

Reinforcement learning in neuroscience - III

Rita Almeida

PhD course 3045, VT 2018



Evidence for action selection in the brain

Evidence for different algorithms:

- ▶ Q-learning
- ▶ SARSA
- ▶ Actor/Critic method

Evidence for Q-learning

Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards

Matthew R Roesch^{1,4}, Donna J Calu^{2,4} & Geoffrey Schoenbaum^{1,3}

NATURE NEUROSCIENCE VOLUME 10 | NUMBER 12 | DECEMBER 2007

Task

Experiment with rats: forced and free-choice trials.

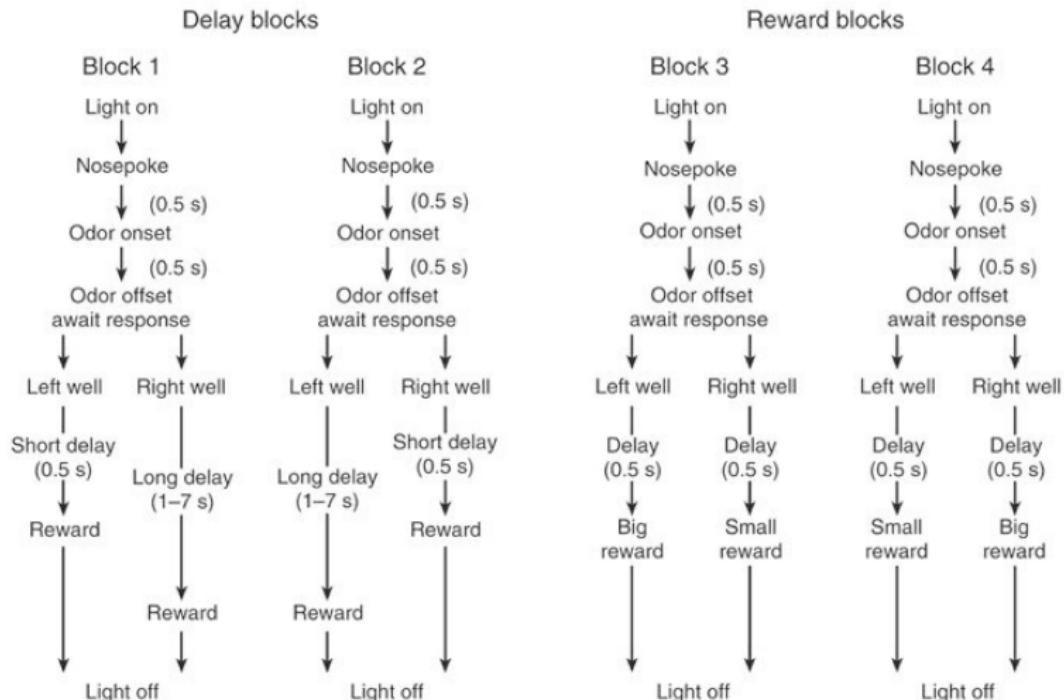


Figure from Roesch et al. 2007.

Results

- ▶ Rats learn the contingencies in the successive blocks: on free choice trials they tend to choose the more immediate and larger rewards.
- ▶ Dopamine neurons in VTA develop selectivity for
 - ▶ cue: firing rate is larger for cues predicting more immediate and larger reward
 - ▶ reward

Evidence for Q-learning

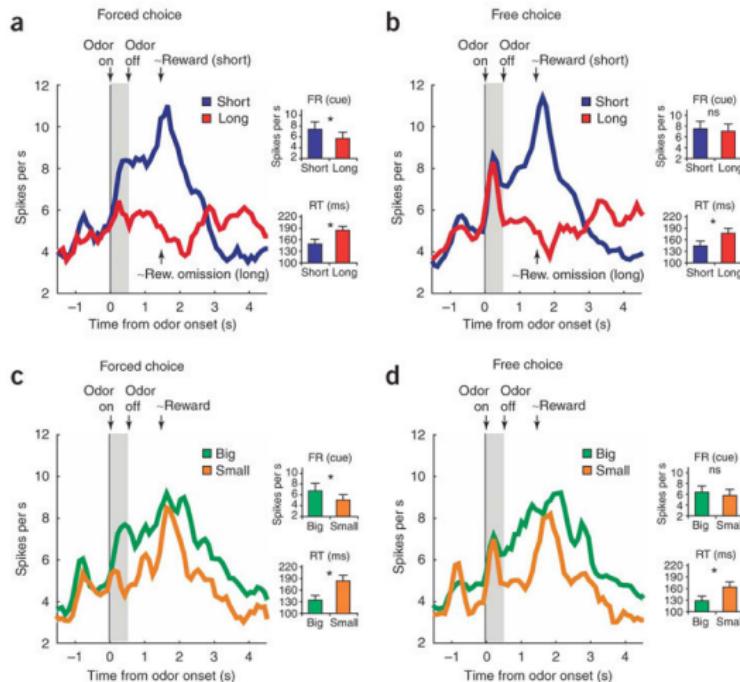


Figure from Roesch et al. 2007.

