Leren in netwerkmodellen 2b

Fleur Zeldenrust

Leren & Geheugen, 2017

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

Vandaag

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

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

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

$$\delta = r - v$$

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)

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?

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

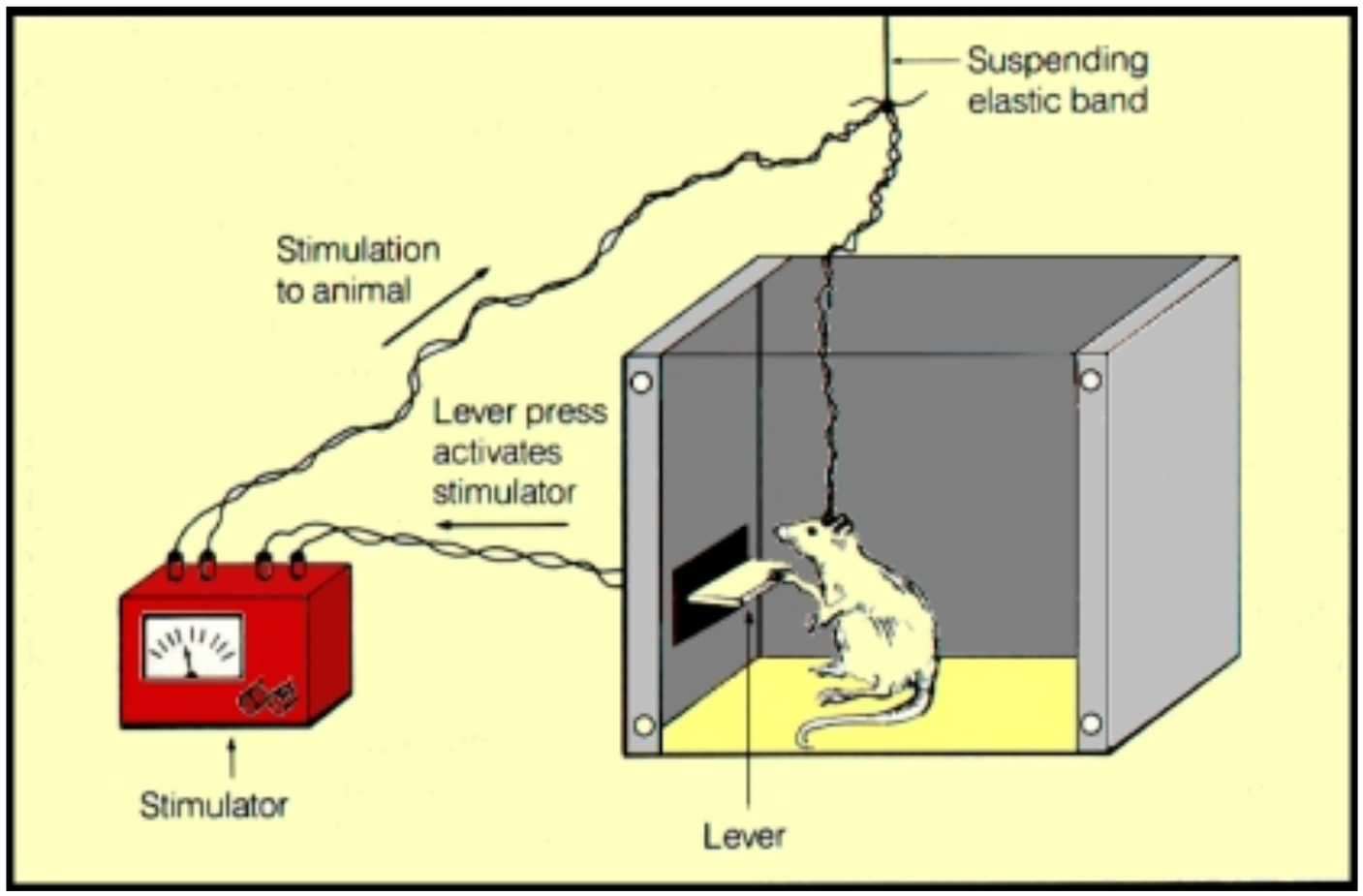
Vandaag

3. **Supervised learning**

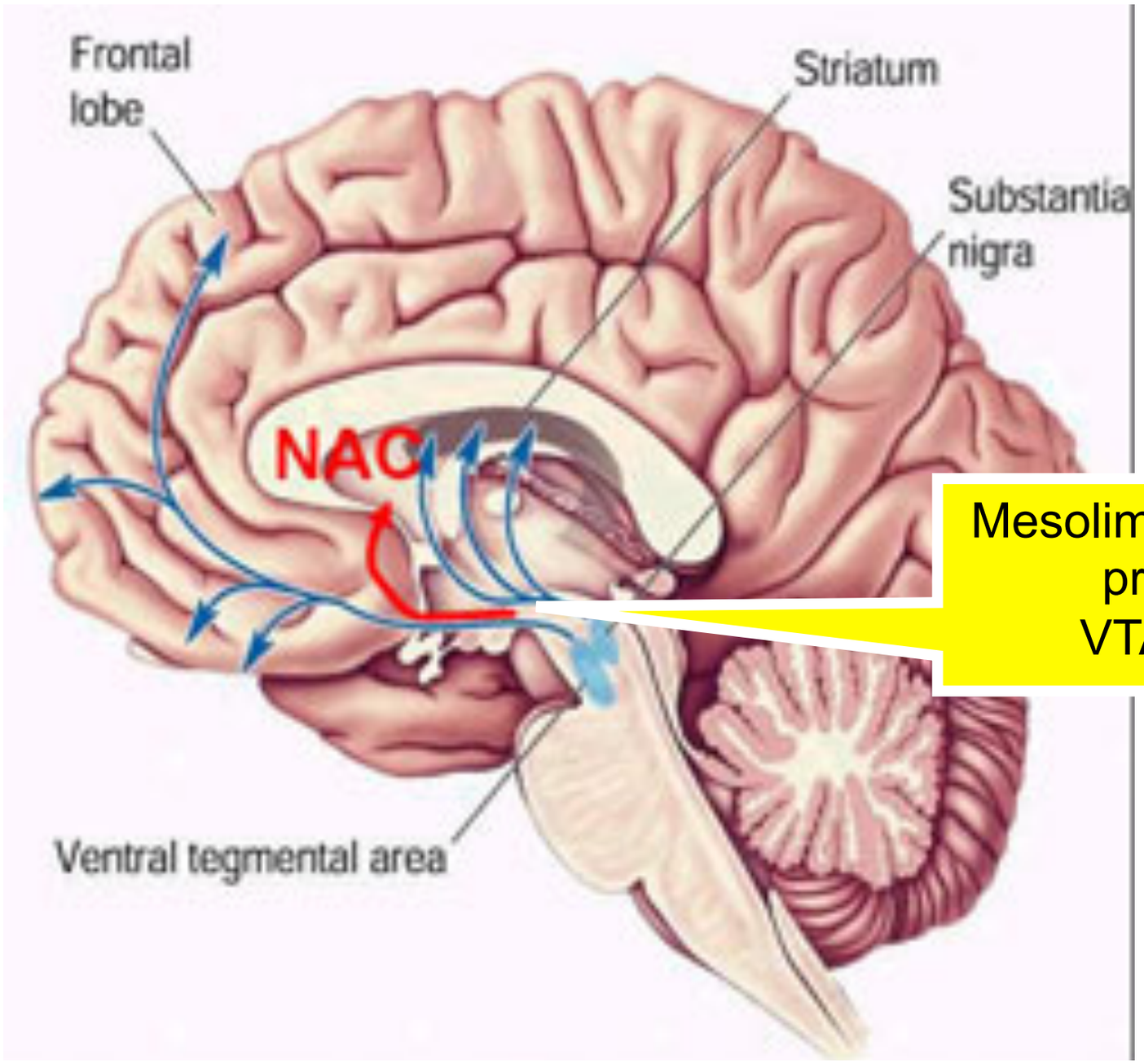
4. **Reinforcement learning**
    - functie van dopamine in de hersenen
