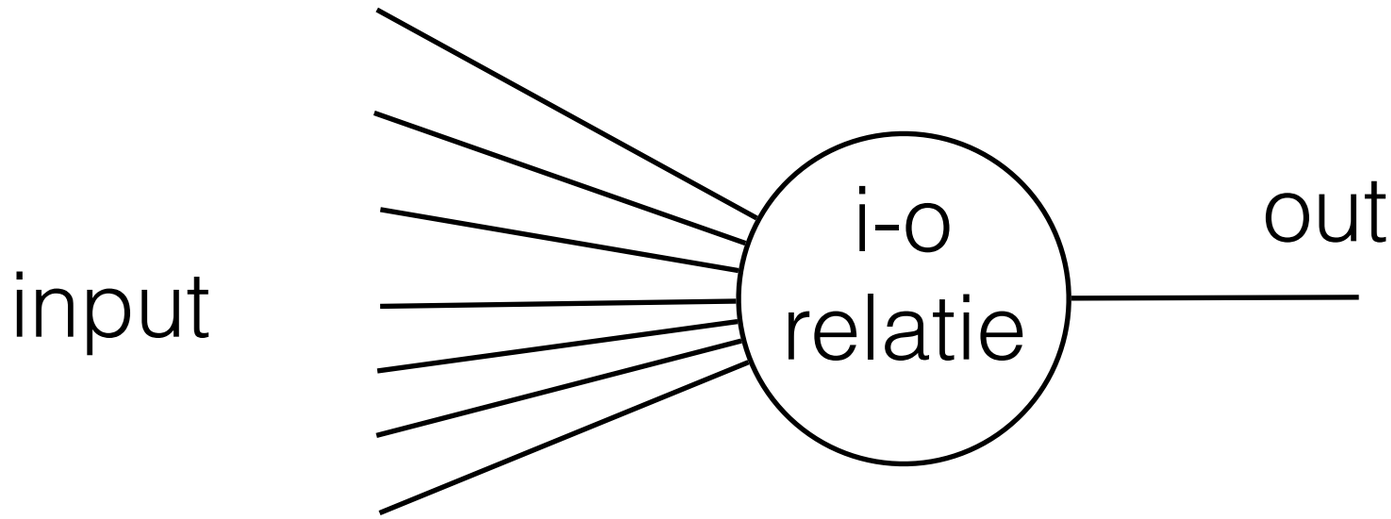
Leren in netwerkmodellen 2a

Fleur Zeldenrust

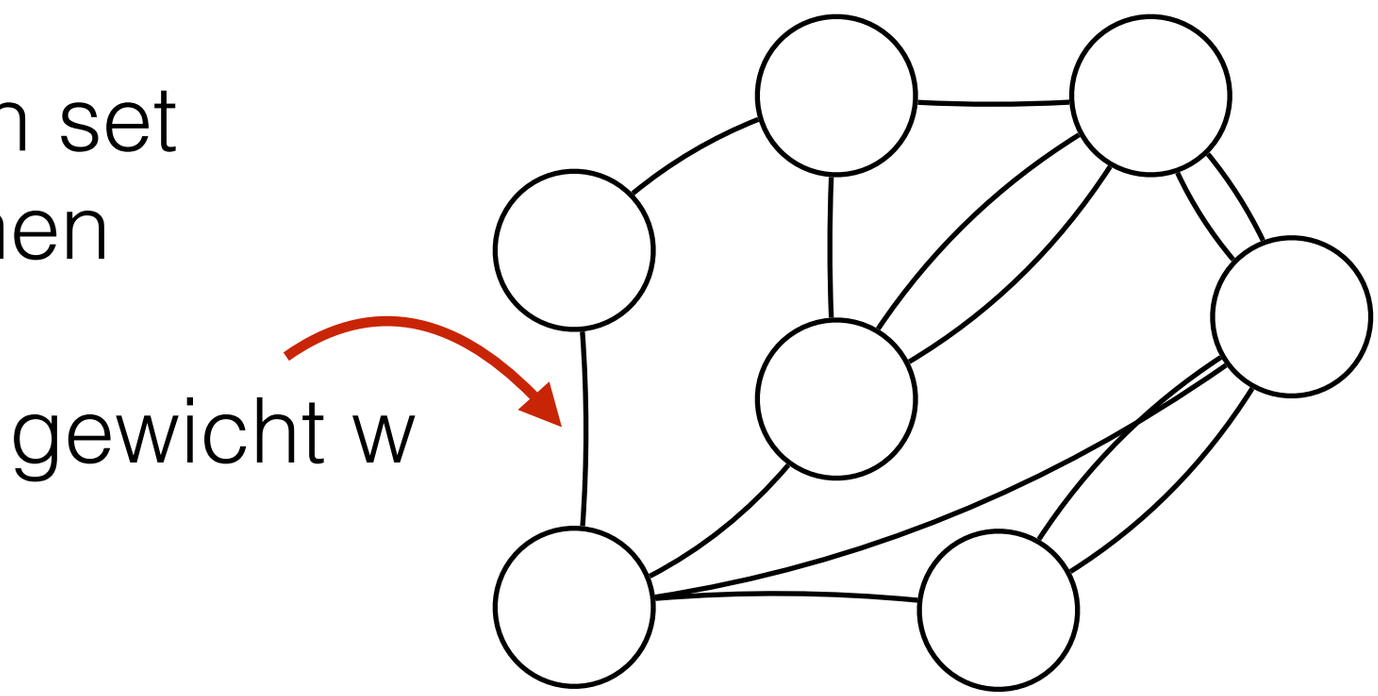
Leren & Geheugen, 2017

Herhaling Perceptie

- Neuron integreert input, zet om naar output



- Een netwerk is een set verbonden neuronen

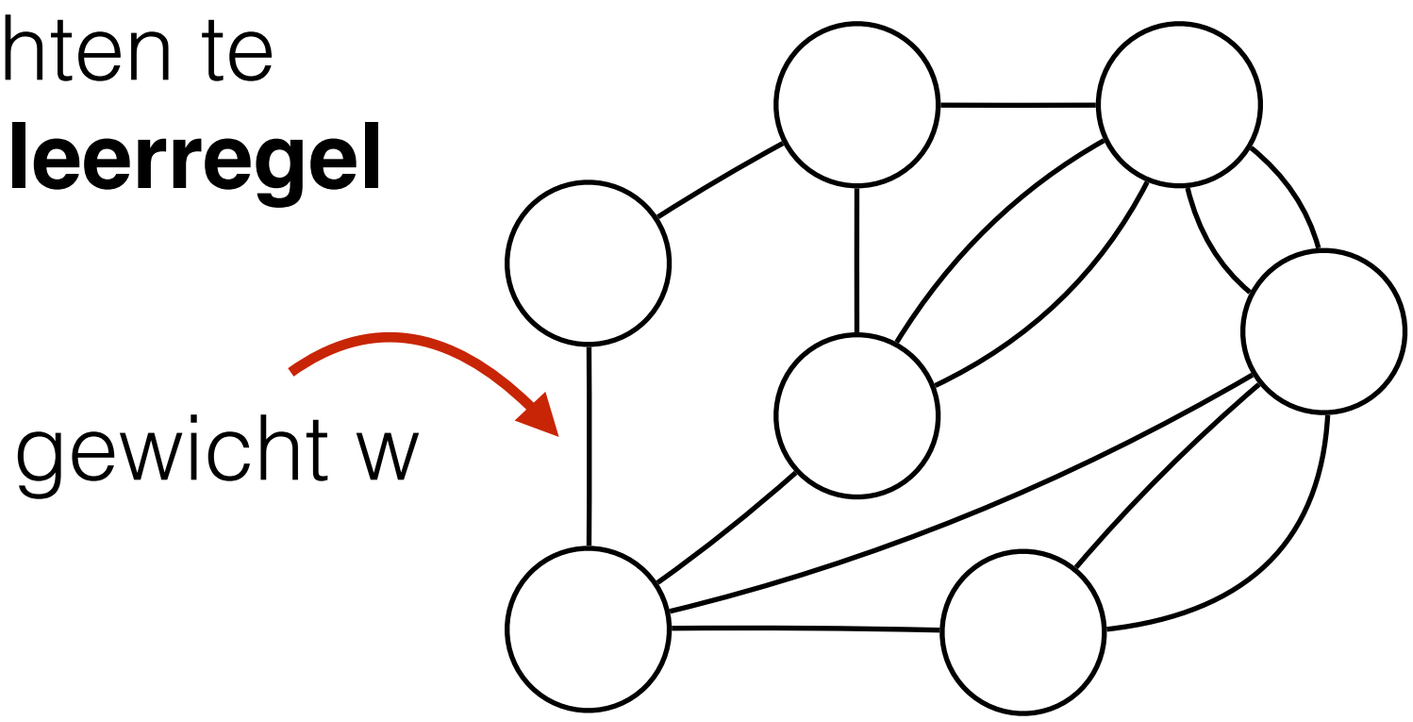


Leren in neurale netwerken

- Aanname: netwerken leren door veranderingen in gewichten (synapsen)

- Dus vraag is: hoe verander ik gewichten zo dat mijn netwerk de correcte output bij de input geeft?

