Adaptation and learning

BME665/565

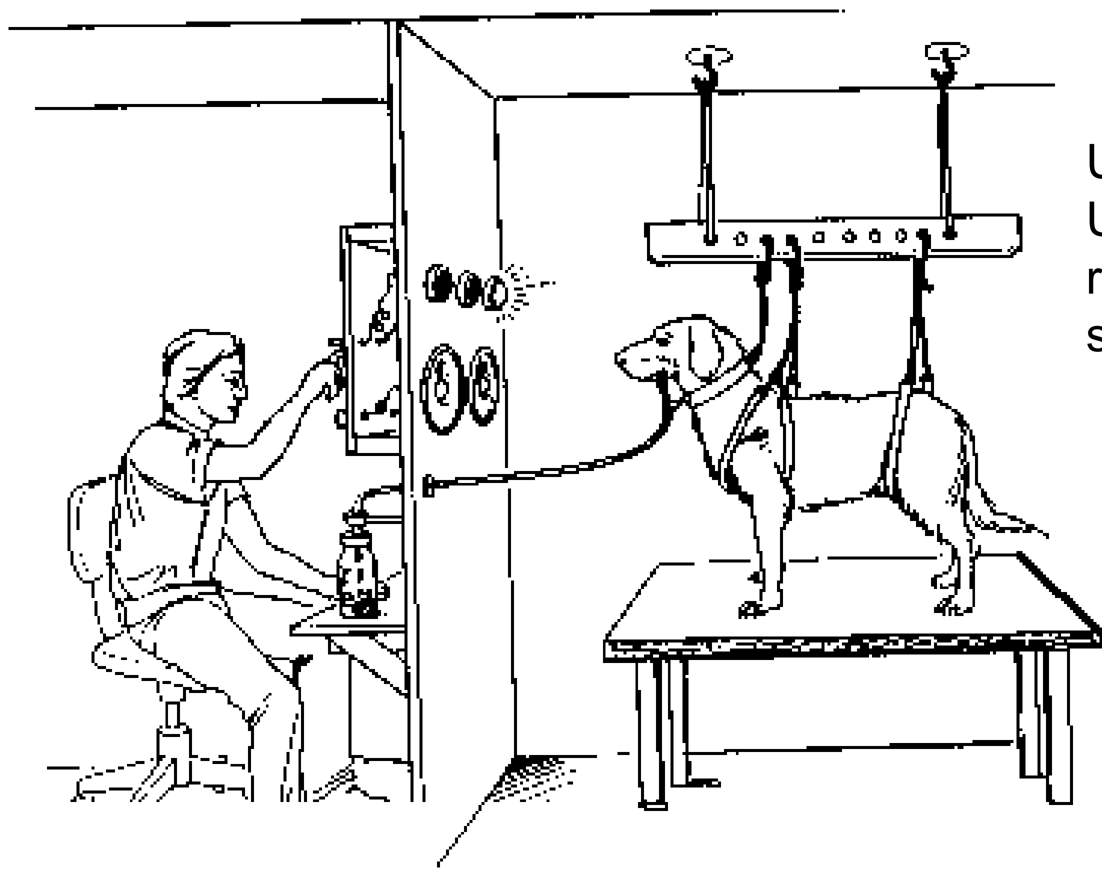
Learning

- Classical conditioning
  - Circuits involved in fear conditioning
  - Role of dopamine in conditioning
- Cortical organization

Phenomenology

UCS:
Unconditioned
stimulus: food

UCR:
Unconditioned
response:
salivation



CS:
Conditioned
stimulus: bell

CR:
Conditioned
response:
salivation

Pavlovian experimental apparatus

Uncontrolled cues are reduced by the wall between dog and experimenter. Saliva drips through the tube into the measuring bottle on the experimenter's side of the wall.