Cue-evoked activity on free-choice trials reflects the value of the best option available.

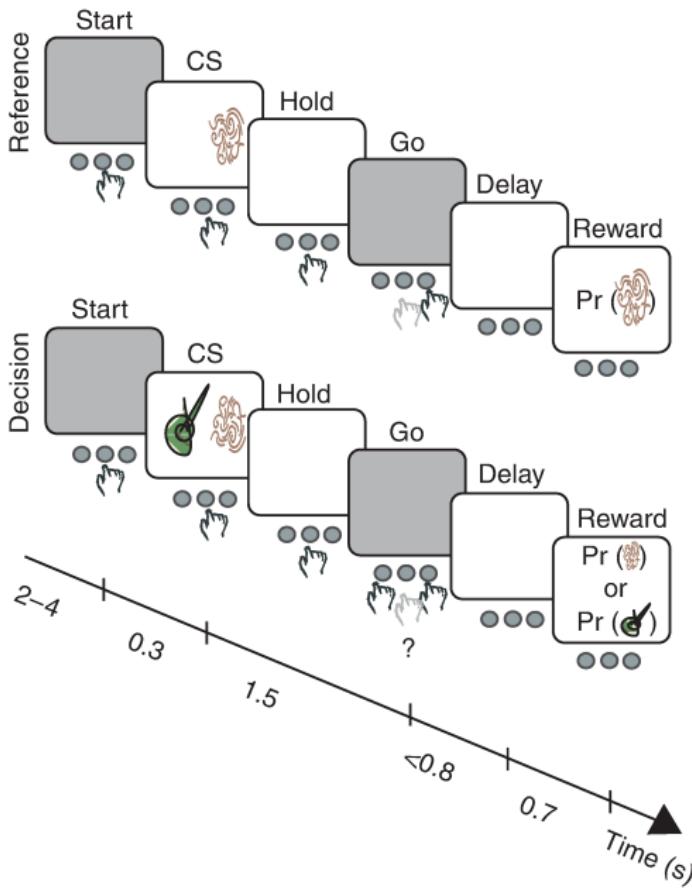
Evidence for SARSA

Midbrain dopamine neurons encode decisions for future action

Genela Morris^{1,2}, Alon Nevet², David Arkadir², Eilon Vaadia^{1,2} & Hagai Bergman^{1,2}

NATURE NEUROSCIENCE VOLUME 9 | NUMBER 8 | AUGUST 2006

Task



Results

Probability of choosing one alternative in the decision trials as a function of quantities derived from the instructed trials:

- ▶ relative frequency of reward for that alternative,
- ▶ relative dopamine response.

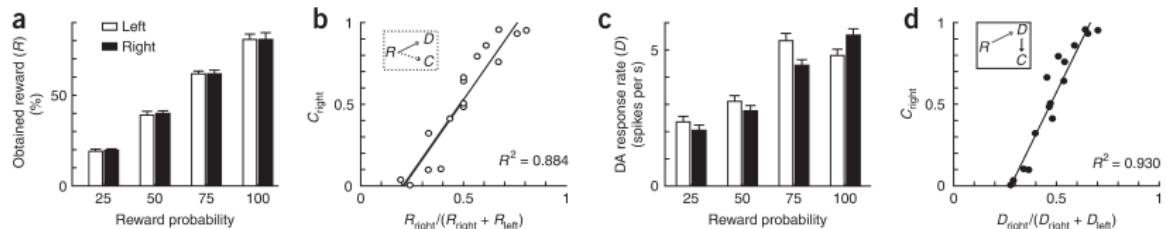
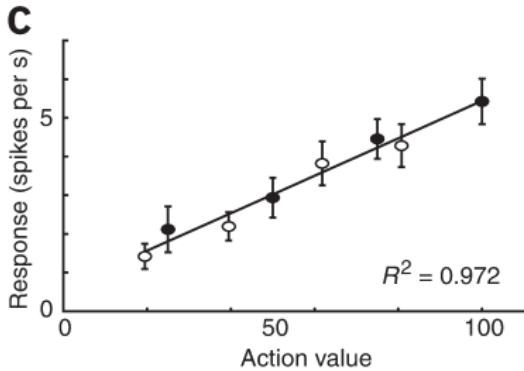
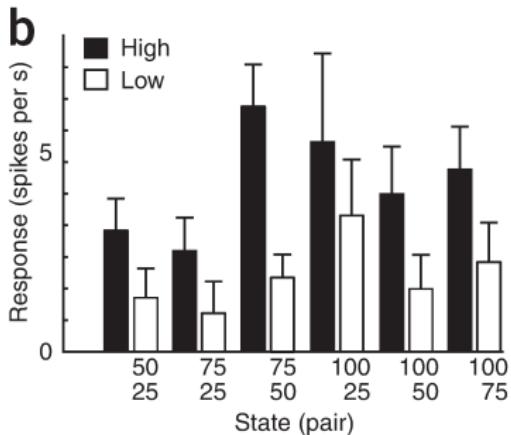


