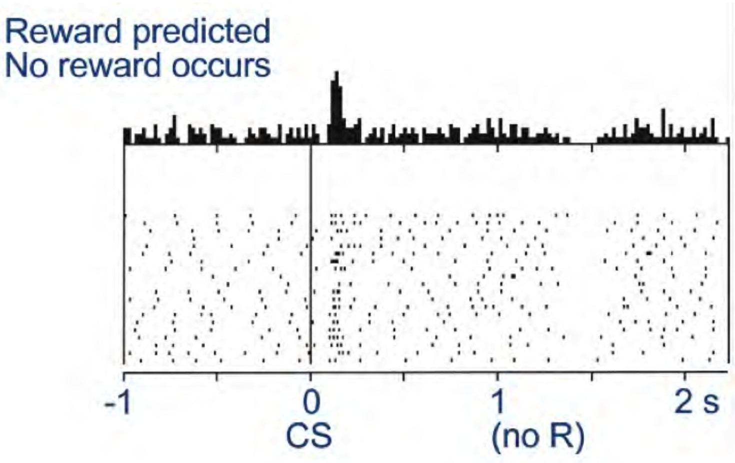
Without the CS, the reward is unexpected. DNs are activated (positive suprise).

DN activity with(out) conditioned stimulus (CS) and reward (R).



With the CS, the reward is expected. DNs are activated by the CS, not the reward (no suprise).

DN activity with(out) conditioned stimulus (CS) and reward (R).



With the CS, a reward is expected by not delivered. DNs are activated by the CS and suppressed after the reward fails to appear (negative surprise).

Points to note

- The activity of dopamine neurons in the midbrain (VTA and SN) is thought to signal *prediction error*.

- Specifically, it seems to signal unexpected reinforcers (positive or negative surprises) rather than expected reinforcers (rewards or punishments).

- This activity could potentially provide a ‘teacher signal’ and enable supervised learning in the cortex and striatum.

- The significance of dopamine activity remains a highly active research area!

Summary classical conditioning

RW rule, binary stimulus patterns, and linear reward predictions are obviously gross oversimplifications. Nevertheless, they summarize and unify an impressive variety of classical conditioning evidence.

- Pavlovian conditioning.
- Extinction
- Partial conditioning
- Blocking
- Overshadowing
- ~~Trace conditioning~~
- ~~Secondary conditioning~~

The RW rule works well when when rewards are delivered together with, or immediately after, stimuli. If rewards are delayed (‘trace conditioning’, ‘secondary conditioning’) the RW rule fails.

Next: Delayed rewards