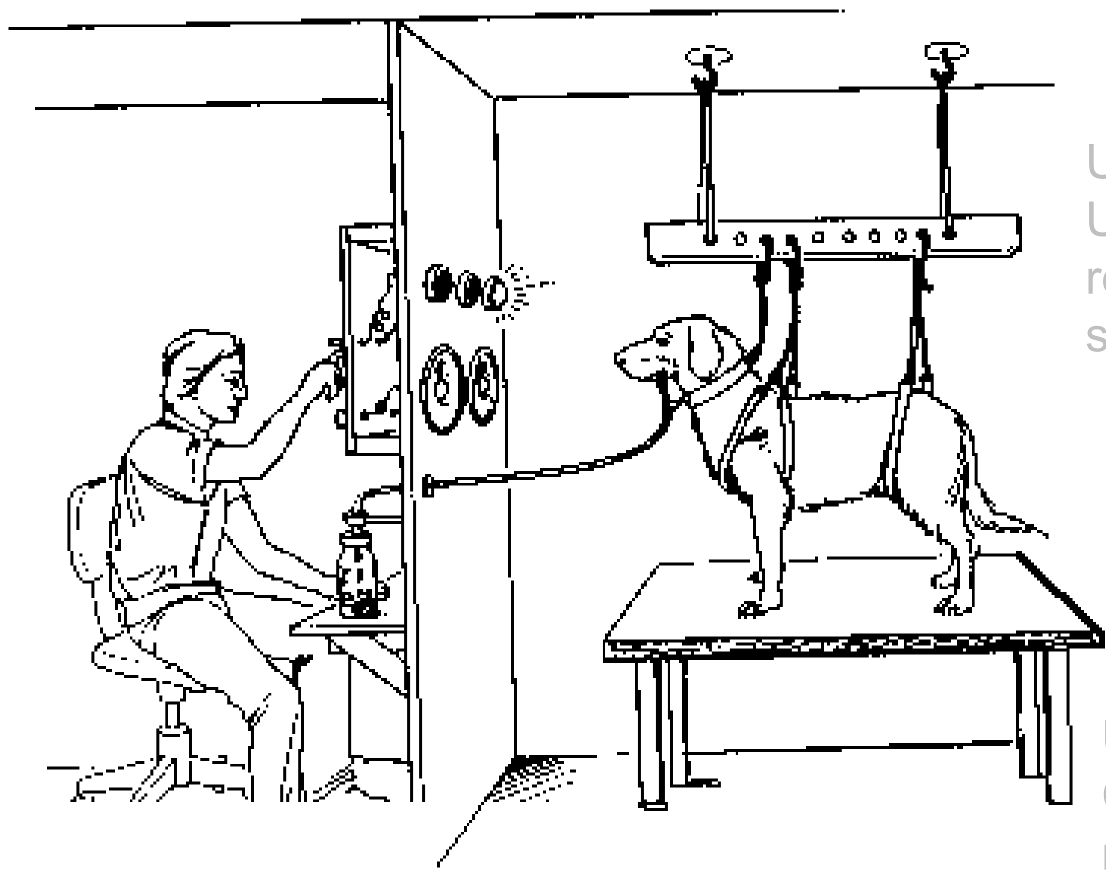
Phenomenology: Reinforcement Learning

UCS: Unconditioned stimulus: food

Reward: food

Stimulus: bell

CS: Conditioned stimulus: bell



UCR: Unconditioned response: salivation

Expectation: salivation

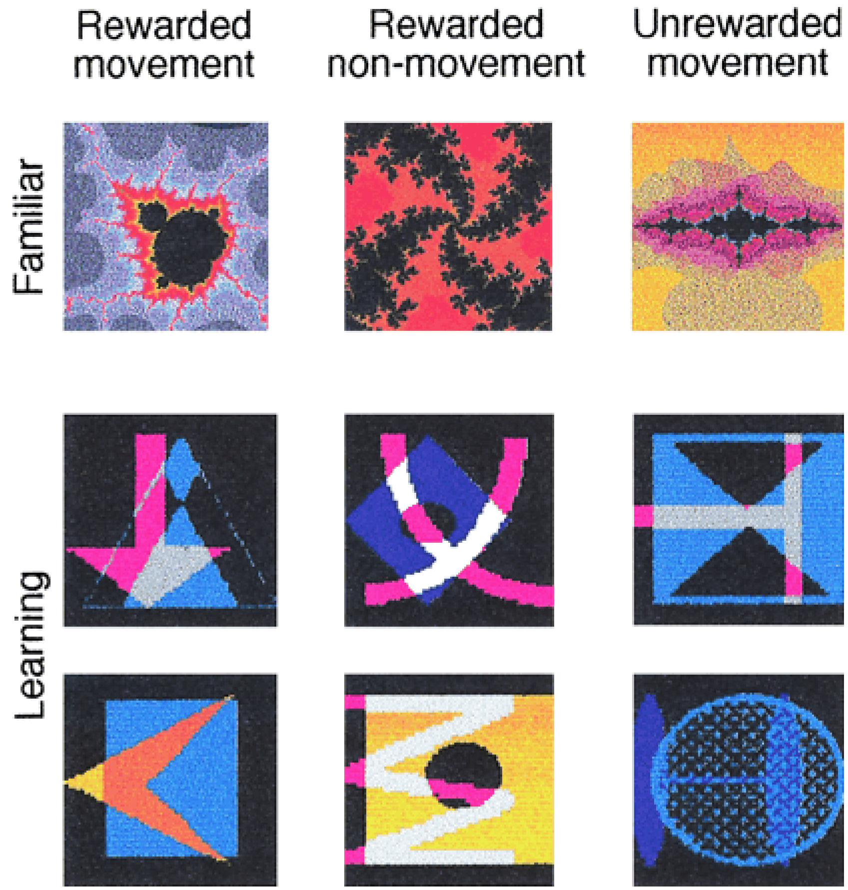
UCR: Conditioned response: salivation

Pavlovian experimental apparatus

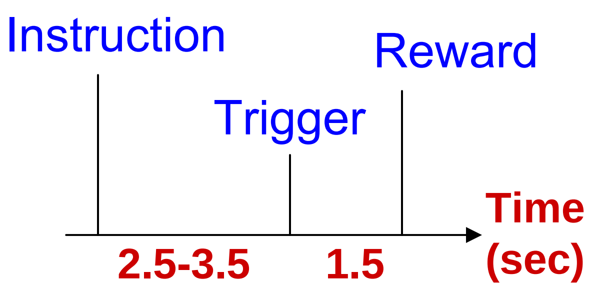
Uncontrolled cues are reduced by the wall between dog and experimenter. Saliva drips through the tube into the measuring bottle on the experimenter's side of the wall.

Reinforcement learning

Apple juice reinforcement on correct behavior

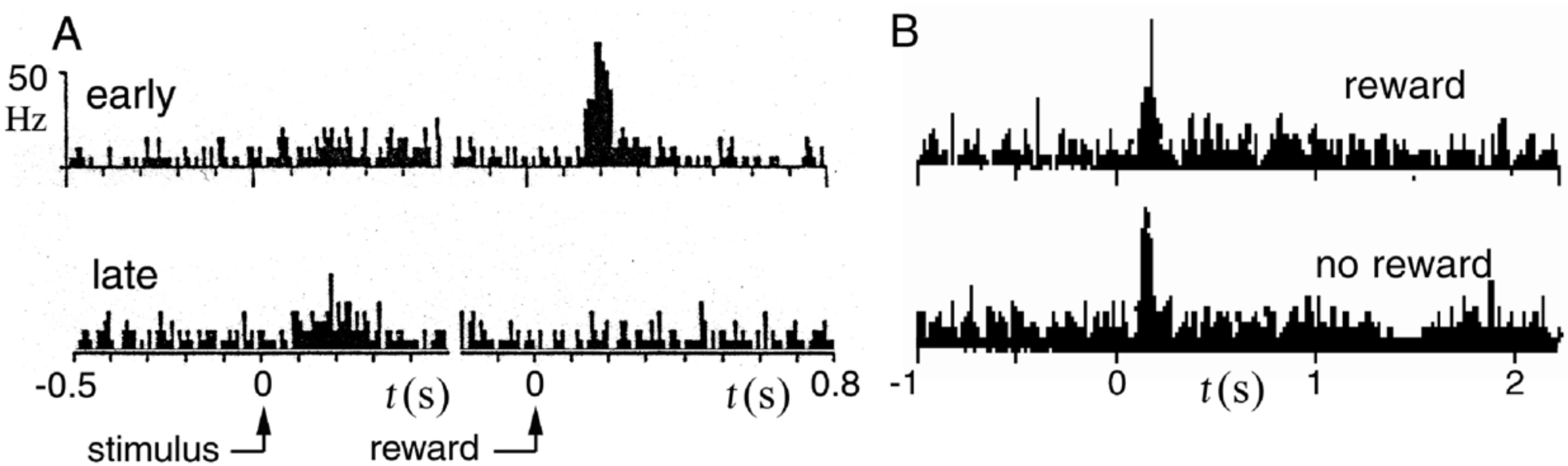


Auditory reinforcement on correct behavior



Tremblay et. al., J Neurophysiol 80: 964-977, 1998;

Reinforcement learning (cont.)



Activity of DA neurons in VTA

Left: neuronal response to reward tied to stimulus

Right: neuronal response of trained neuron when reward is withheld

Variations of Conditioning 1

Extinction:

- Stimulus (bell) repeatedly shown without reward (food):

  conditioned response (salivating) reduced

Partial reinforcement:

- Stimulus only sometimes preceding reward:

  conditioned response weaker than in classical case

Blocking (2 stimuli):

- First: stimulus S1 associated with reward: classical conditioning.
- Then: stimulus S1 and S2 shown together followed by reward:

  Association between S2 and reward not learned

Variations of Conditioning 2

- Inhibitory Conditioning (2 stimuli):
  - Alternate 2 types of trials:

  1. S1 followed by reward
  2. S1+S2 followed by absence of reward

  Result: S2 becomes predictor of absence of reward

- Overshadowing (2 stimuli):
  - Repeatedly present S1+S2 followed by reward

  Result: often, reward prediction shared unequally between stimuli

- Secondary Conditioning:
  - S1 preceding reward (classical case). Then, S2 preceding S1

  Result: S2 leads to prediction of reward

  But: if S1 following S2 showed too often: extinction will occur

Classical conditioning paradigms

Paradigm	Pre-train	Train	Expectation
Pavlovian		$s \rightarrow r$	$s \rightarrow r$
Extinction	$s \rightarrow r$	s → <none>	s → <none>
Partial		$s \rightarrow r$, s → <none>	$s \rightarrow \alpha r$
Blocking	$s_1 \rightarrow r$	$s_1+s_2 \rightarrow r$	$s_1 \rightarrow r$, s_2→<none>
Inhibitory		s_1+s_2 →<none>, $s_1 \rightarrow r$	$s_1 \rightarrow r$, $s_2 \rightarrow -r$
Overshadow		$s_1+s_2 \rightarrow r$	$s_1 \rightarrow \alpha r$, $s_2 \rightarrow \beta r$
Secondary	$s_1 \rightarrow r$	$s_2 \rightarrow s_1$	$s_2 \rightarrow r$

Rescorla-Wagner Theory (1972)

- Organisms only learn when events violate their expectations
- Expectations are built up when ‘significant’ events follow a stimulus complex
- These expectations are only modified when consequent events disagree with the composite expectation

Rescorla-Wagner Rule

$$v = wu \qquad \Delta w = \varepsilon \delta u \qquad \text{where } \delta = r - v$$

- v: Predicted reward
- u: Stimulus (0 or 1)
- ε: Learning rate
- δ: Prediction error
- r: Actual reward

Rescorla-Wagner Rule

$$v = wu \qquad \Delta w = \varepsilon \delta u \qquad \text{where } \delta = r - v$$

- v: Predicted reward
- u: Stimulus (0 or 1)
- ε: Learning rate
- δ: Prediction error
- r: Actual reward

For multiple stimuli, take *u* as a vector:

$$v = \vec{w} \cdot \vec{u}$$

$$\Delta w = \varepsilon \delta \vec{u} = \varepsilon (r - v) \vec{u}$$

Rescorla-Wagner Rule

No stimulus means no change in weight

When there is a reward, the prediction error will drive w to an equilibrium value $< r >$

$$v = wu \qquad \Delta w = \varepsilon \delta u \quad \text{where } \delta = r - v$$

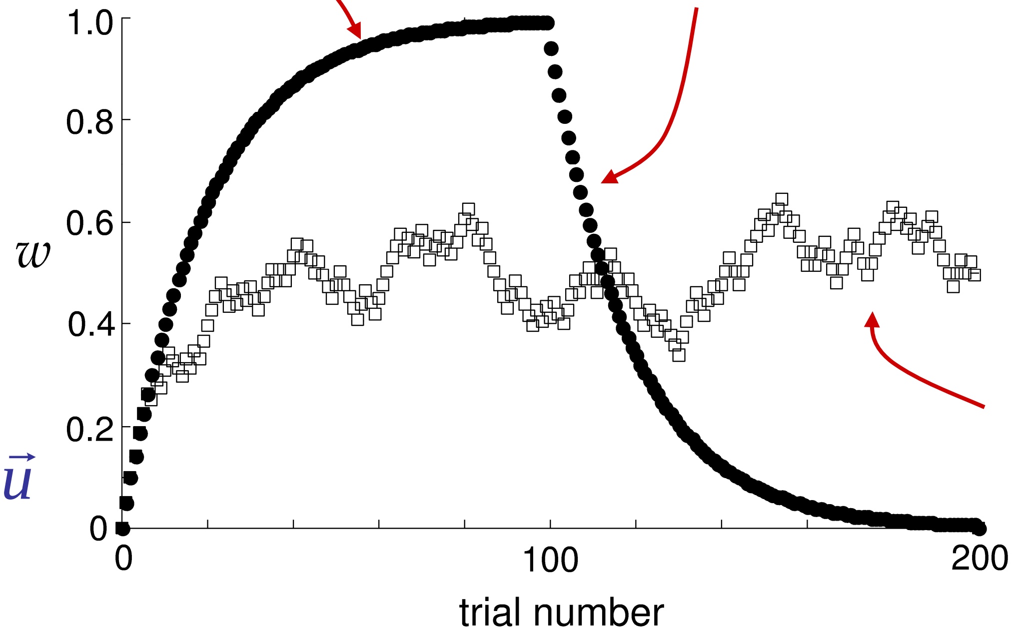
Predicted reward

No reward: reduce weight if there is an expectation of reward

Rescorla-Wagner and classical learning paradigms

Pavlovian:
u=1, r =1 (stimulus and reward present)