- Een recept om gewichten te veranderen heet een **leerregel**



Soorten leren

We kunnen 3 soorten leren onderscheiden:

1. **Unsupervised learning**: netwerk leert op basis van de input alleen (bijvoorbeeld receptive field)

2. **Supervised learning**: er is een leraar die vertelt hoe het netwerk het fout had (veel gebruikt in AI, misschien in cerebellum?)

3. **Reinforcement learning**: er is een (eventueel vertraagde) beloning of straf (bijvoorbeeld leren fietsen)

Overzicht ‘Hebb’-achtige regels

- ‘Naïeve’ regel: alleen LTP (geen LTD), instabiel (synapsen stoppen nooit met groeien)

- Covariance regel: ook LTD, maar nog steeds instabiel

- BCM regel:
    - LTD & stabiel
    - perk gewichten in door dynamische drempel T
    - competitie en selectiviteit

- Oja regel:
    - LTD & stabiel
    - kwadratische som gewichten constant
    - competitie en selectiviteit

Conclusie unsupervised (Hebbian) learning

Rate based

- Klassiek Hebbiaans leren zorgt voor **cell assemblies** en **pattern completion**
- Klassieke Hebb leerregel en covariance geven gewichten die een reflectie zijn van correlaties in de input, maar zijn **instabiel**
- Een activiteit-afhankelijke drempel (BCM rule) of normalisatie gewichten (Oja rule) stabiliseert gewichten en zorgt voor **competitie** en daardoor **selectiviteit**

Spike based: Spike Timing-Dependent Plasticity

- additive (harde grens, competitie, bimodale distributie)
- multiplicative (zachte grens, geen competitie, unimodaal)

Samenvatting supervised learning

- **Perceptron regel:** voor binair neuron, enkellaags perceptron

- **Delta regel:** voor rate neuron, enkellaags perceptron

- **Error-backpropagation:** voor rate neuron, meerlaags perceptron

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

College 2

3. **Reinforcement learning**
    - functie van dopamine in de hersenen
        - herhaling: klassiek conditioneren
    - basal ganglia: action selection
        - exploratie en exploitatie
    - cortex: predictive coding

Reinforcement learning

- op basis van feedback van omgeving (reward/ beloning en straf/punishment)

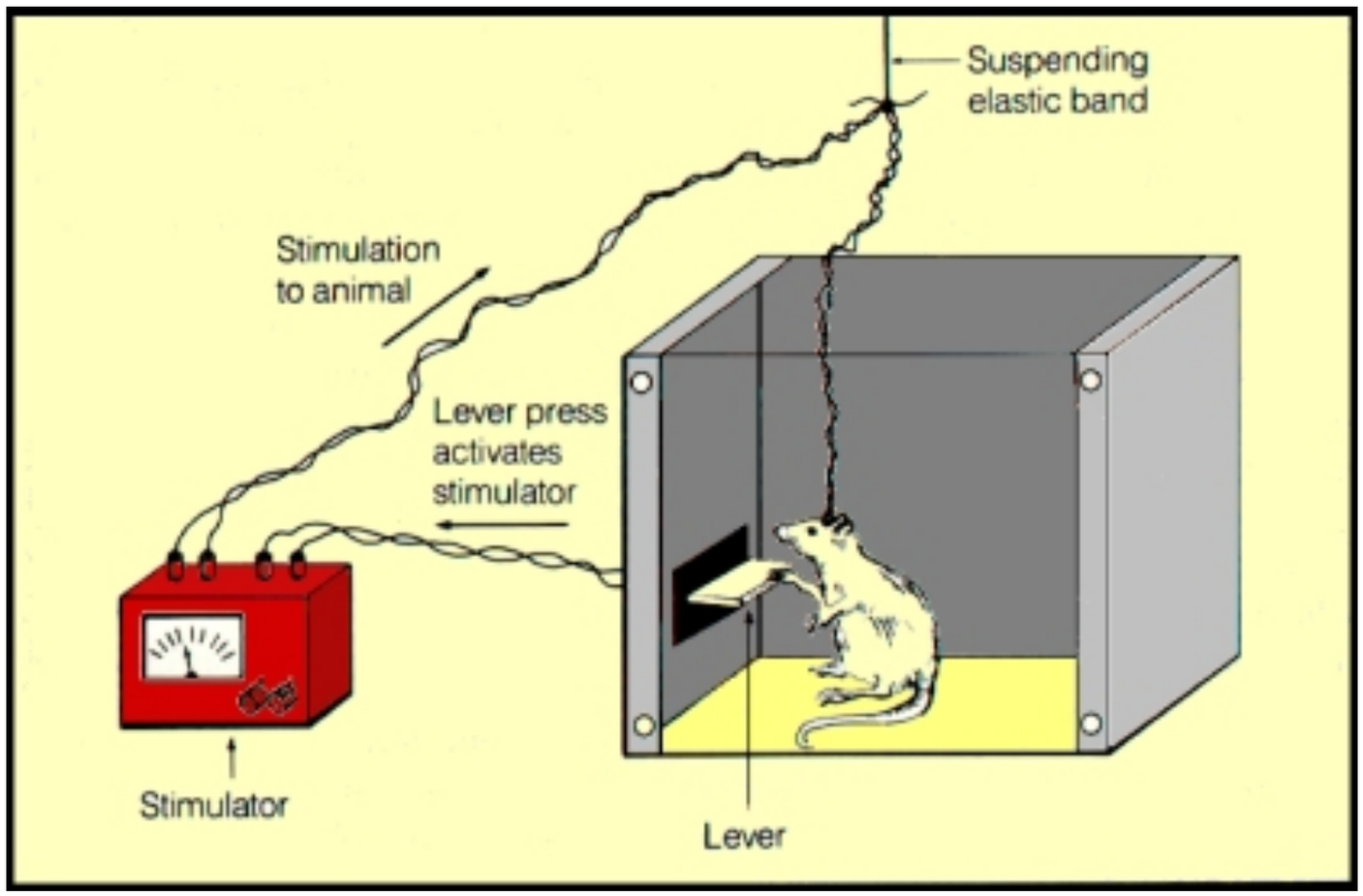
- maar nu niet per cel, maar algemeen (‘van je fiets vallen’) → biologisch realistischer

- ‘maakt me niet uit hoe je het doet, als dit het resultaat maar is’

- nadruk op acties

Beloning in de hersenen

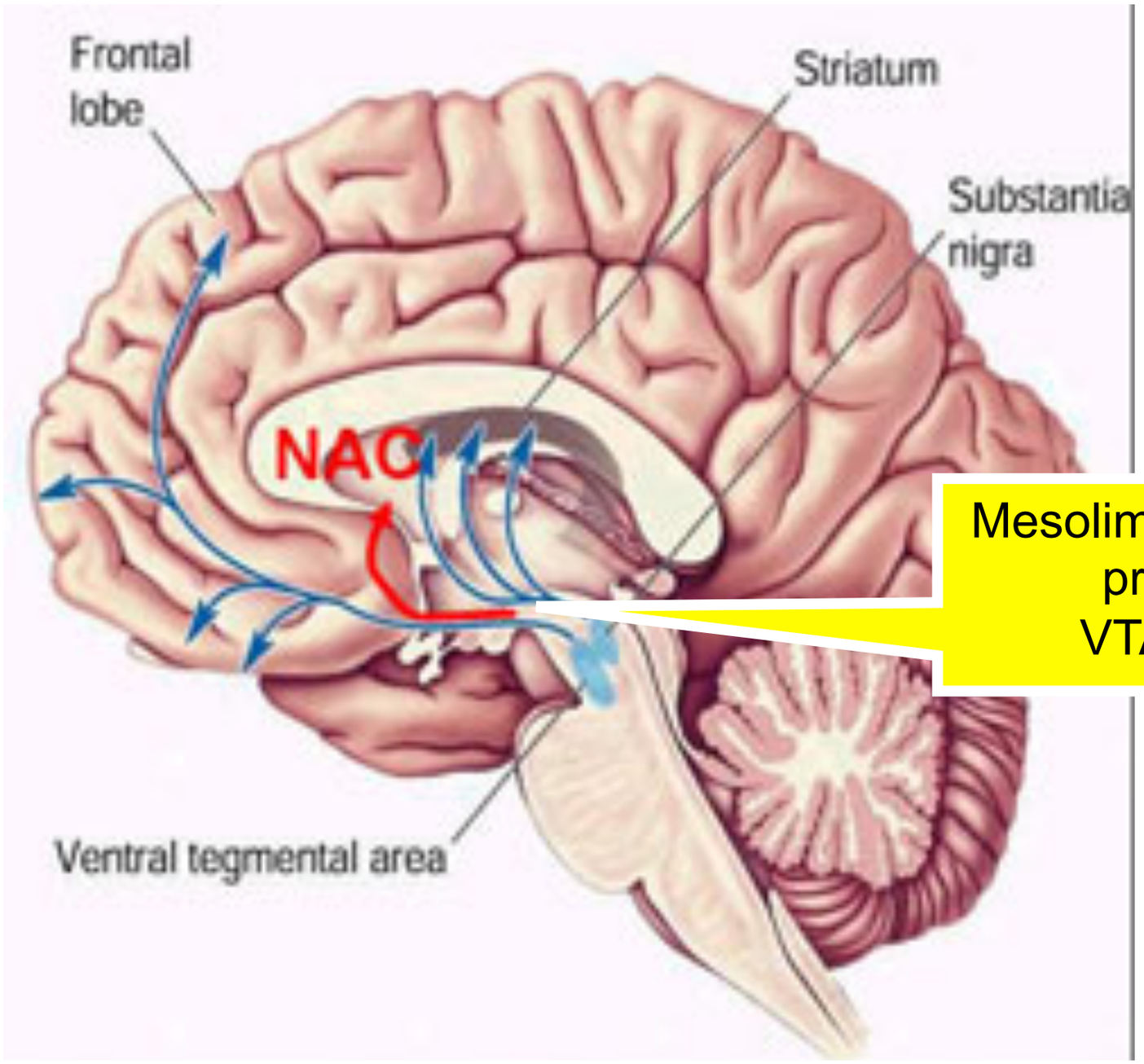
Olds & Milner, 1953/54: dieren voeren gedrag uit om stimulatie van sommige hersengebieden te ontvangen