Figure from Morris et al. 2006.

Relative dopamine response predicts better the choice probabilities.

Evidence compatible with SARSA

- ▶ Dopamine responses to the stimulus pairs in decision trials as a function of action values. Responses are separated by action taken.
- ▶ Dopamine signal is higher when future choice is associated with higher reward probability.
- ▶ Action taken, but not action discarded, has a significant effect on dopamine signal.



Evidence for Actor/Critic

Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning

John O'Doherty,^{1*} Peter Dayan,² Johannes Schultz,¹
Ralf Deichmann,¹ Karl Friston,¹ Raymond J. Dolan¹

16 APRIL 2004 VOL 304 SCIENCE www.sciencemag.org

Evidence for Actor/Critic

fMRI experiment in Humans.

Conditions:

- ▶ Instrumental task - probabilistic (30%, 60%)
 - ▶ reward (2 stimuli)
 - ▶ neutral (2 other stimuli)
- ▶ Pavlovian task - probabilistic (30%, 60%)
 - ▶ reward
 - ▶ neutral

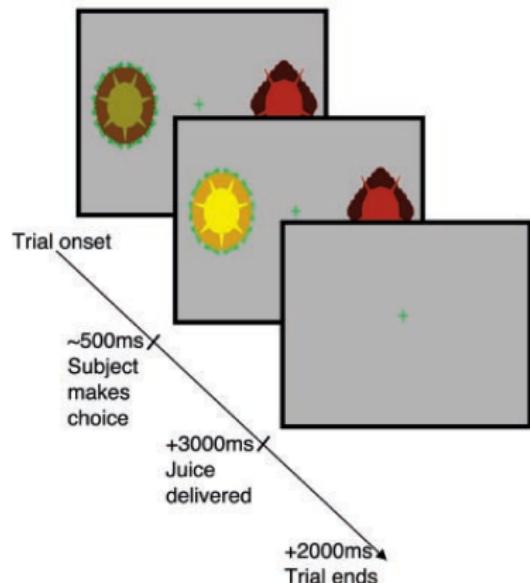


Figure from O'Doherty et al. 2004.

BOLD fMRI to study reinforcement learning

- ▶ Data analysis idea: look for signals that correlate with the predicted bold response.

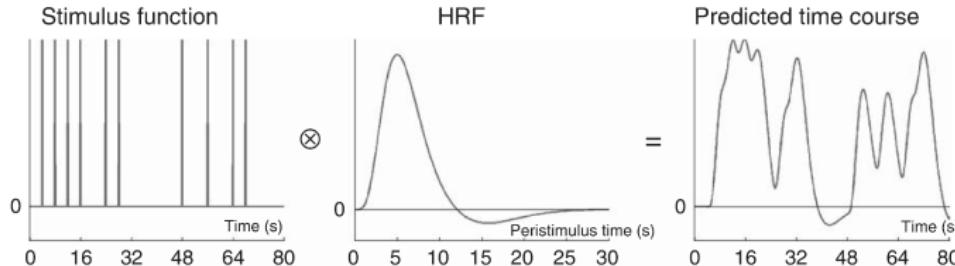
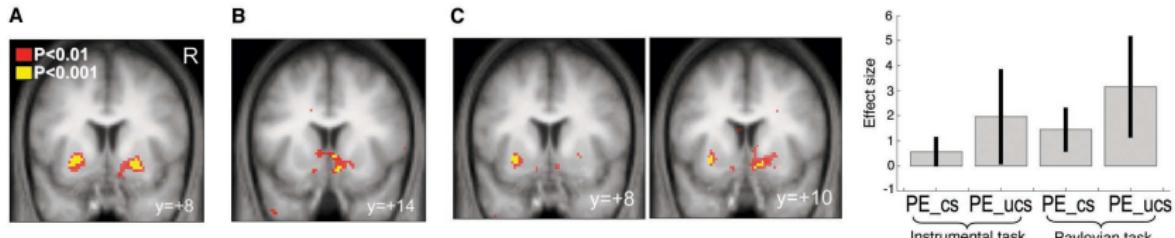


Figure adapted from *Statistical Parametric Mapping* Friston et al. 2007

- For reinforcement learning studies
 - ▶ Estimate model parameters to fit the behavior.
 - ▶ Derive time changing variables describing quantities as: value of actions, prediction error.
 - ▶ Find if/where are signals that correlate with such variables.