        - herhaling: klassiek conditioneren
    - basal ganglia: action selection
        - exploratie en exploitatie
    - cortex: predictive coding

Beloning in de hersenen

Olds & Milner, 1953/54: VTA, **basal ganglia**



Dopamine (DA)!


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

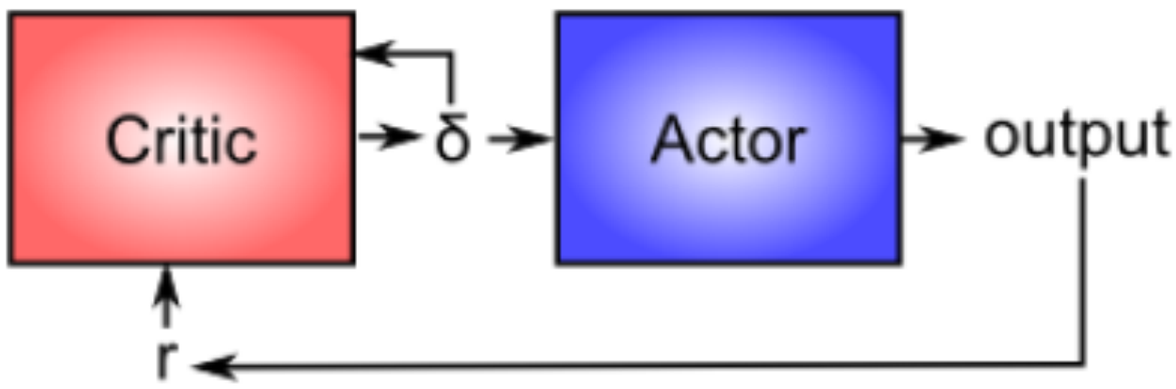
Actor-critic model

- Tot nu toe: systeem leren voorspellen (associëren) input met reward

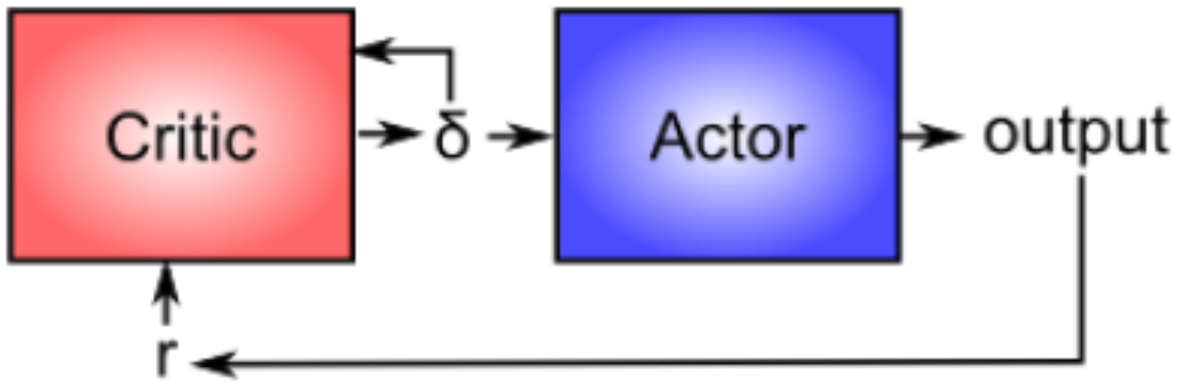
- Maar hoe beslis ik wat ik moet doen op basis van mijn voorspelling van reward?

- Oftewel: hoe leer ik beter fietsen als de enige feedback die heb vallen en me pijn doen is?