Beloning door stimulatie

- allerlei soorten gedrag: terugkeren bepaald gebied kooi, lever pressing
    - blijkbaar geeft die stimulatie ‘beloning’: motivatie
- in vrijwel alle diersoorten
- in verschillende hersengebieden: hypothalamus, ventral tegmental area (VTA), nucleus accumbens (NAc, basal ganglia), amygdala, prefrontal cortex,…

Dopamine (DA)!


Frontal lobe
Striatum
Substantia nigra
NAC
Ventral tegmental area
Mesolimbic dopamine projection
VTA → NAc

Dopamine == Beloning?

1978, Roy A. Wise: anhedonia hypothese:

- ‘prettige' ervaringen zorgen voor toename DA in NAc
- drugs zorgen voor verandering DA activiteit
- toedienen pimozide (DA antagonist) ratten:
    - stoppen geleerd gedrag
    - niet meer leren nieuw gedrag

Dopamine == Beloning? Nee!

- 6-OHDA (doodt DA neuronen) injectie (of ratten zonder DA neuronen):
    - ratten stoppen ‘lever-pressing’ voor eten, maar stoppen niet met eten
    - ratten hebben nog steeds preferenties
- Neuronen in NAc activeren voor ‘straf'
- Lesie NAc: avoidance learning werkt niet meer
- Neuronen in VTA zijn alleen actief voor **onverwachte** beloning!

Dus wat doet dopamine dan wel?

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

College 2

3. **Reinforcement learning**
    - functie van dopamine in de hersenen
        - herhaling: klassiek conditioneren
    - basal ganglia: action selection
        - exploratie en exploitatie
    - cortex: predictive coding

Herhaling: klassiek conditioneren

1. Vóór conditioneren:

Unconditioned stimulus (US)

Stimulus die op natuurlijke wijze Een respons veroorzaakt

Unconditioned response (UR)

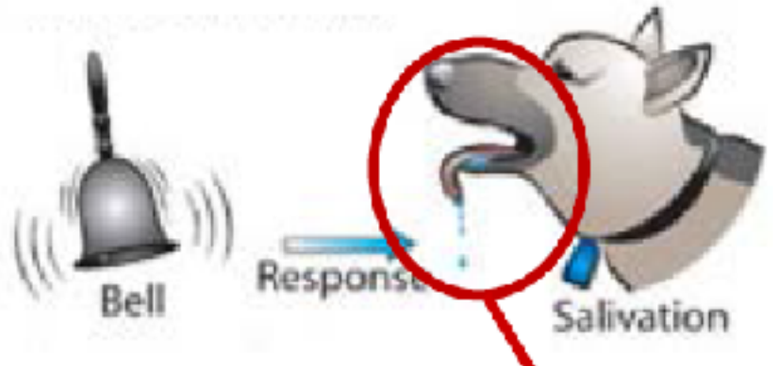
Unconditioned = Unconditional
'Zonder voorwaarde'

3. Tijdens conditioneren:

Neutrale stimulus

Geen Unconditioned Response (UR)

4. Na conditioneren:



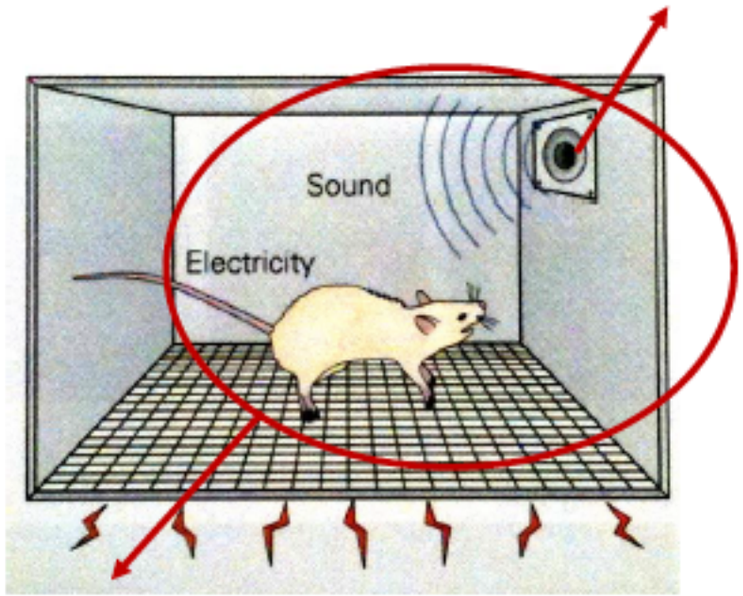
Conditioned stimulus (CS)

Conditioned response (CR)

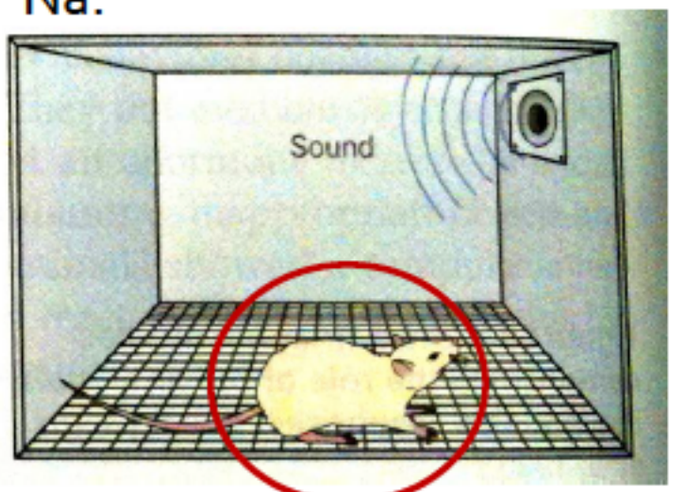
Herhaling: klassiek aversief conditioneren

Toon: Conditioned Stimulus (CS)

Voor / tijdens:



Na:



footshock: unconditioned Stimulus (US)

Startle response: unconditioned response (UR)

Gezamenlijk aanbieden
van CS en US (pairing)

Vriesgedrag / freezing:

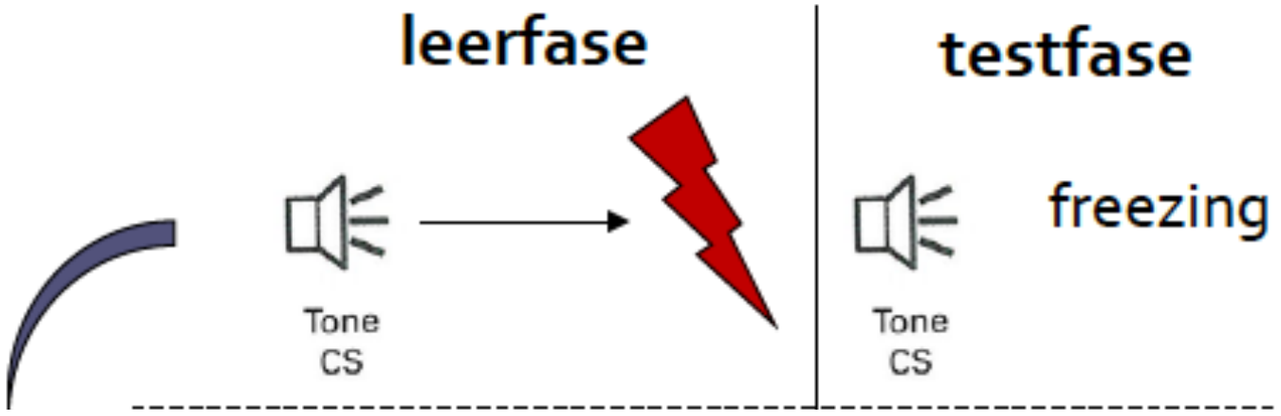
Conditioned Response (CR)

Kan kwalitatief verschillen van UR

Herhaling: blocking

Kamin: blocking effect:

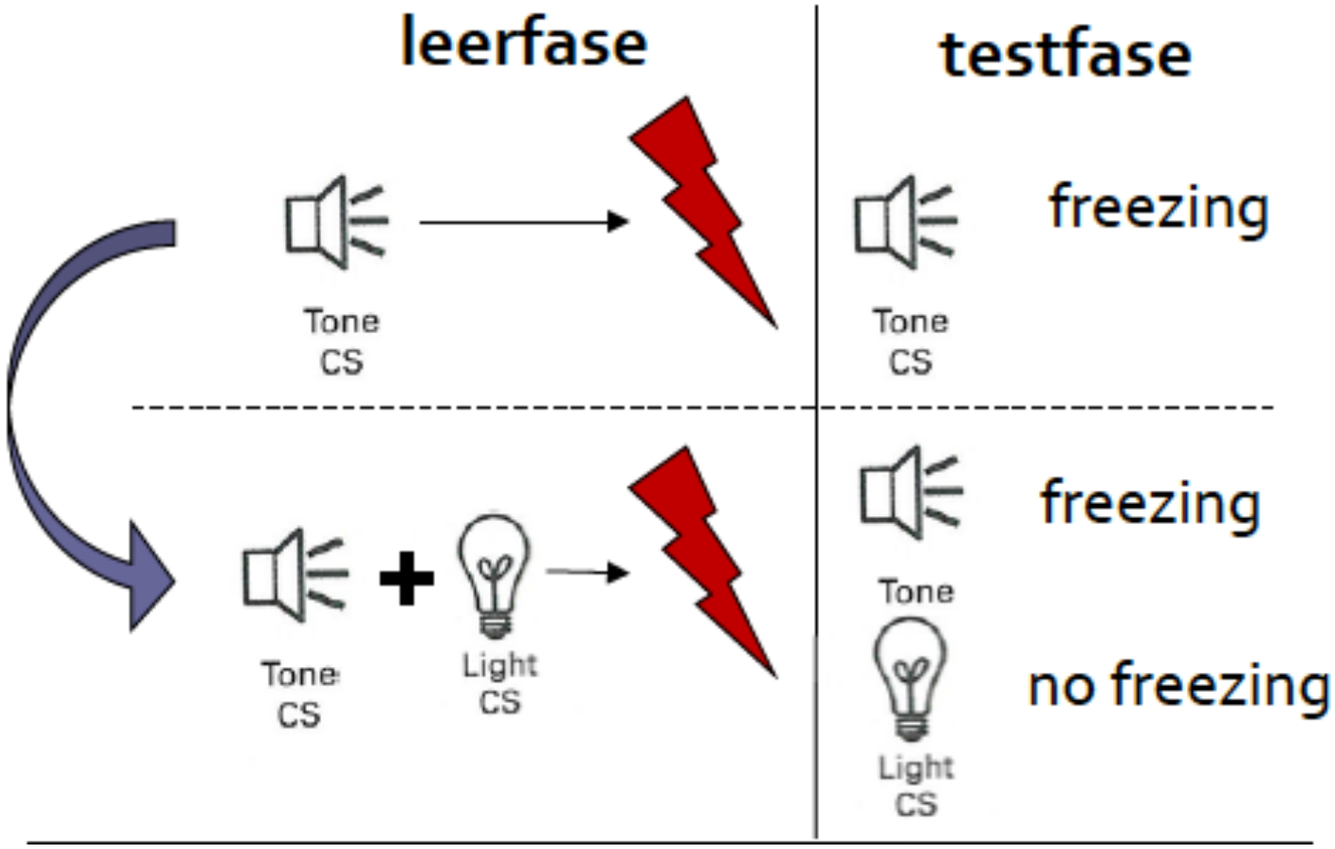
- Als een CS voor 100% de US voorspelt…



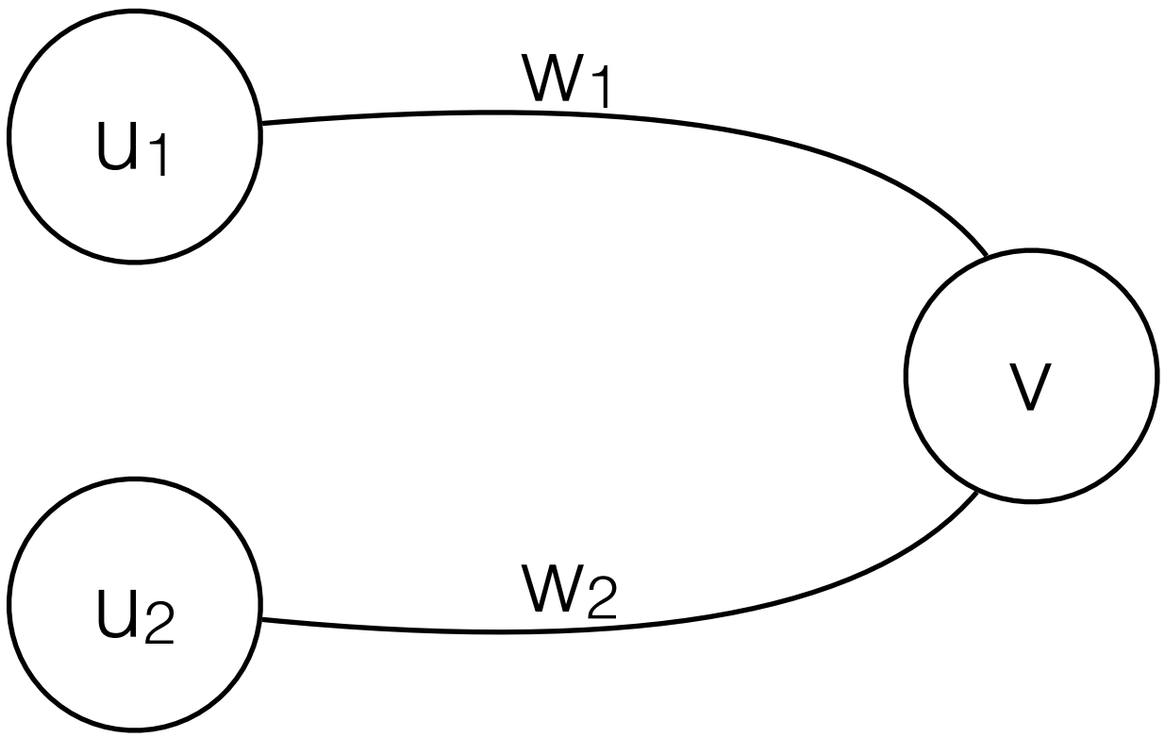
Herhaling: blocking

Kamin: blocking effect:

- Als een CS voor 100% de US voorspelt…
- Geeft 2e CS geen extra voorspellende informatie over US
  - 1e CS geeft wel CR
  - 2e CS niet



Hebbian (unsupervised) learning kan blocking niet verklaren!
Er is correlatie 2e CS en schok

Stimuli	Verwachte beloning	Ontvangen beloning	Prediction error
	v	r	δ=r-v

$$v = u_1 w_1 + u_2 w_{12}$$

Update gewichten

$$w_i \rightarrow w_i + \Delta w_i$$

$$\Delta w_i = \epsilon \delta u_i$$

Leerregel: Rescorla-Wagner (RW)

- u: stimulus; r: beloning (of straf!)
- v: verwachte beloning (of straf!) voor elke stimulus
- w_i = associatie stimulus met beloning
- error = r-v verschil ontvangen en verwachtte beloning
- Update elke keer gewicht om error te minimaliseren:

$$w_i \rightarrow w_i + \Delta w_i$$

$$\Delta w_i = \epsilon \delta u_i$$

$$\delta = r - v$$

Leerregel: Rescorla-Wagner (RW)

- Rescorla-Wagner is hetzelfde als delta rule supervised learning!