Extinction:
u=1, r =0 (stimulus present, no reward)



Partial:
u=1, r =random (stim present, reward variable)

$$\Delta w = \varepsilon (r - v)\vec{u}$$

$$v = wu$$

Paradigm	Pre-train	Train	Expectation
Pavlovian		s → r	s → r
Extinction	s → r	s → <none>	s → <none>
Partial		s → r, s → <none>	s → *α* r

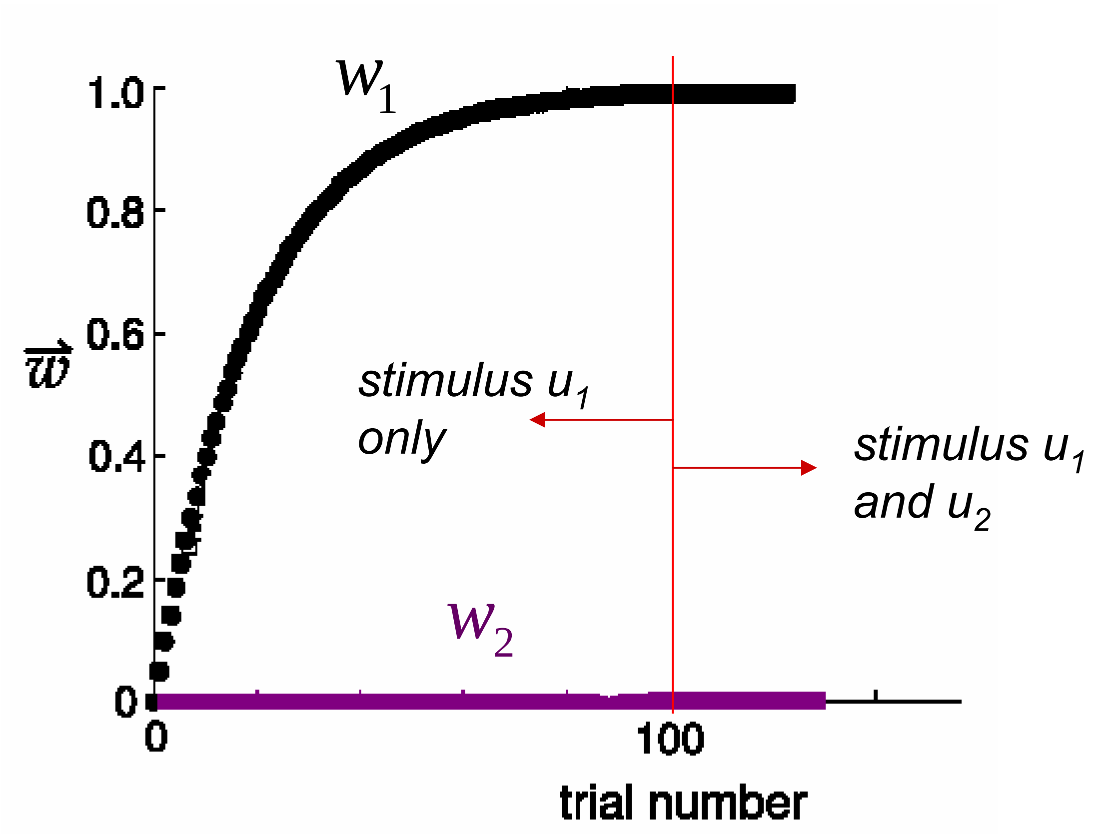
Rescorla-Wagner and classical learning paradigms (cont.)

Paradigm	Pre-train	Train	Expectation
Blocking	$s_1 \rightarrow r$	$s_1+s_2 \rightarrow r$	$s_1 \rightarrow r$, $s_2 \rightarrow$ <none>

Blocking:
Reward present; stimuli change partway through

$$v = w_1 u_1 + w_2 u_2$$

$$\vec{w} \rightarrow \langle r \rangle$$



$$\Delta w = \varepsilon (r - v)\vec{u} \quad v = \vec{w} \cdot \vec{u}$$

Rescorla-Wagner and classical learning paradigms (cont.)

Paradigm	**Pre-train**	**Train**	**Expectation**
Inhibitory		$s_1+s_2 \rightarrow$ <none>, $s_1 \rightarrow r$	$s_1 \rightarrow r$, $s_2 \rightarrow -r$

Inhibition:

Reward present when stimulus 1 is present;

Reward absent when stimulus 1 + stimulus 2 presented together

$$w_1 u_1 \rightarrow \langle r \rangle \qquad w_1 u_1 + w_2 u_2 \rightarrow 0$$

$$w_2 u_2 \rightarrow -w_1 u_1$$

$$\Delta w = \varepsilon(r - v)\vec{u} \qquad v = \vec{w} \cdot \vec{u}$$

Rescorla-Wagner and classical learning paradigms (cont.)

Paradigm	Pre-train	Train	Expectation
Overshadow		$s_1+s_2 \rightarrow r$	$s_1 \rightarrow \alpha r$, $s_2 \rightarrow \beta r$

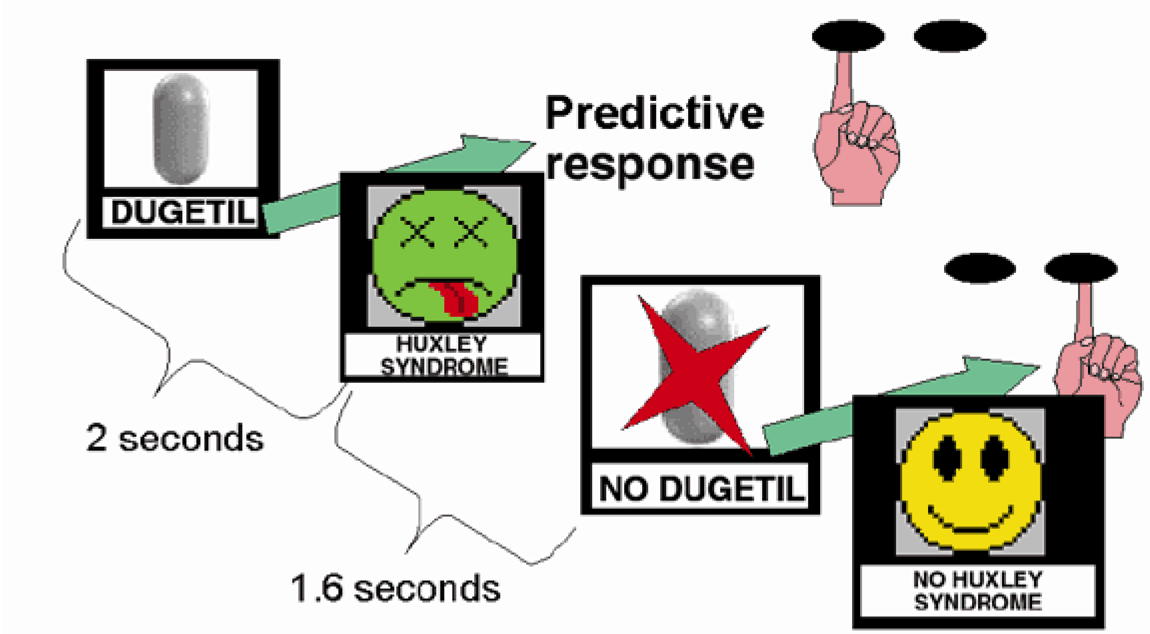
Overshadow:

$v=w_1+w_2$ goes to r, but w_1 and w_2 may become different if there are different learning rates ε_i for them

$$w_1u_1 + w_2u_2 \rightarrow \langle r \rangle$$

$$\Delta w = \varepsilon (r - v)\vec{u} \quad v = \vec{w}\cdot\vec{u}$$

Example: Responses in human DLPFC to surprise events



Subjects learned associations between cues (fictitious drugs) and outcomes (fictitious syndromes)