- Actor-critic model



Actor-critic model

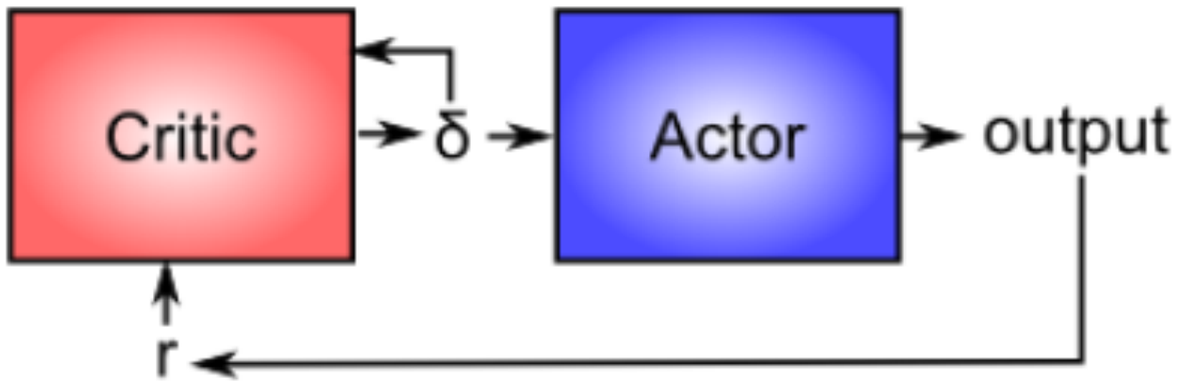


- **critic**: vergelijkt beloning r met verwachte beloning, en maakt een prediction error δ

- **actor**: selecteert gedrag (motor output) gebaseerd op sensory input en verwachte beloning

- zowel actor als critic leren met behulp van prediction error δ

Actor-critic model



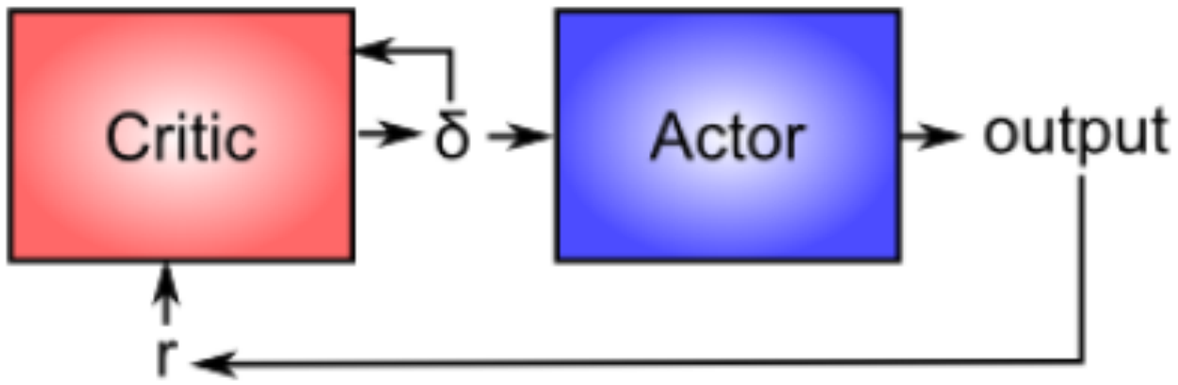
- **critic**: temporal difference learning

- **actor**: ‘policy improvement’ learning:
  - houd bij voor elke actie volgend op elke stimulus wat de waarde is: value tabel!

Value tabel

stimulus→ actie ↓	stim 1	stim 2	…	stim n
actie 1	3	-2	…	0
actie 2	0	1	…	-4
…	…	…	…	…
actie n	-1	2	…	4

Actor-critic model



- **critic**: temporal difference learning

- **actor**: ‘policy improvement’ learning:
  - houd bij voor elke actie volgend op elke stimulus wat de waarde is: value tabel!
  - kies actie op basis van deze tabel
  - update de waarden in de matrix met prediction error van critic

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

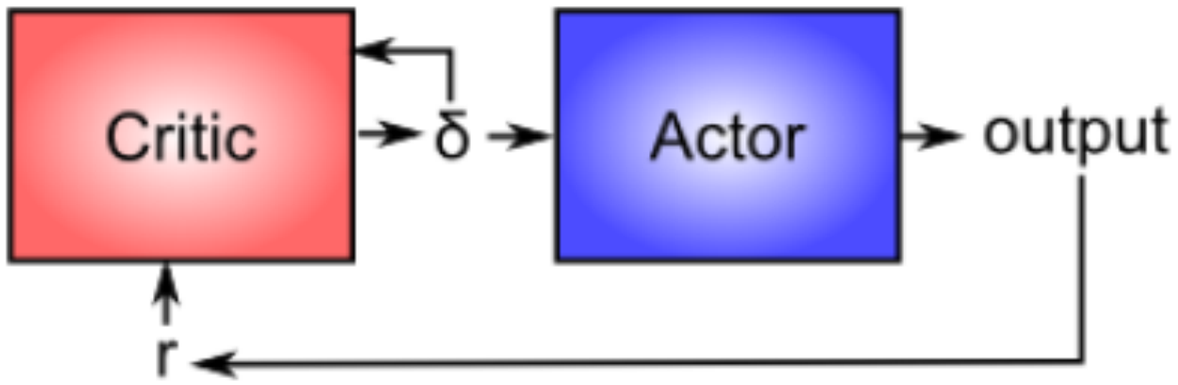
Vandaag

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

Exploratie - exploitatie

- De wereld verandert, rewards blijven niet altijd gelijk
- Value tabel alleen updaten met acties die ik uitvoer
- **Dilemma** voor gegeven stimulus
    1. Kies ik actie met hoogste reward? **Exploitatie**
        - Maar dan weet ik niet of andere actie misschien inmiddels meer oplevert
    2. Kies ik willekeurige andere actie? **Exploratie**
        - Maar dan zou ik lage beloning of straf kunnen krijgen
- Dit gedrag wordt meestal gemodelleerd door één parameter tussen 0 (alleen exploitatie) en 1 (alleen exploratie)

Basal Ganglia == actor?



(te) simpel model

- **critic**: VTA
- **prediction error δ:** dopamine
- **actor**: ontvangt error signaal van VTA, selecteert output: basal ganglia?