- Verschil: gewichten w niet tussen neuronen, maar tussen ‘associaties’. Dus: als 1 neuron = 1 ‘concept’ hetzelfde (‘grandmother cell’)

- **Prediction error**: verschil tussen verwachting en observatie, (hier: ‘geobserveerde beloning’ r): r-v

- RW: leren gebeurt niet door beloning, maar door **onverwachte** beloning

- RW verklaart blocking: het gewicht voor de nieuwe stimulus wordt niet aangepast als er geen prediction error (δ) is

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

$$\delta = r - v$$

Prediction error

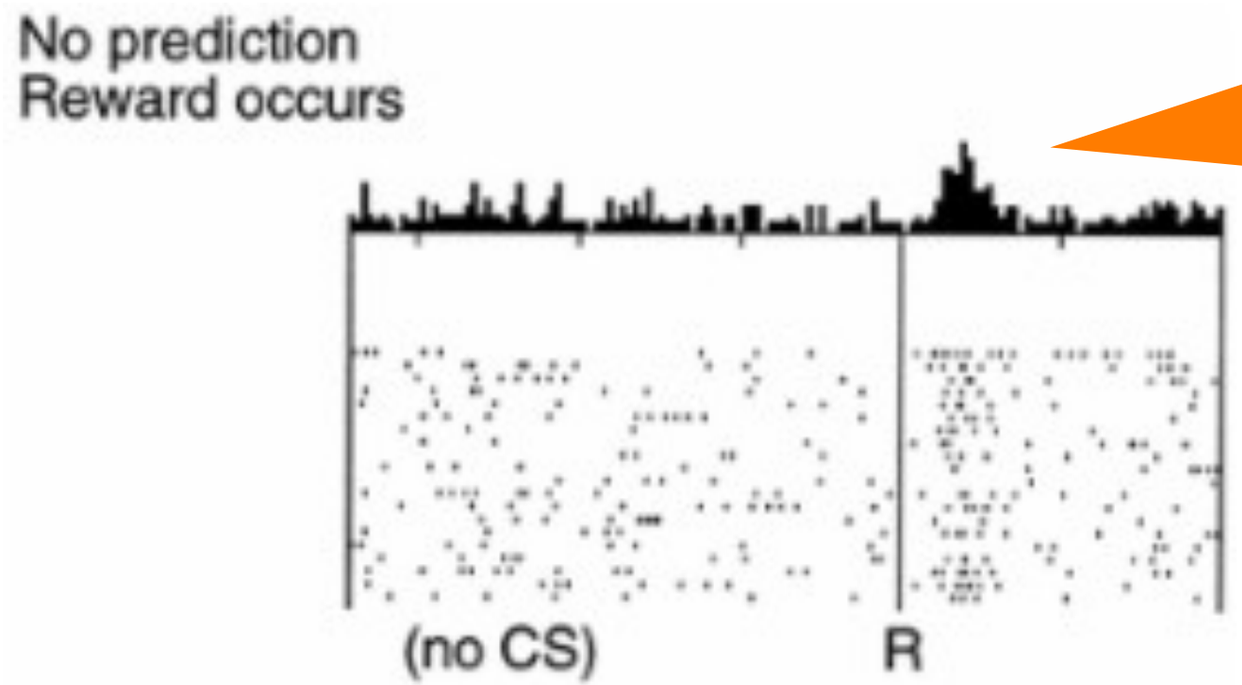
Een systeem dat **prediction errors** codeert moet:

- een verwachting of voorspelling van beloning hebben voor elke stimulus
- een positief signaal geven voor onverwachte beloning
- niet reageren op verwachte beloning
- negatief reageren op minder dan verwachte beloning (en positief reageren op kleiner dan verwachte straf)

Terug naar dopamine

- Dopamine neuronen in VTA zijn alleen actief voor **onverwachte** beloning

- Schulz & al (1998): in vivo VTA neuronen, apen



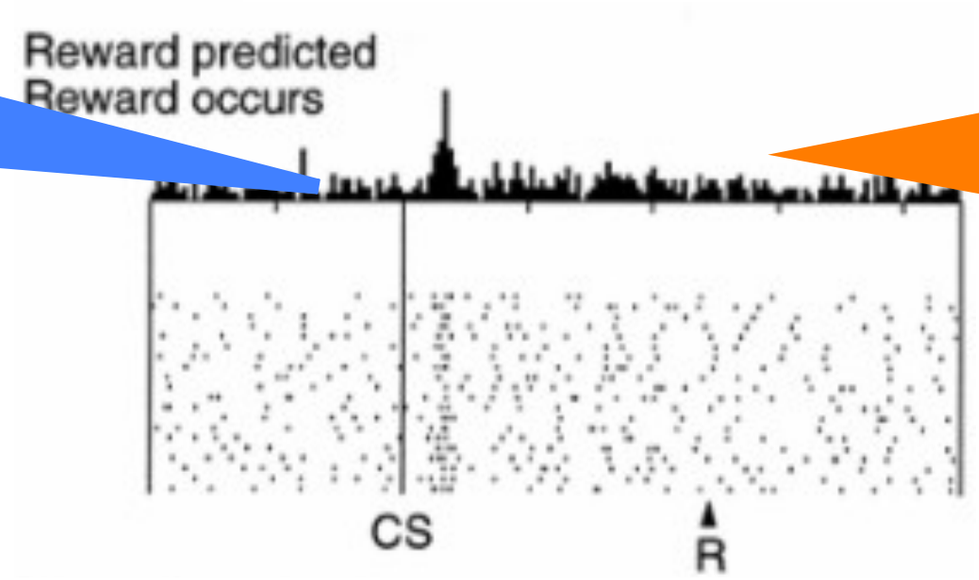
VTA neurons fire during reward delivery.
Hedonic value?

Terug naar dopamine

- Dopamine neuronen in VTA zijn alleen actief voor **onverwachte** beloning

- Schulz & al (1998): in vivo VTA neuronen, apen

VTA neurons now fire to the reward-predicting cue.

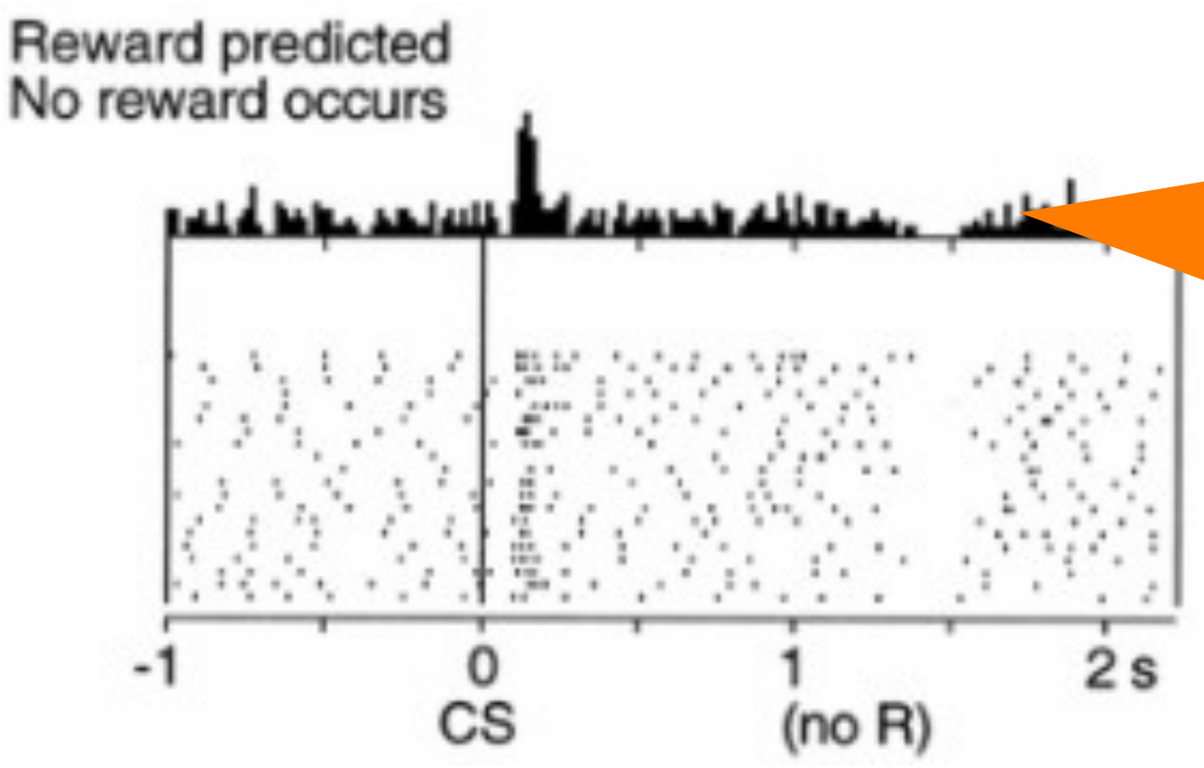


No hedonic value, because no activation when predicted reward occurs as expected.