Fletcher et al., ***Nature Neuroscience*** 4, 1043 - 1048 (2001)

Asymmetry of Rescorla-Wagner rule

$$v = wu \quad \Delta w = \varepsilon \delta \vec{u} = \varepsilon (r - v) \vec{u}$$

$$\varepsilon = f(\alpha \beta)$$

Learning rate associated with the outcome

Learning rate associated with the stimulus

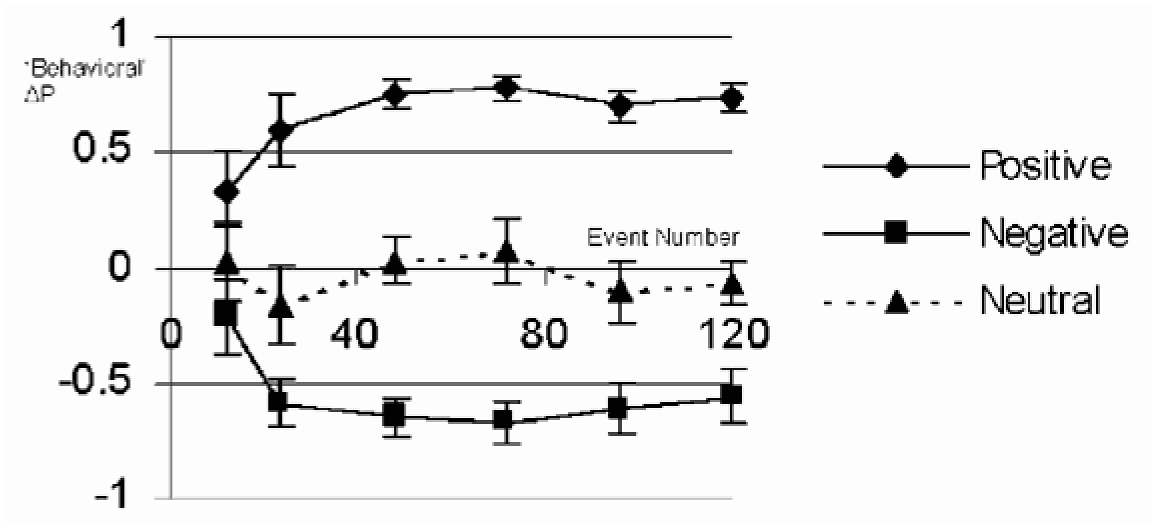
Positive contingency: the presence of 'drug' is a strong predictor of 'syndrome,' a surprise event is 'drug−no syndrome'

For a learned negative contingency (‘no drug’ then ‘syndrome’), 'drug−syndrome' is unexpected

According to the Rescorla−Wagner rule, these two types of unexpectedness should induce different weight adjustments

Fletcher et al., ***Nature Neuroscience*** 4, 1043 - 1048 (2001)

Learned associations

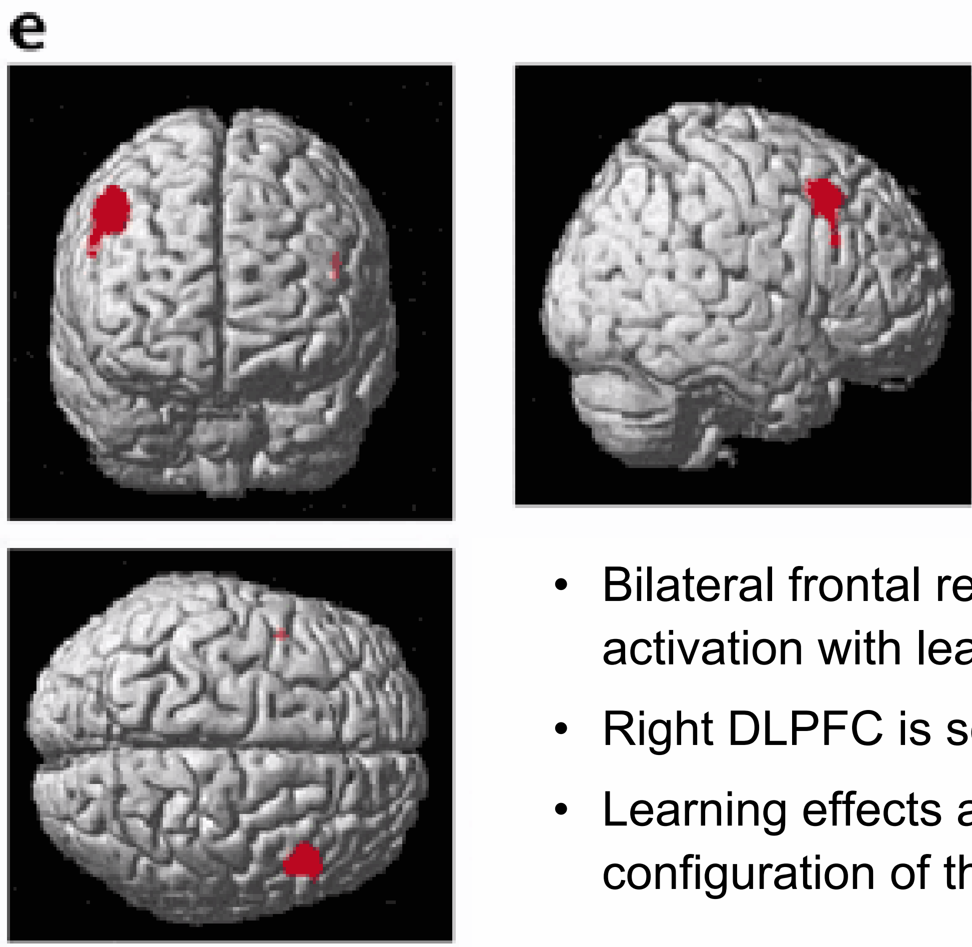


$$\Delta P = P(\text{'syndrome' following 'drug'}) - P(\text{'syndrome' following 'no drug'})$$

Subjects were sensitive to both positive and negative causal relationships, but were more sensitive to positive relationships

Fletcher et al., ***Nature Neuroscience*** 4, 1043 - 1048 (2001)

Neuronal activity patterns reflect the same differences



- Bilateral frontal regions show decreased activation with learning
- Right DLPFC is sensitive to unpredictability
- Learning effects are modulated by the configuration of the surprise event

Fletcher et al., ***Nature Neuroscience*** 4, 1043 - 1048 (2001)

Application of Rescorla-Wagner Rule to fMRI data

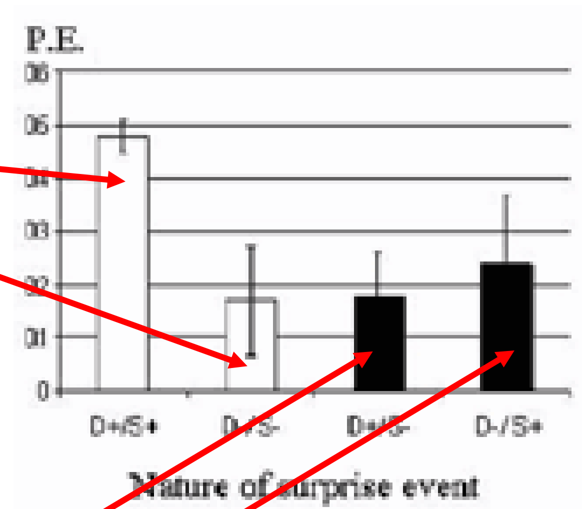
Learned response: (-ve contingency)

Syndrome follows No Drug

Surprise events:

Syndrome follows Drug

No syndrome follows No Drug



Learned response: (+ve contingency)

Syndrome follows Drug

Surprise events:

No Syndrome follows Drug

Syndrome follows No Drug

$$\varepsilon = f(\alpha\beta)$$

Fletcher et al., ***Nature Neuroscience*** 4, 1043 - 1048 (2001)

Predicting future reward : Temporal Difference Learning

- Try to predict the total future reward expected from time *t* onward to the time *T* of end of trial
- Assume time is in discrete steps

$$R(t) = \left\langle \sum_{\tau=0}^{T-t} r(t+\tau) \right\rangle$$

- Predicted total future reward from time t (one stimulus case):

$$v(t) = \sum_{\tau=0}^{t} w(\tau) u(t-\tau)$$

Problem: how to adjust the weight? Would like to adjust *w(τ)* to make *v(t)* approximate the true total future reward *R(t)* (reward that is yet to come) but this is unknown since lying in future