Go-NoGo taak

- Bij stimulus 1: ’Doe iets’ voor reward (druk op knop als je appel ruikt)

- Bij stimulus 2: ’Doe niets’ (doe niets als je tomaat ruikt)

- Gebruikt voor: impulsiviteit, attention, etc

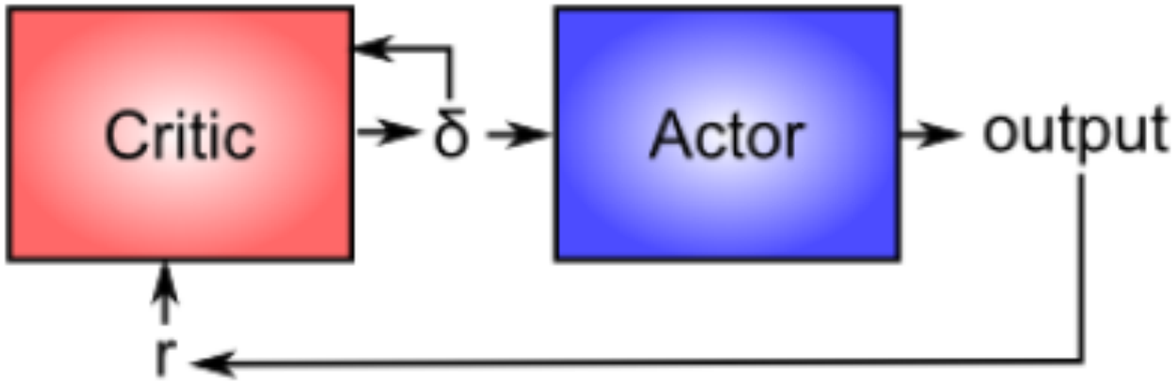
Basal Ganglia: action selection

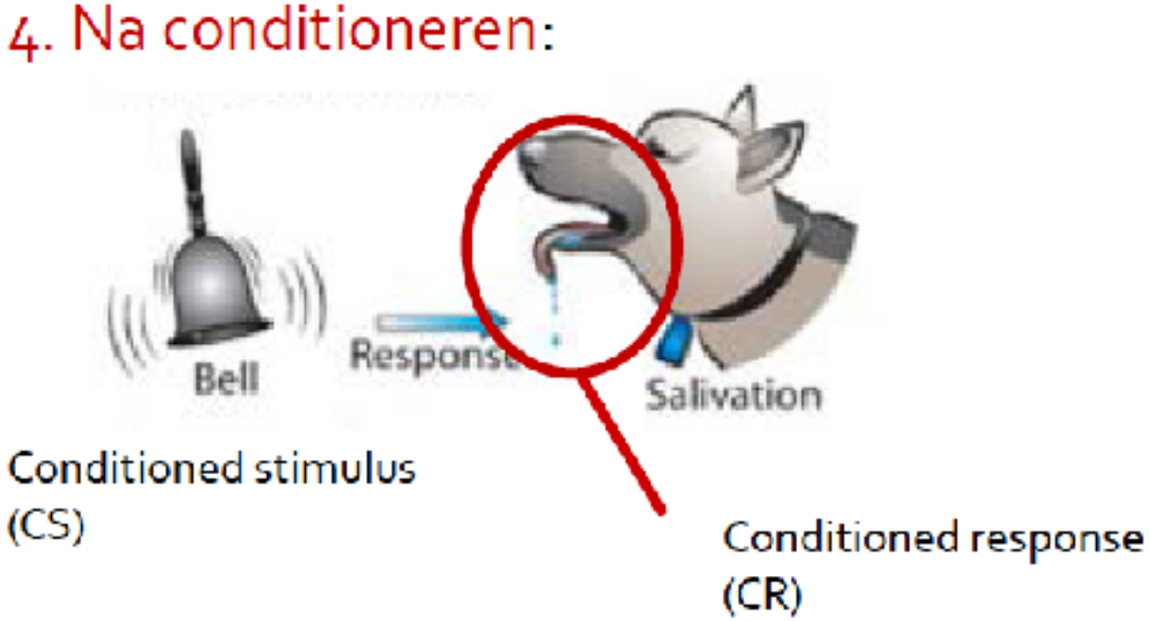
2 verschillende pathways (striatum):

1. D1 receptoren, dopamine excitatoir effect (direct pathway): Go

2. D2 receptoren, dopamine inhibitior effect (indirect pathway): NoGo

actor-critic model



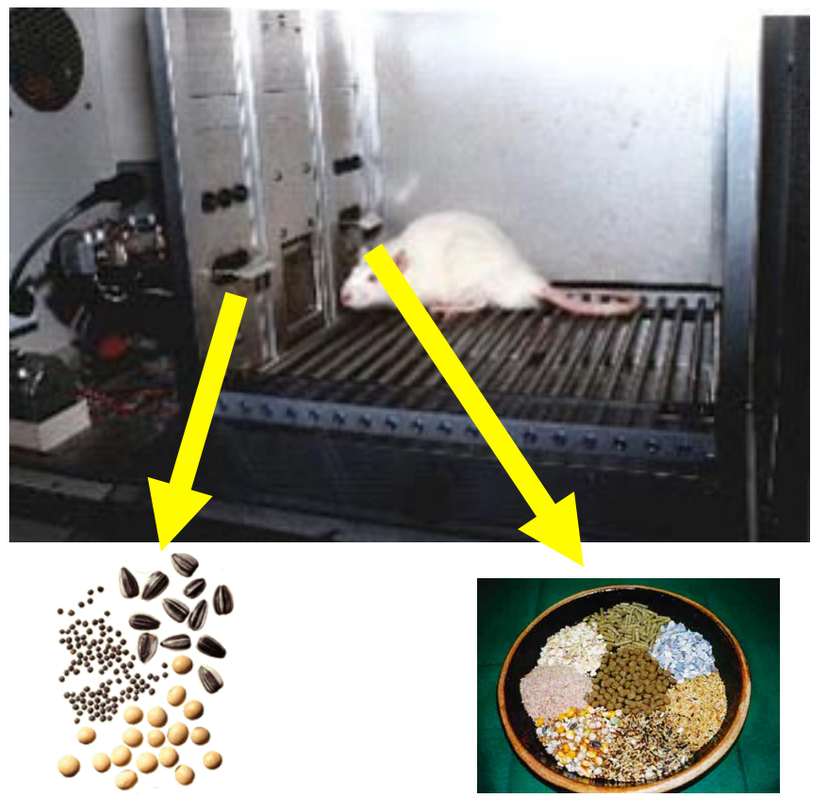


Klassiek conditioneren:

- wel critic, geen actor

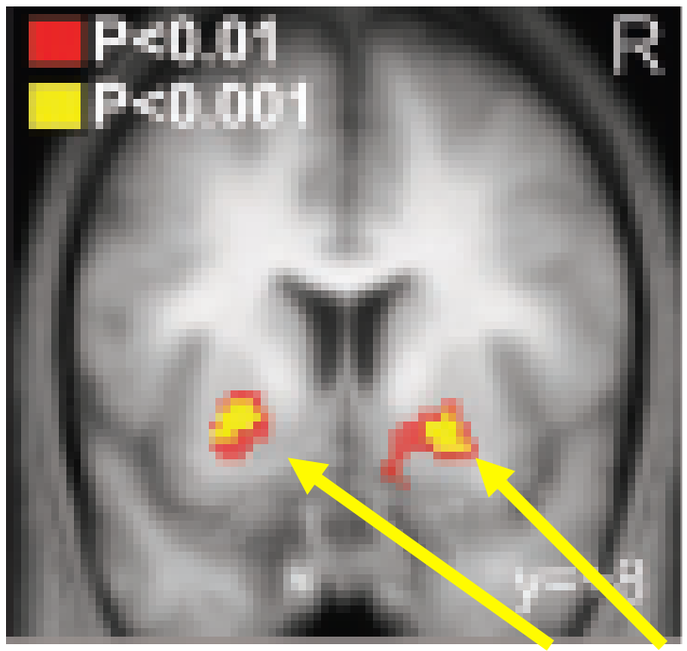
Actieve taak (operant conditioning)

- critic en actor



actor-critic model

Klassiek conditioneren



Operant conditioneren

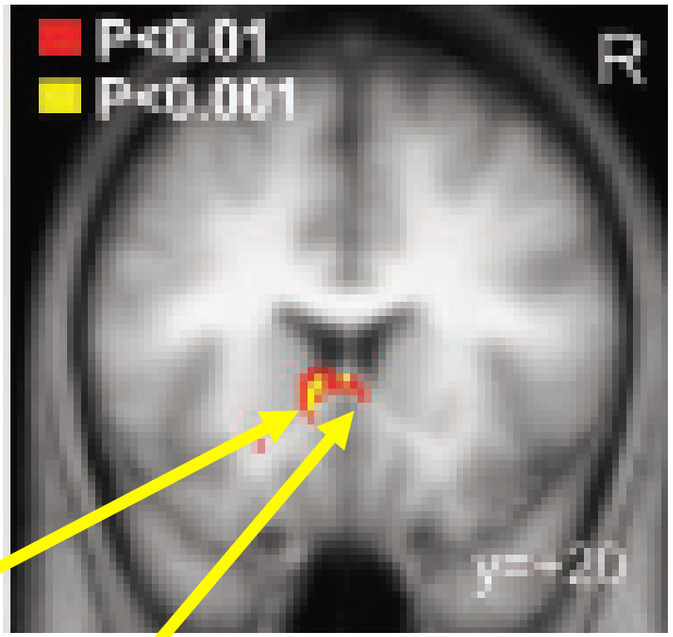
- NAc (ventral striatum, basal ganglia) actief bij beide taken

O'Doherty et al., 2004

actor-critic model

Klassiek conditioneren

Operant conditioneren



- NAc (ventral striatum, basal ganglia) actief bij beide taken
- dorsal striatum (basal ganglia) alleen actief bij operant conditioning

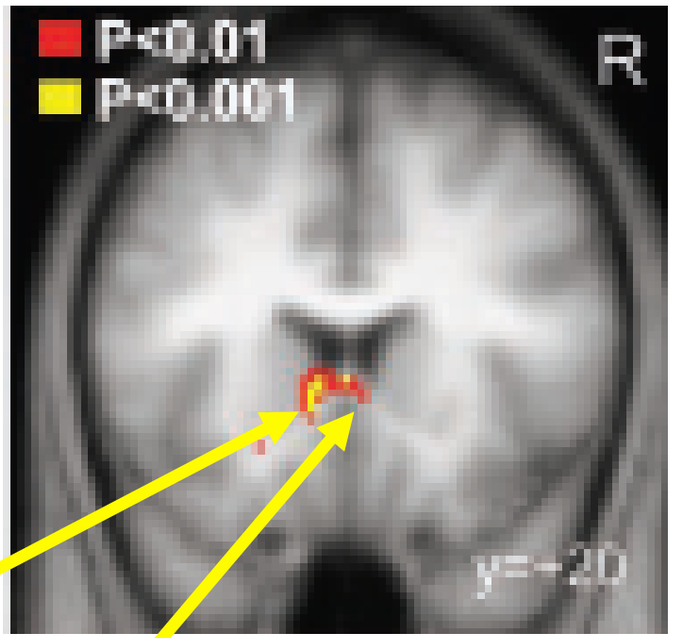
Dus wat hoort waarbij?

O'Doherty et al., 2004

actor-critic model

Klassiek conditioneren

Operant conditioneren



- NAc (ventral striatum, basal ganglia) actief bij beide taken (**critic?**)
- dorsal striatum alleen actief bij operant conditioning (**actor?**)

Dus wat hoort waarbij?