Terug naar dopamine

- Dopamine neuronen in VTA zijn alleen actief voor **onverwachte** beloning

- Schulz & al (1998): in vivo VTA neuronen, apen

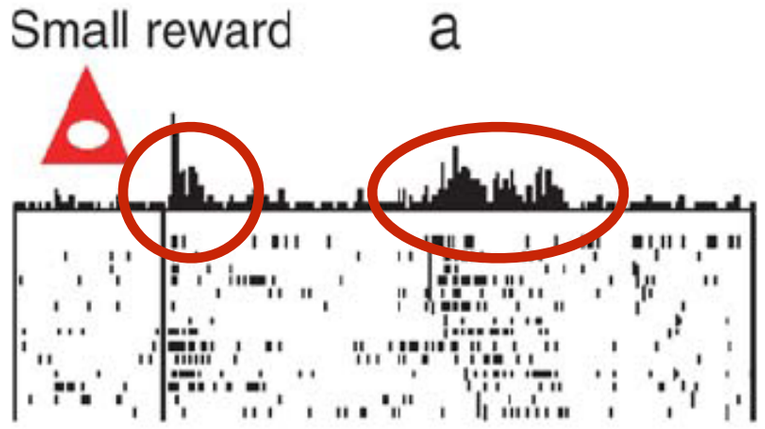


When the predicted reward is omitted, the cells stop firing at the timepoint of expected reward delivery.

Wat is vroege respons?

- dus: late respons VTA neuron lijkt op prediction error in RW regel

- maar: er is ook een vroege respons (net na stimulus) na het leren



Wat is vroege respons?

- dus: late respons VTA neuron lijkt op prediction error in RW regel

- maar: er is ook een vroege respons (net na stimulus) na het leren

- hangt samen met grootte beloning

- Dit kan niet verklaard worden door RW regel

Conclusie VTA neuronen

Late respons lijkt prediction error weer te geven

Vroege respons lijkt het soort beloning weer te geven:

- grootte
- kwaliteit
- waarschijnlijkheid / onzekerheid
- delay

Dit zit niet in RW model, maar wel in Temporal Difference (TD) model

Temporal Difference (TD)

- Neem ook beloning in de toekomst (**future rewards**) mee
- Discounting: Hoe verder in de toekomst belonong, hoe ‘minder waard’: γ (0< γ<1)
- ‘total future reward’ = $R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$
- ‘total expected future reward’ = $V_t = \sum_{\mathrm{i}=0}^{\infty} \gamma^i w u_{t+i}$

Temporal Difference (TD)

- Beloning op tijdstip t is het verschil in voorspellingen total future reward op tijdstip t en op tijdstip t+1

$$R_t - \gamma R_{t+1} = -\Delta R = r_t$$

- Als mijn verwachtingen kloppen, dan gelt dit ook voor verwachte future reward V: ΔR = ΔV

- Dus de **prediction error** is nog steeds verschil ontvangen en verwachte reward:

$$\delta = \Delta V + r_t$$

- Zo kun je weer een leerregel maken:

RW

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

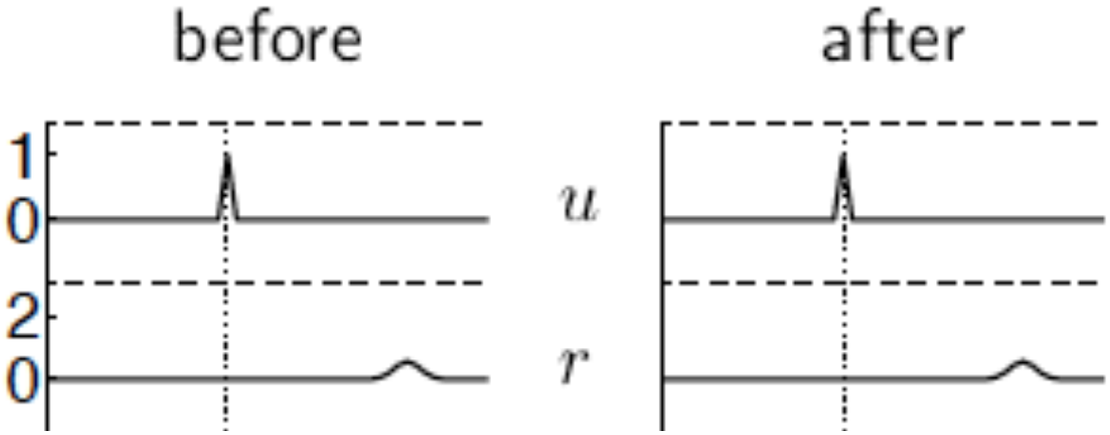
$$\delta = r - v$$

TD

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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)

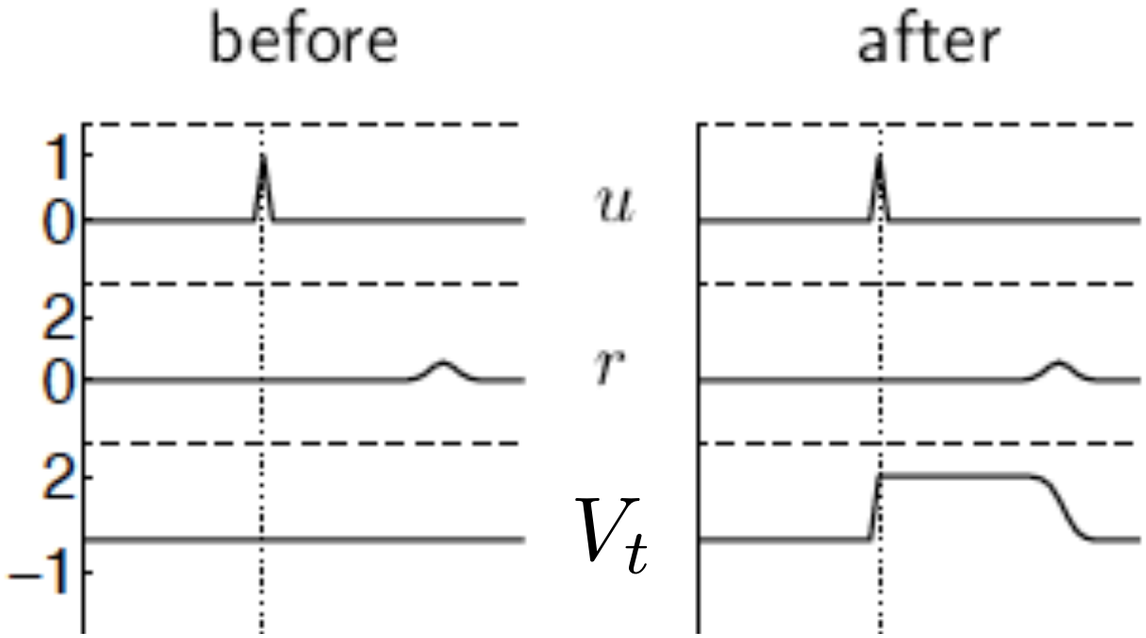


ontvangen reward

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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)