Predicting future reward

- If the time within a trial is taken to be discrete and all variables are functions of time, then *v(t)* can be taken as the expectation of reward later in the trial, and so for a trial of length T:

$$v(t) = \left\langle \sum_{\tau=0}^{T-t} r(t+\tau) \right\rangle$$

- For a single time-dependent stimulus:

$$v(t) = \sum_{\tau=0}^{t} w(\tau) u(t-\tau)$$

$$\Delta w(\tau) = \varepsilon \delta(t) u(t-\tau) \qquad \delta(t) = \sum_{\tau} r(t+\tau) - v(t)$$

Predicting future reward

- Approximate

$$\sum_{\tau=0}^{T-t} r(t+\tau) = r(t) + \sum_{\tau=0}^{T-t-1} r(t+1+\tau)$$

$$\approx r(t) + v(t+1)$$

$$v(t) = \left\langle \sum_{\tau=0}^{T-t} r(t+\tau) \right\rangle$$

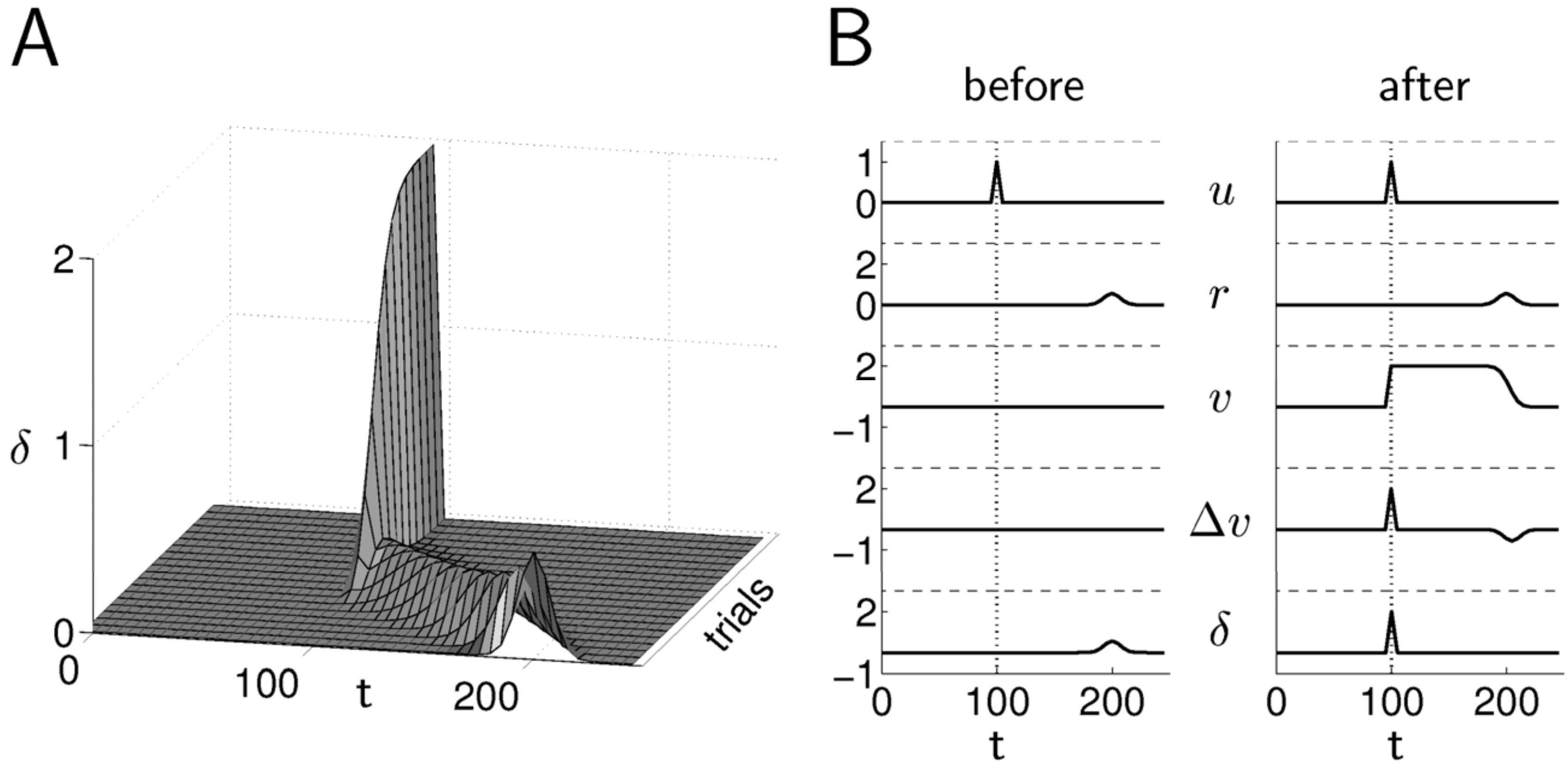
So:

$$\delta(t) = \sum_{\tau} r(t+\tau) - v(t)$$

$$\underbrace{\delta(t)}_{\text{temporal difference error}} = r(t) + \underbrace{v(t+1) - v(t)}_{\text{temporal difference}}$$

Temporal difference learning rule

Temporal difference learning rule



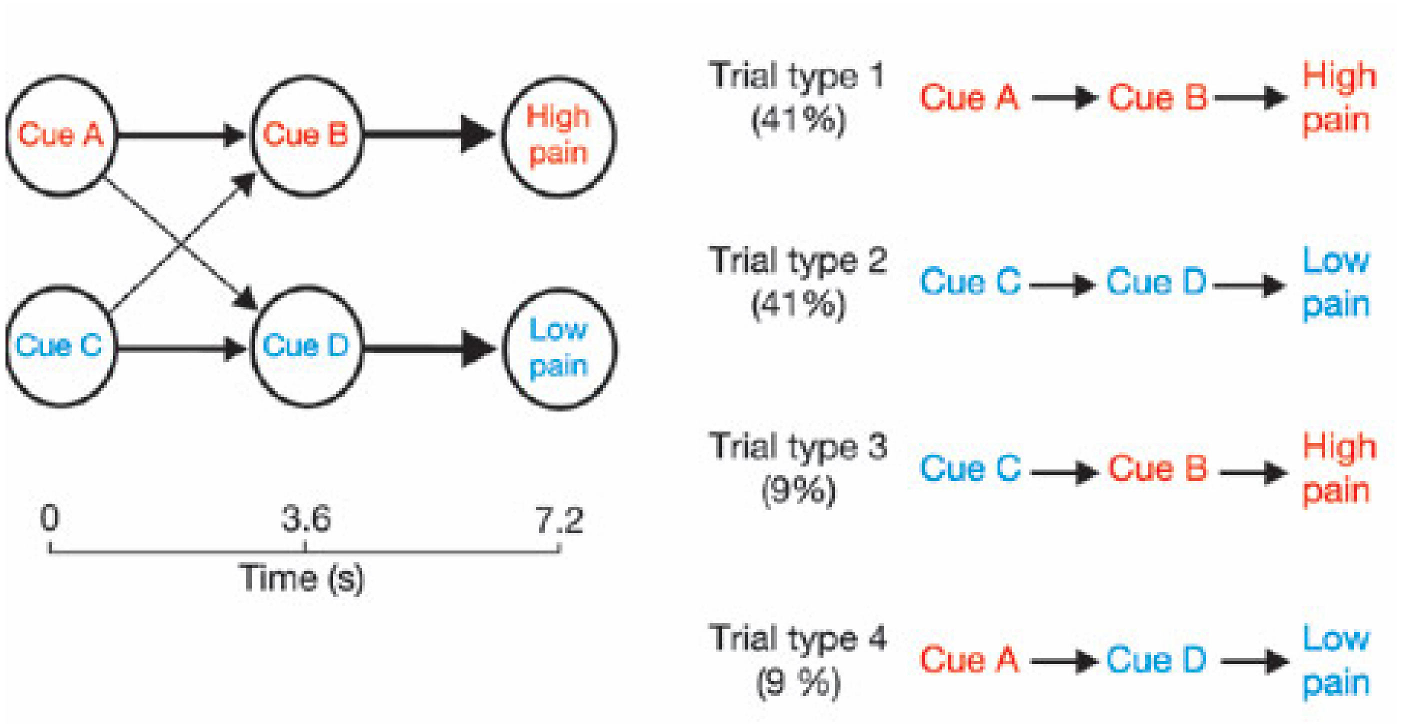
Example: Understanding pain conditioning in humans

- Animals (including humans) use environmental stimulus to predict future danger
- Seymour and colleagues used the temporal difference model to identify brain regions involved in the processing of aversive conditioning to sequential stimuli
- Using fMRI data gathered during pain conditioning, they identified those regions with strong negative or positive correlations with the temporal difference and the temporal difference errors predicted by their stimulus protocol

Seymour et al., ***Nature*** 429, 664-667 (2004)