O’Doherty et al., 2004

Conclusie

- Het lijkt of VTA een dopamine/**prediction error** naar BG stuurt
- Maakt dat VTA == **critic**, BG == **actor**?
- Structuur BG lijkt hiervoor gemaakt (D1/Go en D2/NoGo pathways)
- Maar ligt complexer: delen BG lijken zowel critic als actor

Programma

College 1

1. **Unsupervised learning**

2. **Supervised learning**

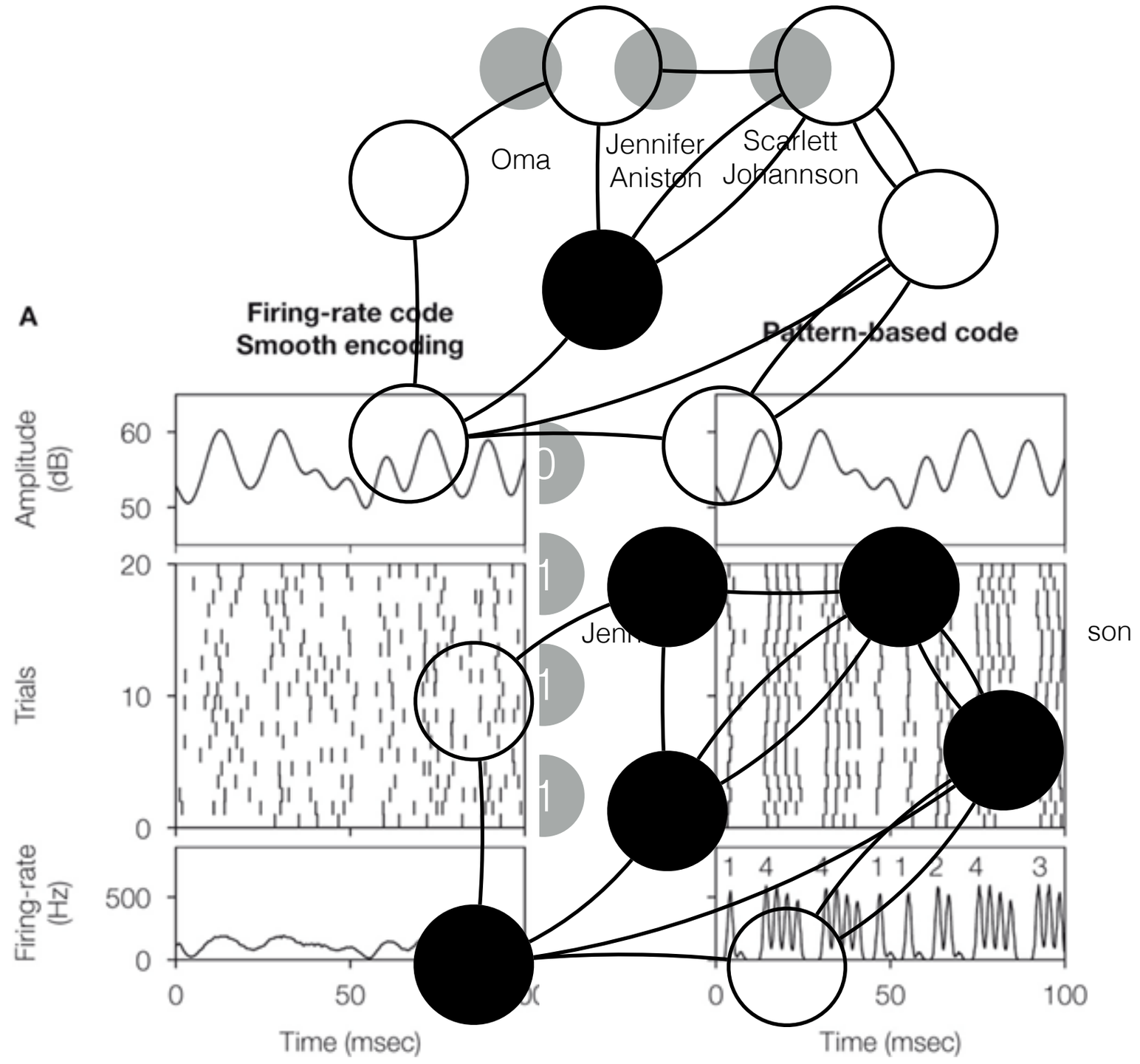
Vandaag

3. **Supervised learning**

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

Neural Coding

- local
- distributed
  - sparse
  - dense
- rate
- temporal



Predictive coding

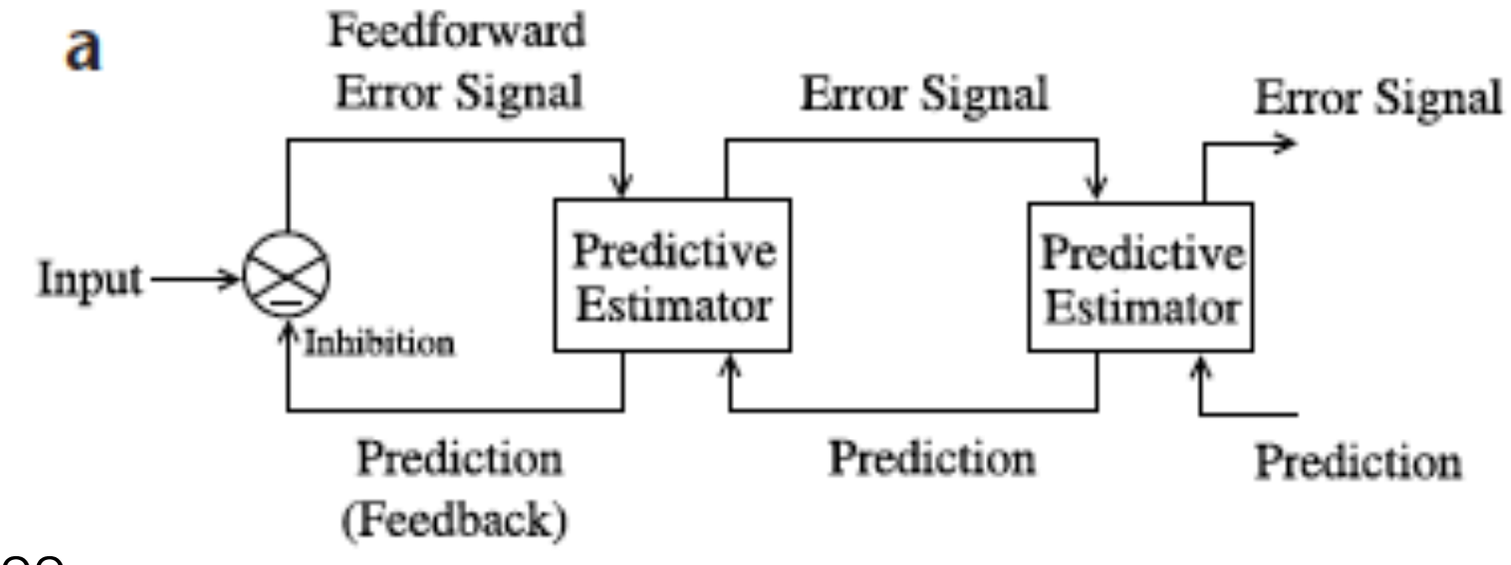
- Tot nu toe: een neuron/netwerk probeert zo goed mogelijk de input die hij ontvangt weer te geven: **representatie**