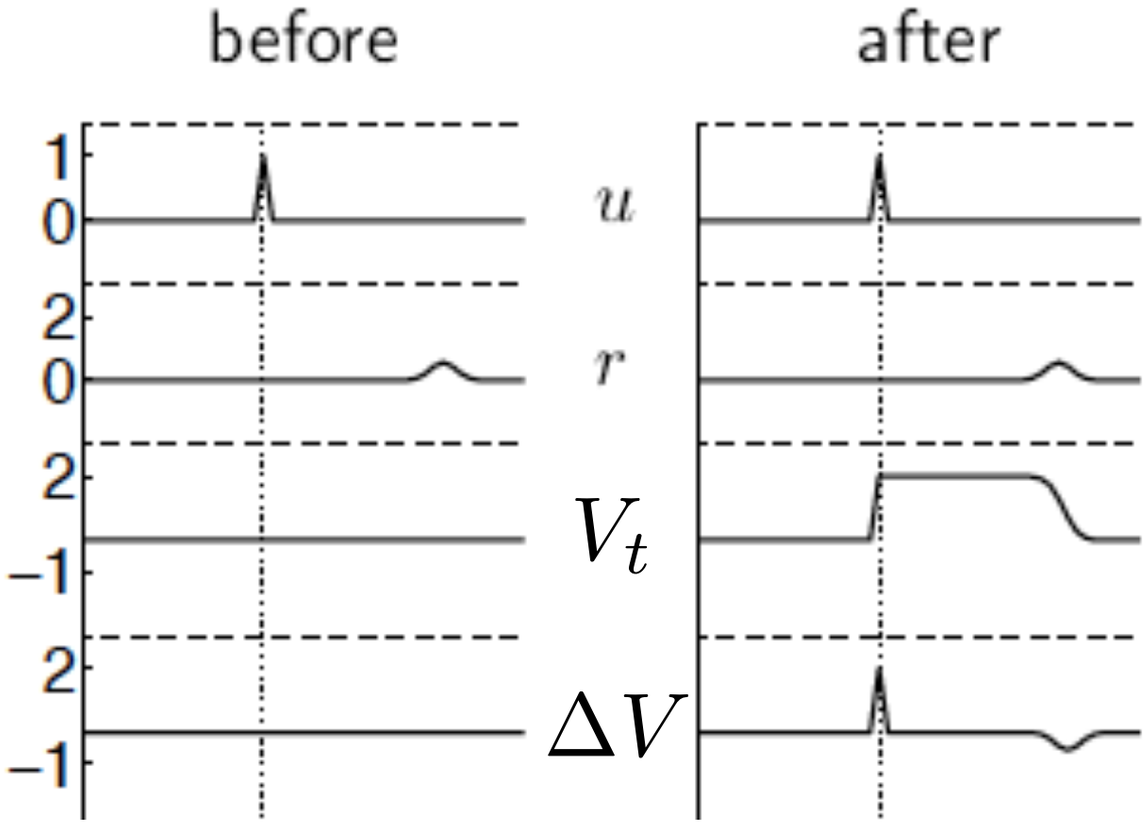
ontvangen reward

total expected future reward

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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)



ontvangen reward

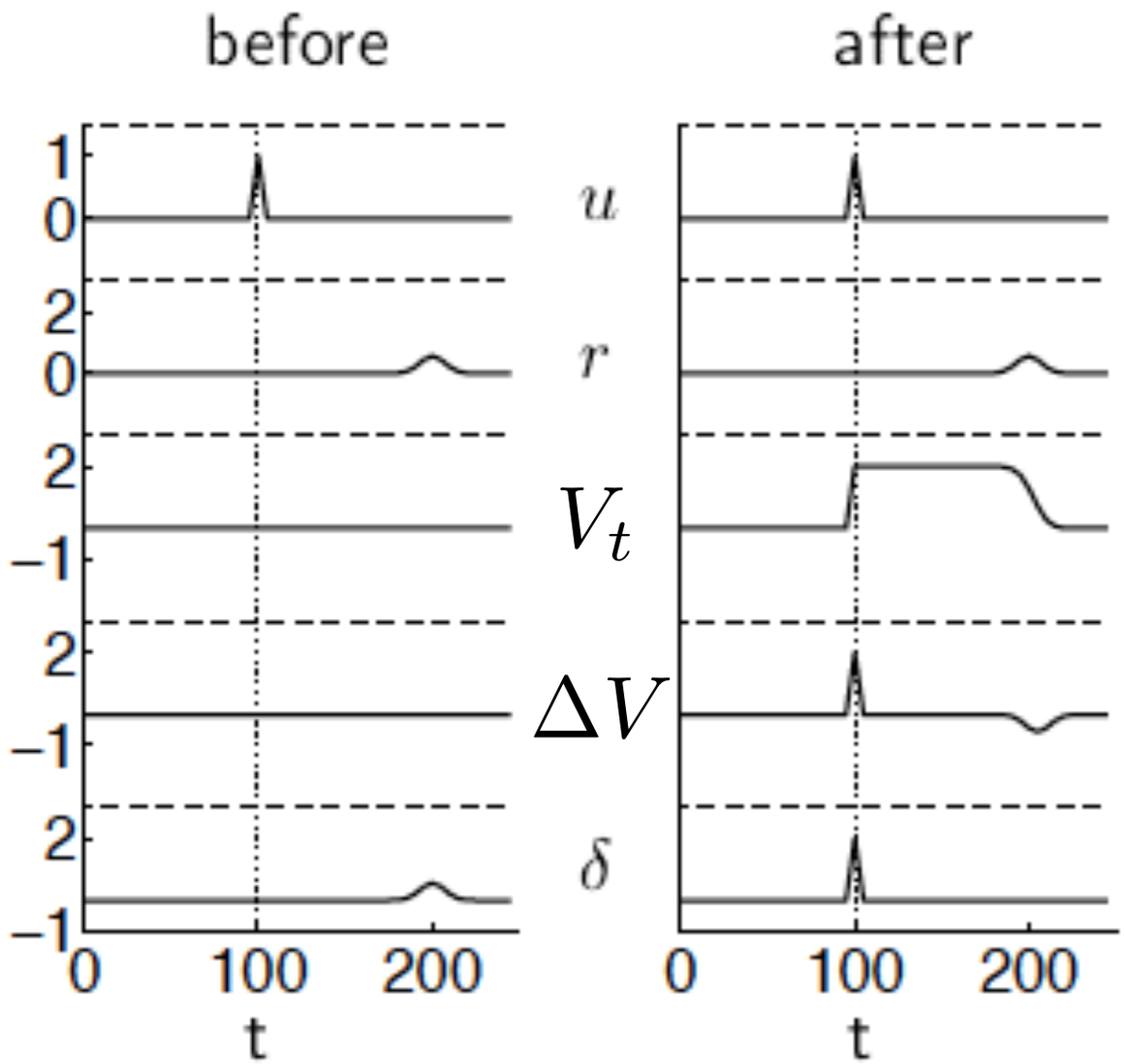
total expected future reward

verschil expected future reward

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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)



Waar lijkt het vuren VTA neuronen op?