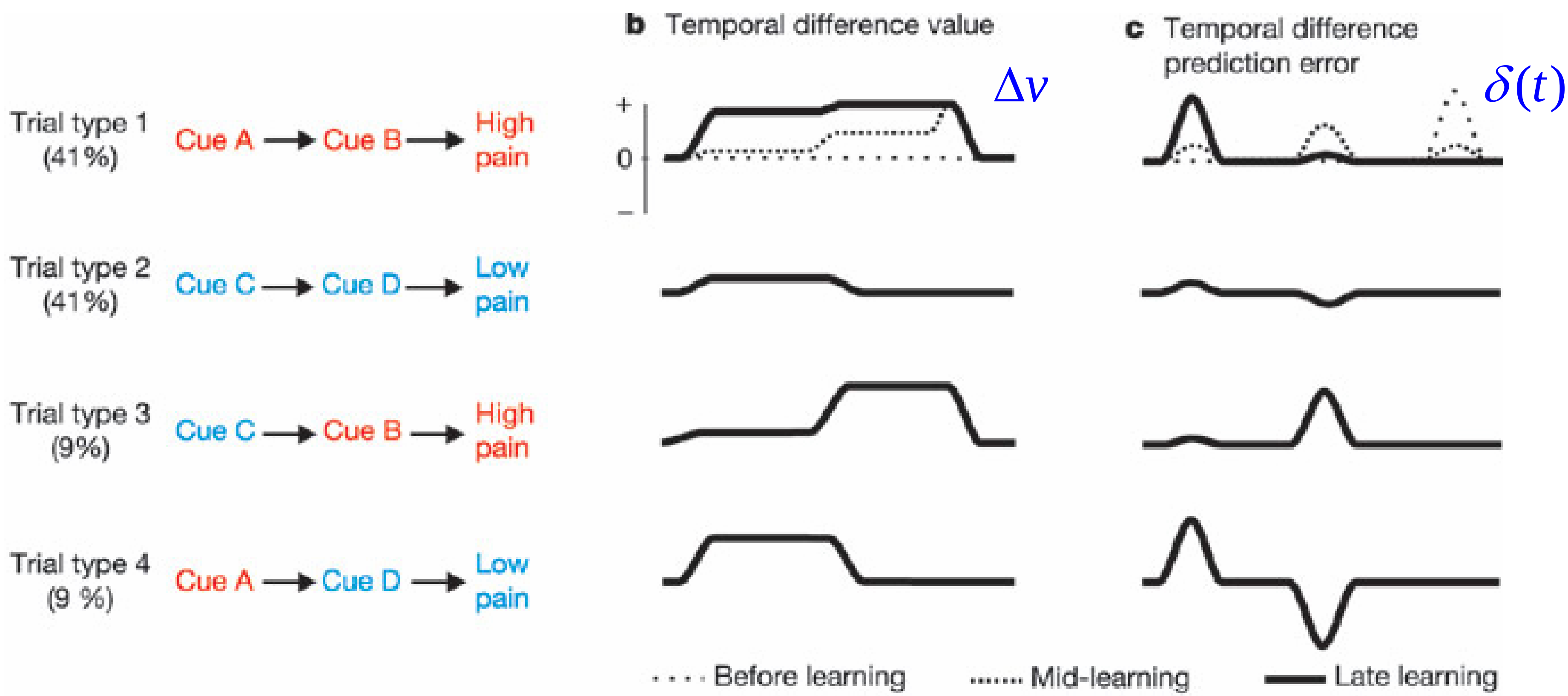
Pain conditioning protocol

- Subjects were asked to judge if cue were on the left or the right
- Second cue completely predicted intensity of pain stimulus
- First clue probabilistically predicted the intensity – in a small percentage of trials, the second cue would reverse the prediction of the first trial



Seymour et al., ***Nature*** 429, 664-667 (2004)

Response predicted by Temporal Difference Rule



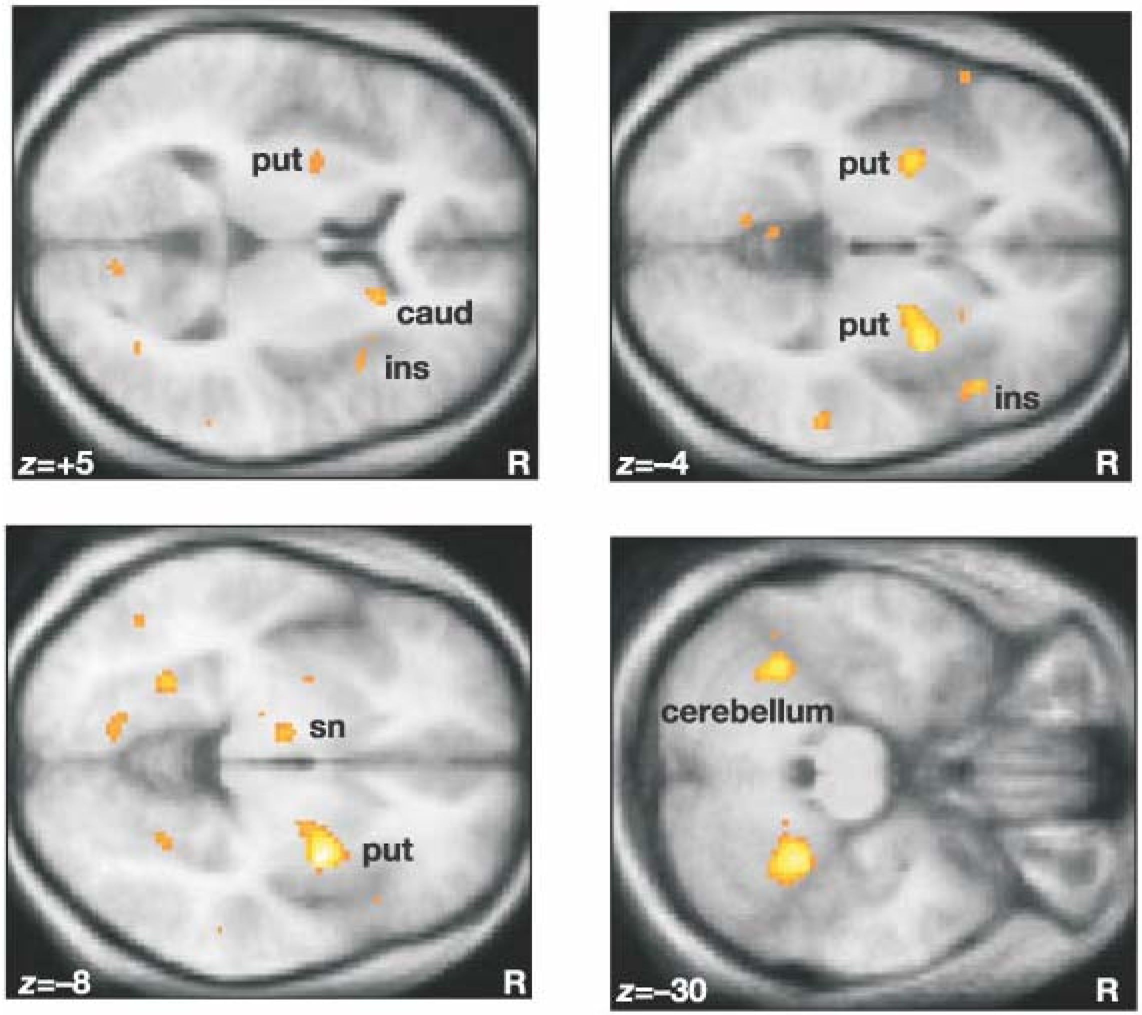
$$v(t) = \sum_{\tau=0}^{t} w(\tau) u(t-\tau)$$

$$\Delta w(\tau) = \varepsilon \delta(t) u(t-\tau)$$

$$\delta(t) = r(t) + v(t+1) - v(t)$$

Seymour et al., ***Nature*** 429, 664-667 (2004)

Regions showing significant correlation with the temporal difference error

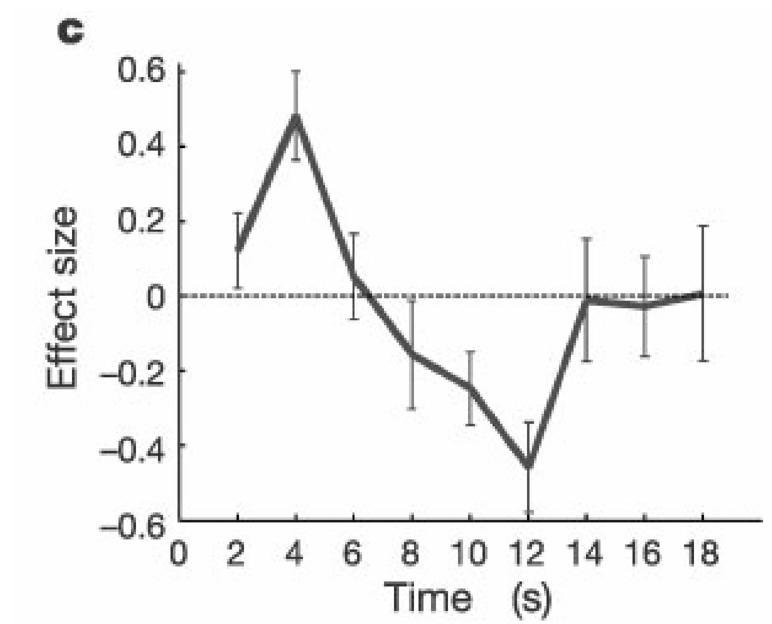


Prediction error was highly correlated with activity in both the right and the left ventral putamen, as well as caudate, cerebellum, right insula, left substantia nigra