- Dit is echter erg **inefficiënt**: netwerken die dit doen zijn voortdurend bezig informatie door te geven die niet wezenlijk verandert

- Alternatief: **predictive coding**

- Hersenen gebruiken een model van de wereld, voorspellen verwachte input, en geven **prediction errors** door

Predictive coding



Rao & Ballard 1999

- Input wordt vergeleken met **voorspelling** (**feedback**) van hoger gelegen lagen

- Een **prediction error** wordt vooruit gestuurd (**feed-forward**)

- Recursief: De volgende ‘laag' vergelijkt deze prediction error weer met de voorspelling van de laag die daarop volgt

Predictive coding

- Dus: netwerk is zowel feed-forward als feedback (realistischer dan puur feed-forward perceptron)

- Netwerk reageert alleen op onverwachtse input, dus **efficiënter** (hoeft bekende input niet te coderen)

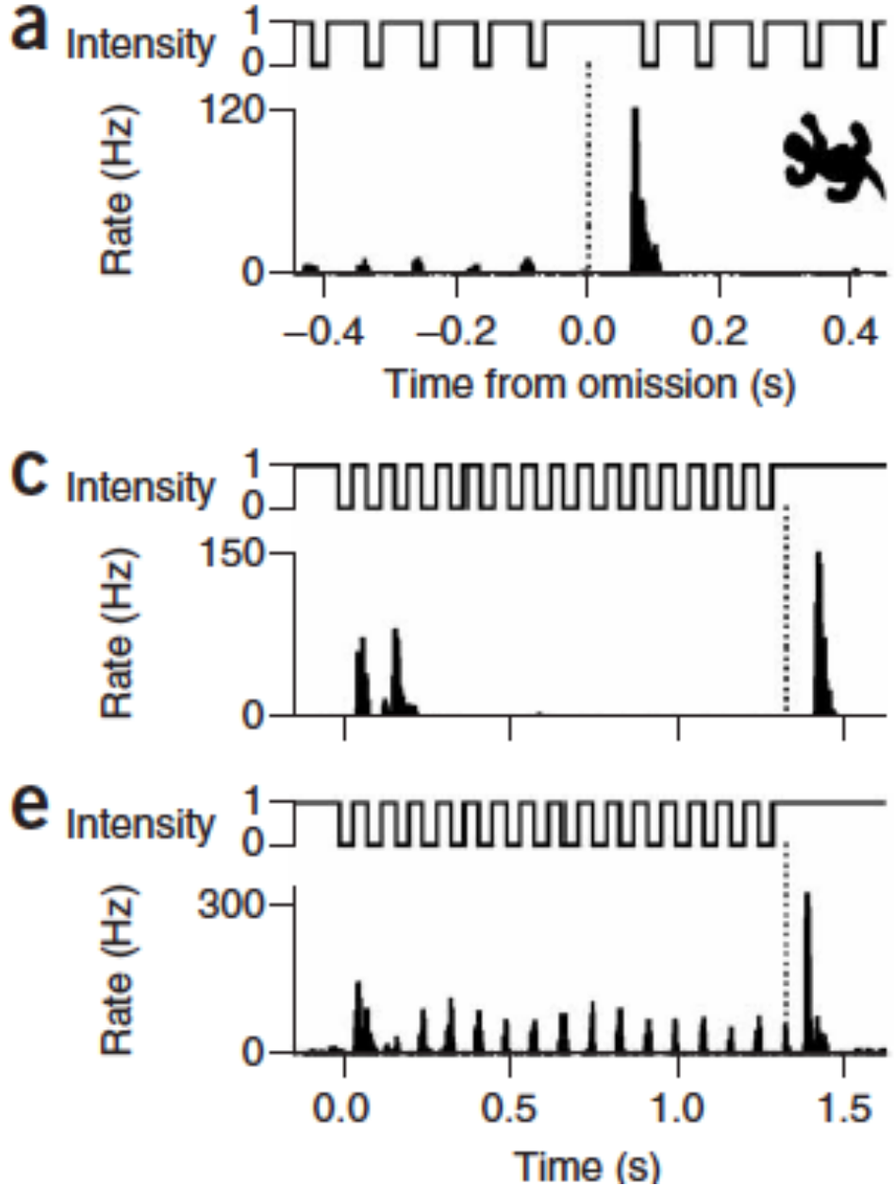
- het model dat de voorspellingen genereert wordt voortdurend geüpdatet dmv prediction error

- Oude theorie: eerste formulering door Helmholtz (1863)!

Predictive coding in cortex?

Er lijkt bewijs te zijn dat corticale neuronen vooral reageren op ‘nieuwe' stimuli:

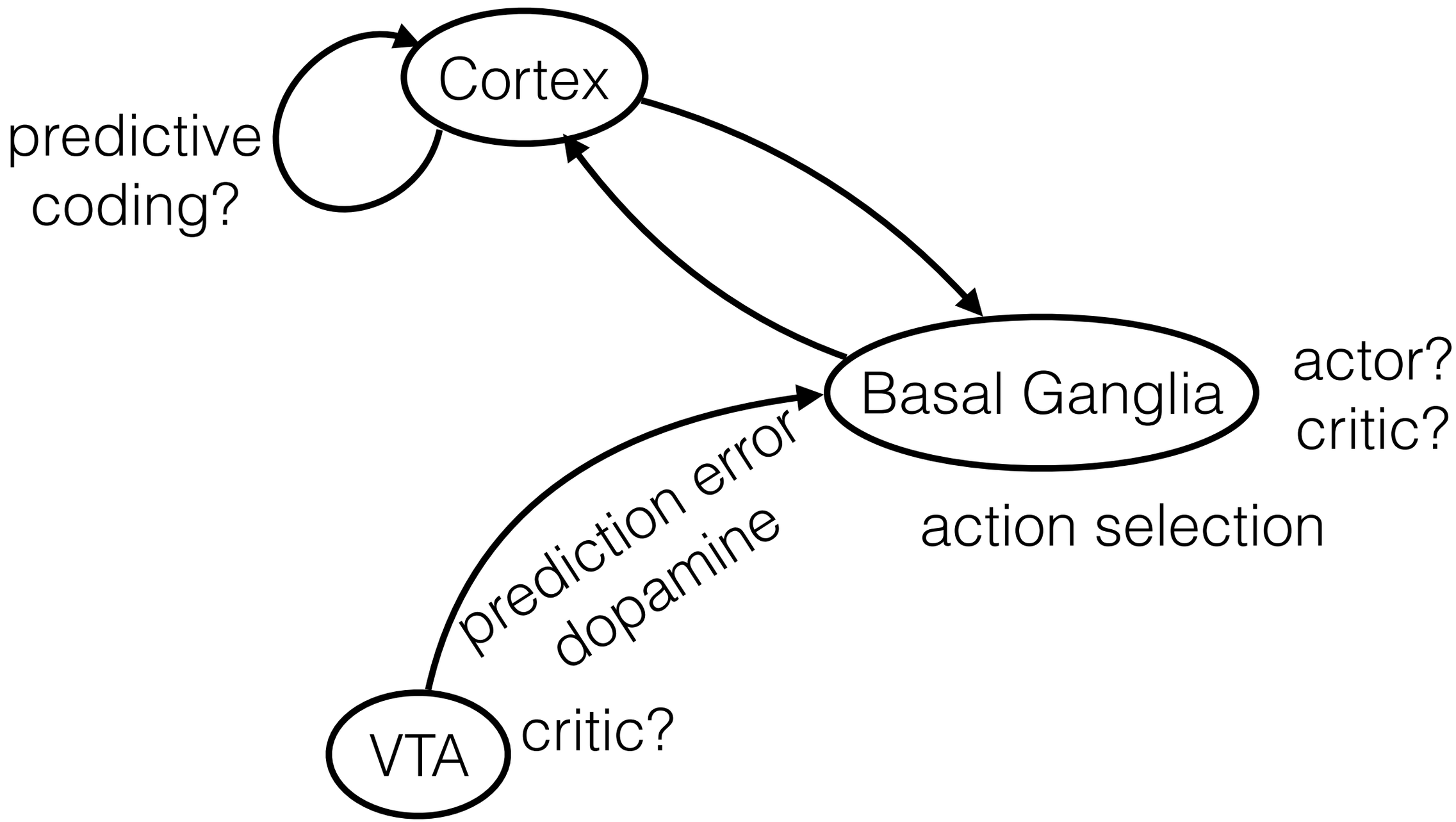
- hiernaast: neuron reageert op ‘weglaten' stimulus



Schwartz et al., 2007

predictive coding	**feed-forward coding**
afwijkingen input worden weergegeven **(prediction error)**	input signaal wordt weergegeven
feedback verbindingen geven voorspelling hogere lagen	geen feedback verbindingen
feed-forward verbindingen geven prediction error	feed-forward verbindingen geven representaties door
hersenen hebben model van de wereld	hersenen hebben geen model van de wereld, filteren alleen
efficiënt	minder efficiënt

Samenvatting
Reinforcement Learning



Programma

College 1

1. **Unsupervised learning**
    - LTP/LTD modellen: rate-based Hebbian learning
        - voorbeeld: Leren in Hopfield netwerk
    - Andere Hebb-achtige regels
    - STDP: spike-based Hebbian learning

2. **Supervised learning**
    - perceptron regel

Vandaag

3. **Supervised learning**
    - delta regel
    - error-backpropagation

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

Voorbeeld tentamenvragen

- Is de covariance rule van unsupervised learning stabiel?
  - Nee

- Kun je een temporal code leren met de BCM regel?
  - Nee

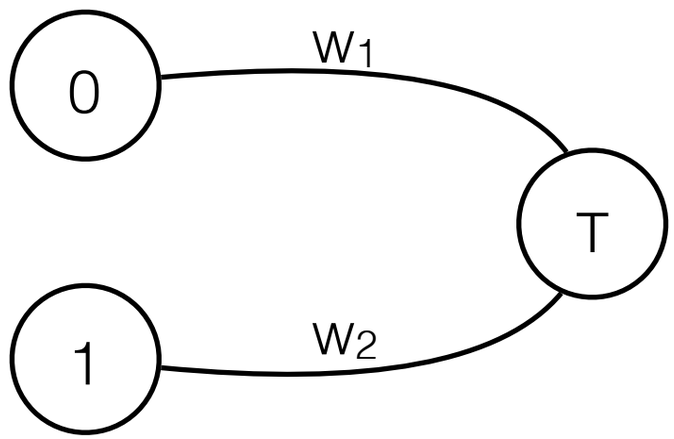
- Noem een probleem van error-backpropagation

- In welk hersengebied denkt men dat de ‘actor' van het actor-critic model zit?
  - Basal Ganglia

- Noem een verschil tussen feed-forward en predictive coding

- Gegeven een leerregel en een netwerk, reken de nieuwe gewichten uit

$$\Delta w_i = \gamma(oa_i - \alpha o^2 w_i)$$



Tot slot

- Als je dit soort modellen leuk vindt:
    - signaalanalyse
    - computational cognitive neuroscience
- Als je reinforcement learning interessant vindt:
    - systems neuroscience

Literatuur

Online (bij Leren & Geheugen)

- Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). Neuronal Dynamics. Staat online: http://neuronaldynamics.epfl.ch/online/index.html

- O'Reilly, R. C., Munakata, Y., Franks, M. J., Hazy, T. E., & Contributers. (2012). Computational cognitive neuroscience (first.). Wiki Book, online op http://ccnbook.colorado.edu

Papier (extra materiaal)

- Dayan & Abbott: 'Theoretical Neuroscience'

- Izhikevich: 'Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting'

- Rieke et al.: 'Spikes: exploring the neural code'