ontvangen reward

total expected future reward

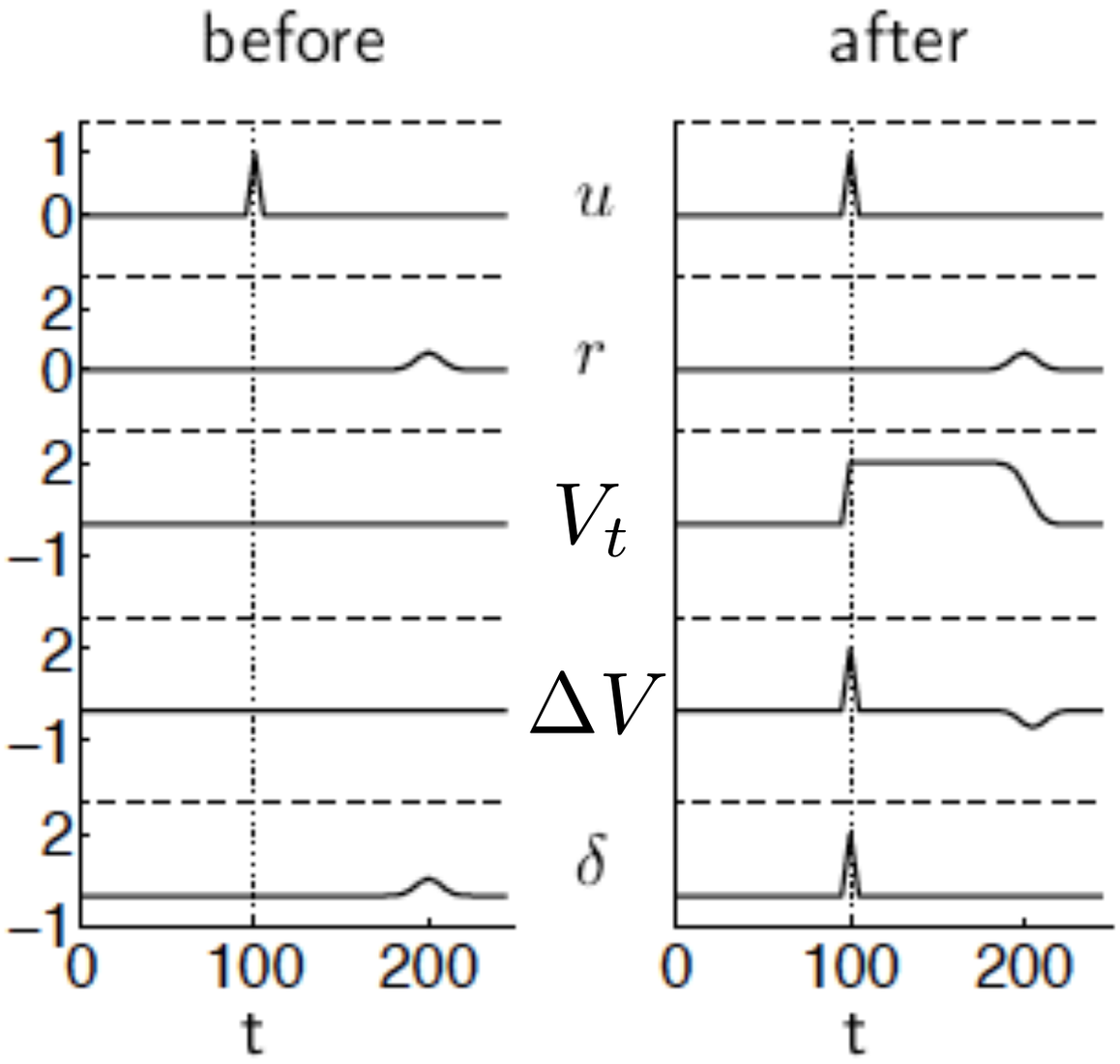
verschil expected future reward

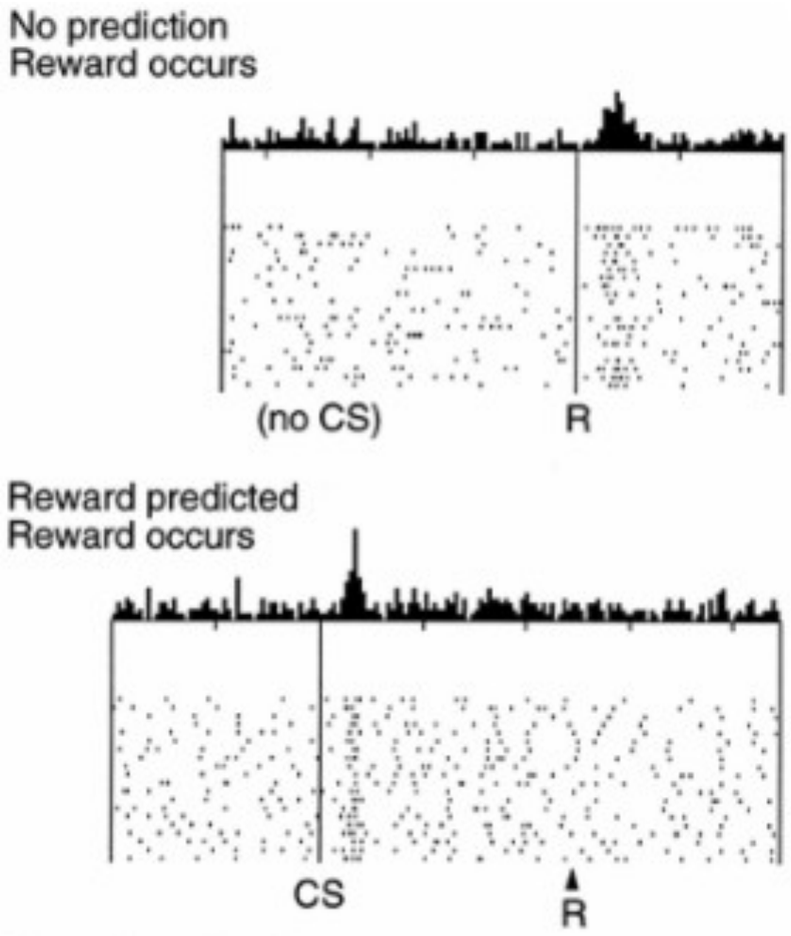
error

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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)





Temporal Difference (TD)

Conclusie:

Vuren VTA dopamine neuronen lijkt verdacht veel op de prediction error δ in temporal difference learning!

Ook hier weer: temporal difference learning vergelijkbaar met de ‘delta' regel van supervised learning!

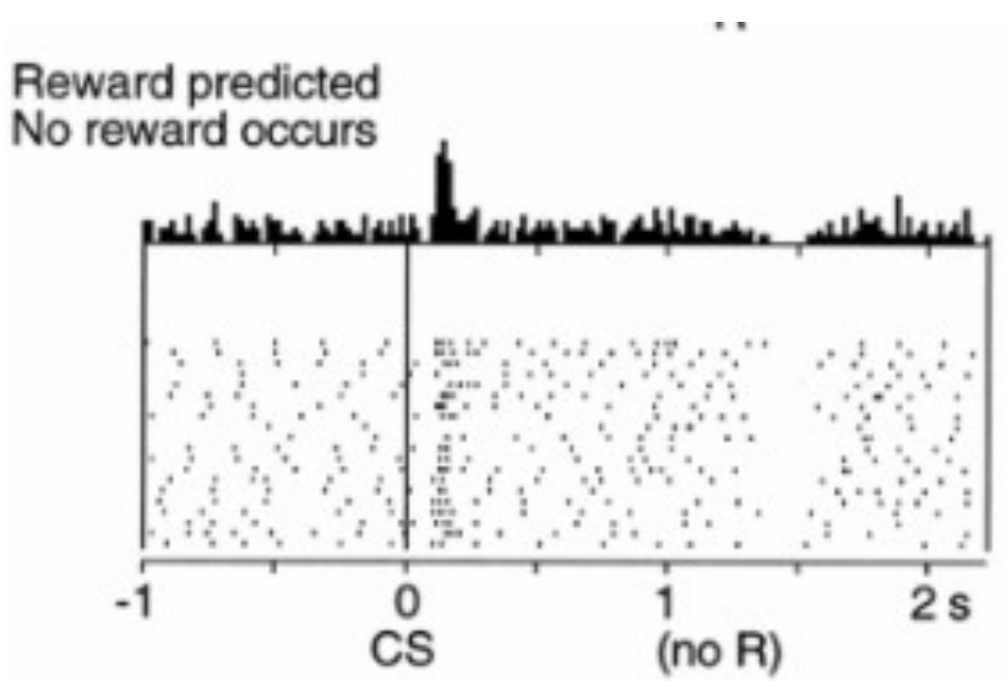
VTA geeft verschil verwachtte en gekregen beloning weer (prediction error).

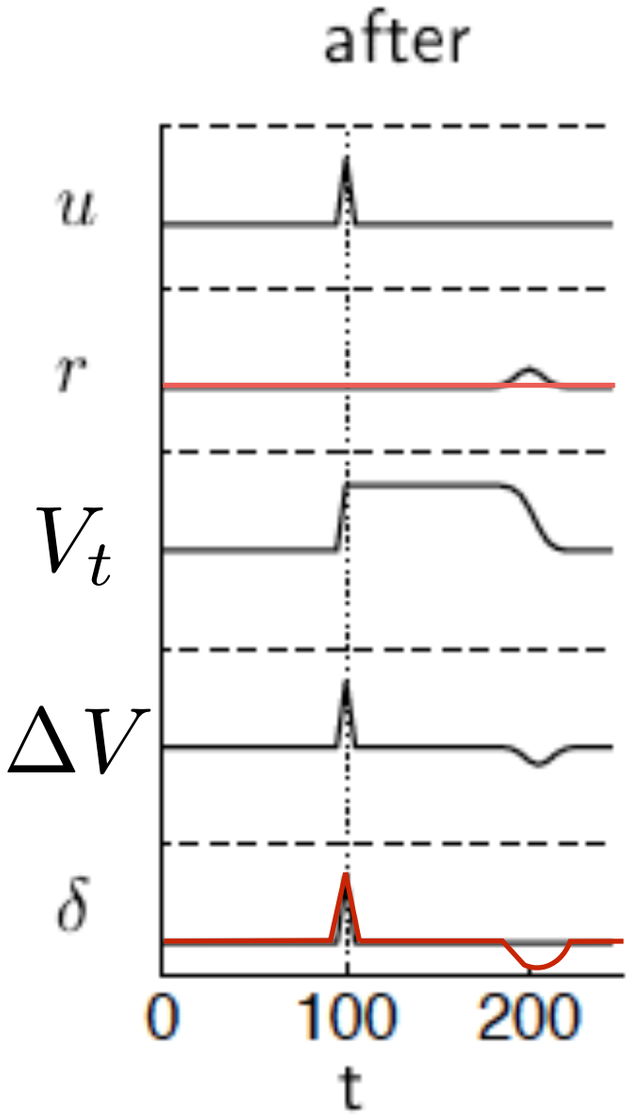
Hoe wordt dit verwerkt door de rest van de hersenen?

Pauze

Temporal Difference (TD)

Wat gebeurt er als je de beloning weglaat na leren?





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

$$\delta = \Delta V + r_t$$

Temporal Difference (TD)

- Neem ook beloning in de toekomst (**future rewards**) mee, maar discounted met γ
- 'total future reward' = $R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$
- NB j = i-1

$$
\begin{aligned}
&= r_t + \sum_{i=1}^{\infty} \gamma^i r_{t+i} \\
&= r_t + \sum_{j=0}^{\infty} \gamma^{j+1} r_{t+j+1} \\
&= r_t + \gamma \sum_{j=0}^{\infty} \gamma^j r_{t+j+1} \\
&= r_t + \gamma R_{t+1}
\end{aligned}
$$