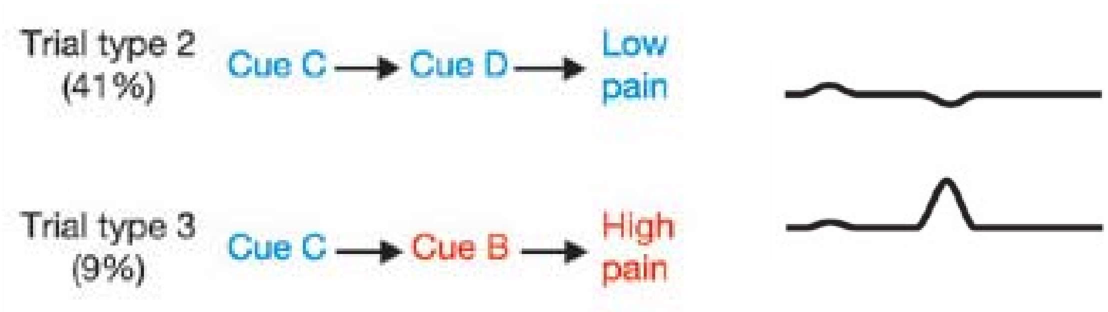
Seymour et al., ***Nature*** 429, 664-667 (2004)

Ventral Putamen showed biphasic response

Activity in ventral putamen showed a biphasic response, similar to that predicted by the temporal difference error

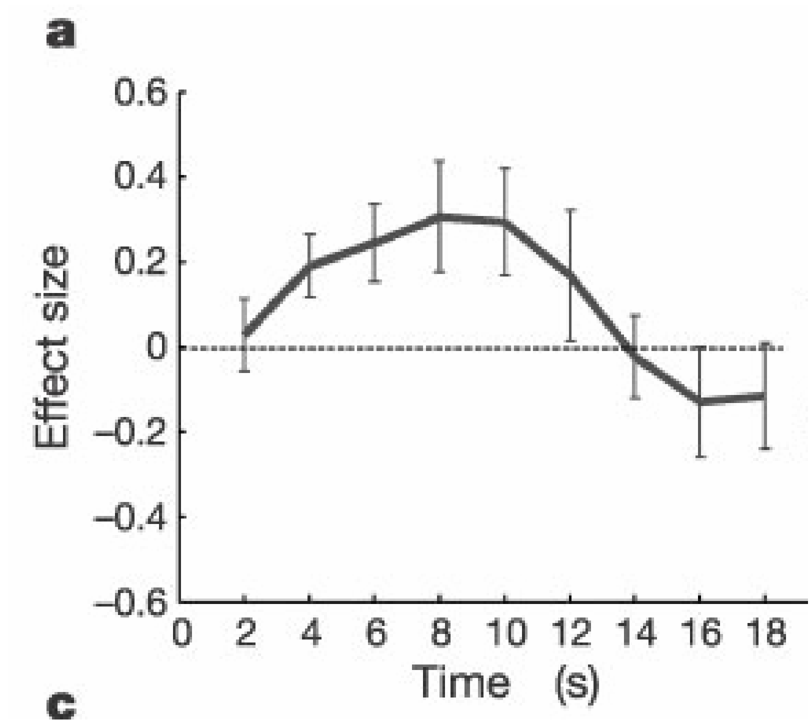


Comparison of Ventral Putamen activity between conditions

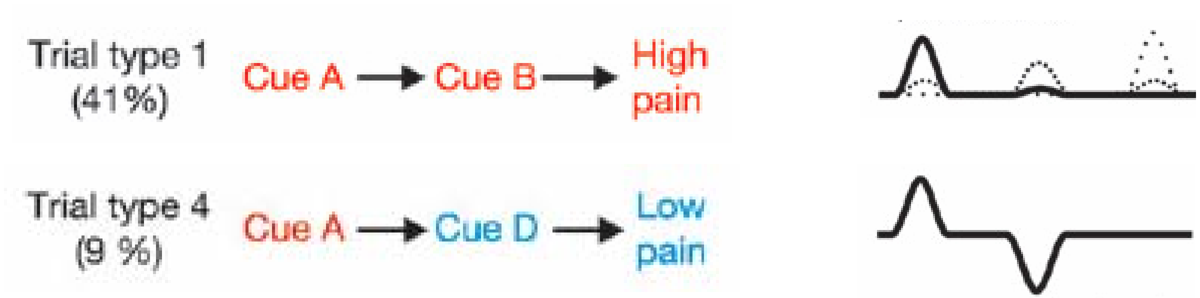


Positive prediction error: trial 3 – trial 2

Expectation was the first cue would predict stimulus, but it does not

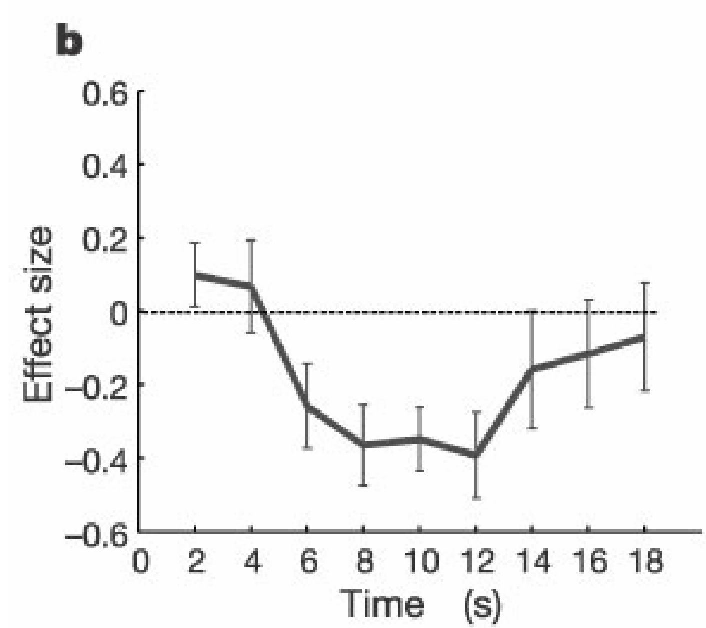


Comparison of Ventral Putamen activity between conditions

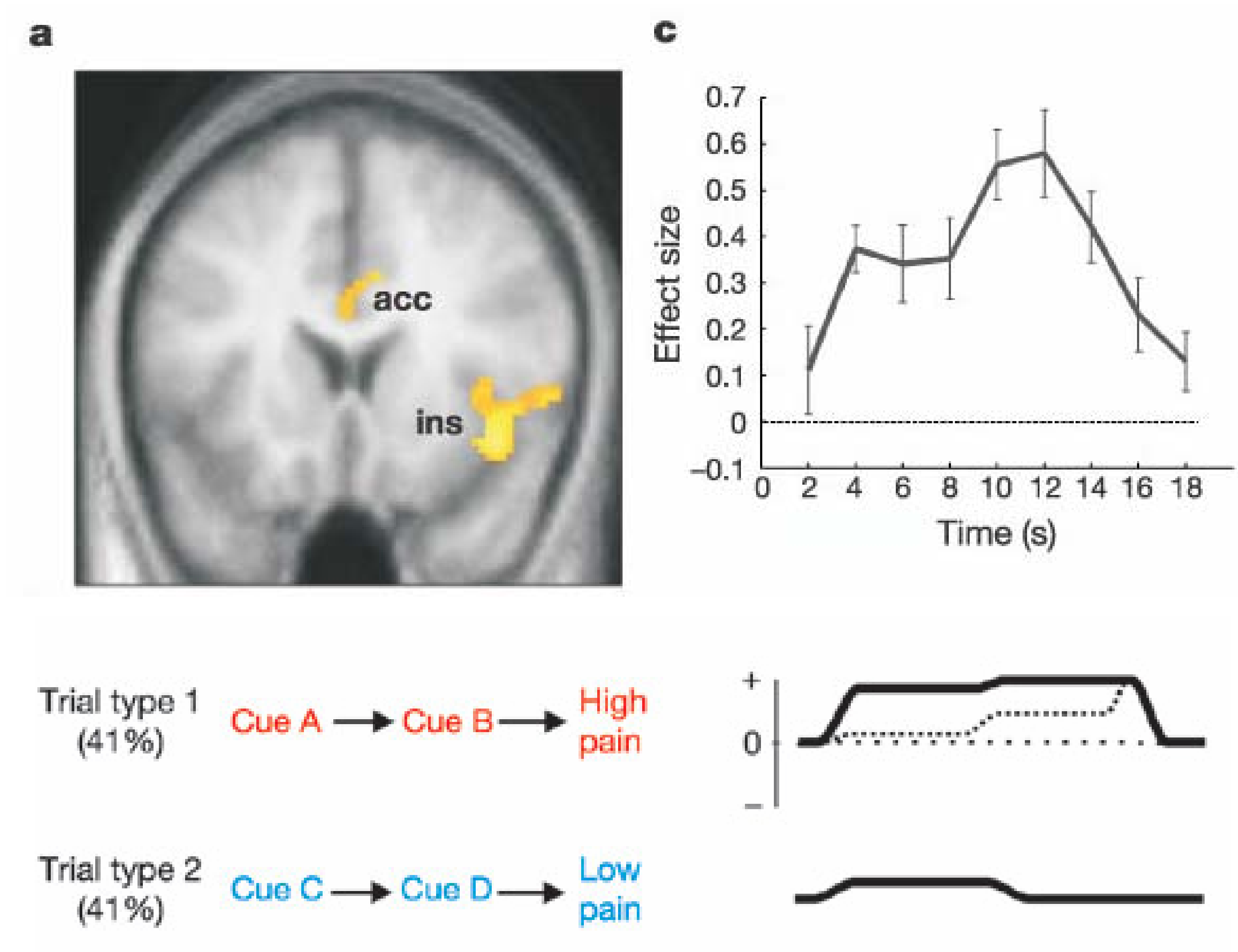


Negative prediction error: trial 4 – trial 1

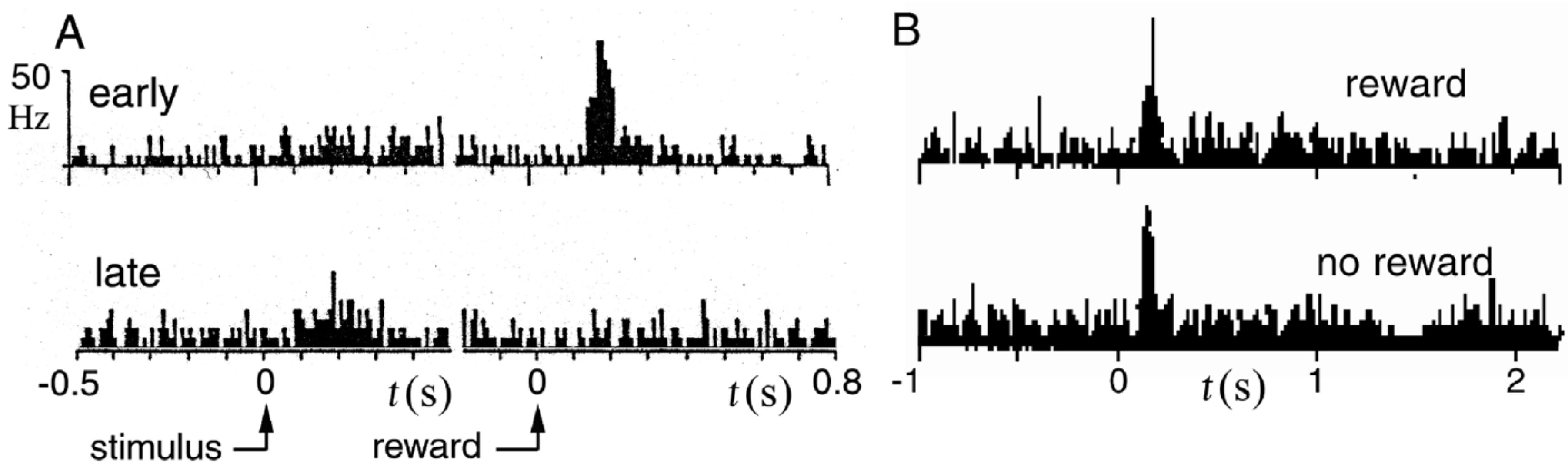
Expectation was that first cue would not predict stimulus, but it does

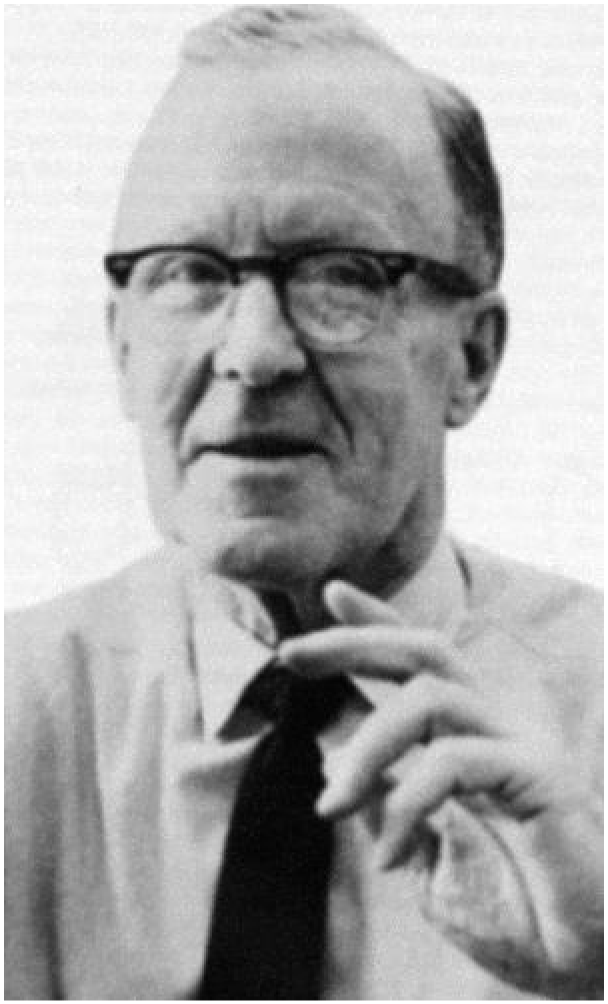


Right anterior insula showed correlations with temporal difference



How do the neurons know?





"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

Unsupervised learning

- Let *u* represent the pre-synaptic activity level, *v* the post-synaptic activity level
- Using a linear integrate-and-fire model:

$$\tau_r \frac{dv}{dt} = -v + \sum_{b=1}^{N_u} w_b u_b$$

- If the stimuli are presented slowly w.r.t. the neuron dynamics, then set *v* to the asymptotically steady-state value:

$$v = \vec{w} \cdot \vec{u}$$

Unsupervised learning (cont.)

- Basic plasticity rule (based on Hebb's conjecture):

$$\tau_w \frac{d\vec{w}}{dt} = v\vec{u}$$

- Averaged over all input patterns during training:

$$\tau_w \frac{d\vec{w}}{dt} = \langle v\vec{u} \rangle \qquad v = \vec{w} \cdot \vec{u}$$

$$= \vec{Q} \cdot \vec{w} \;\; \text{where} \;\; \vec{Q} = \langle uu \rangle$$

Correlation-based plasticity rule

Stability

- Problem: what constrains the weights?

$$\tau_w \frac{d\vec{w}}{dt} = v\vec{u}$$

- Take the dot product of *w* with both sides:

$$\tau_w \vec{w} \cdot \frac{d\vec{w}}{dt} = v\vec{w} \cdot \vec{u}$$

- And note that: $\frac{d\left|\vec{w}\right|^2}{dt} = 2\vec{w}\frac{d\vec{w}}{dt}$ so $\tau_w \vec{w}\frac{d\vec{w}}{dt} = \frac{\tau_w}{2}\frac{d\left|\vec{w}\right|^2}{dt}$

Stability (cont.)

- So, given that $\tau_w \vec{w} \cdot \frac{d\vec{w}}{dt} = v\vec{w} \cdot \vec{u}$ and

$$\tau_w \vec{w} \frac{d\vec{w}}{dt} = \frac{\tau_w}{2} \frac{d\left|\vec{w}\right|^2}{dt} \qquad v = \vec{w} \cdot \vec{u}$$

- We note that: $\frac{\tau_w}{2} \frac{d\left|\vec{w}\right|^2}{dt} = v^2$

- So, the weight vector grows continuously, and therefore we have unbounded growth – we need to constrain it
  - Many (not very biologically plausible) saturation constraints have been proposed