Evidence for Actor/Critic

Critic - ventral striatum



Actor - dorsal striatum

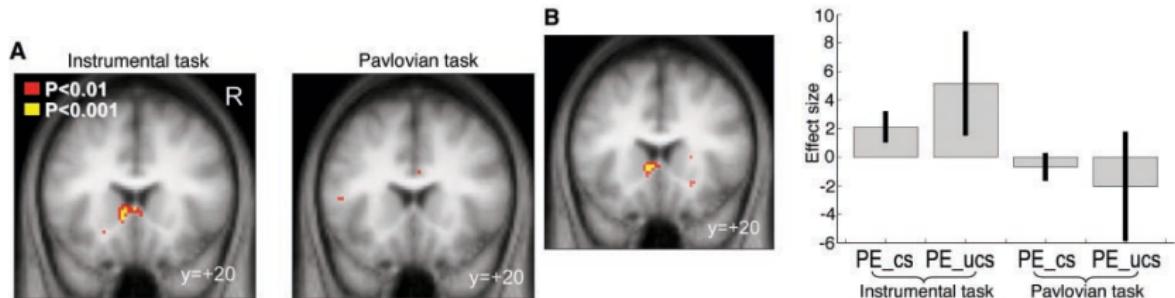


Figure from O'Doherty et al. 2004.

Causal role of dopamine neurons activity in RL

A causal link between prediction errors, dopamine neurons and learning

Elizabeth E Steinberg^{1,2,11}, Ronald Keiflin^{1,11}, Josiah R Boivin^{1,2}, Ilana B Witten^{3,4}, Karl Deisseroth^{5–8} & Patricia H Janak^{1,2,9,10}

VOLUME 16 | NUMBER 7 | JULY 2013 NATURE NEUROSCIENCE

Causal role of dopamine neurons activity in RL

Blocking

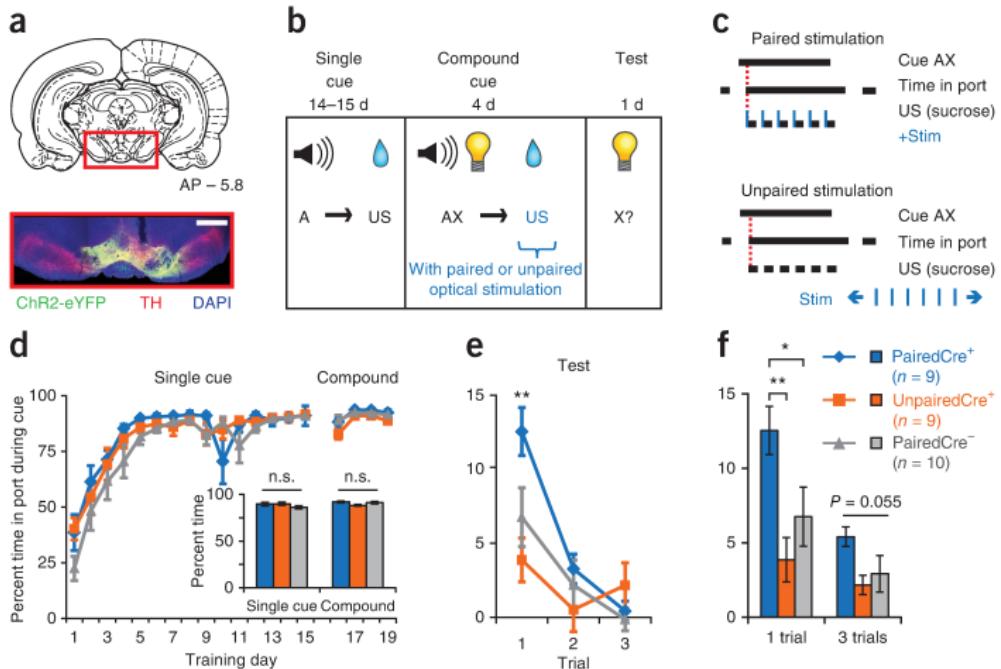


Figure from Steinberg et al. 2013.

Causal role of dopamine neurons activity in RL Extinction

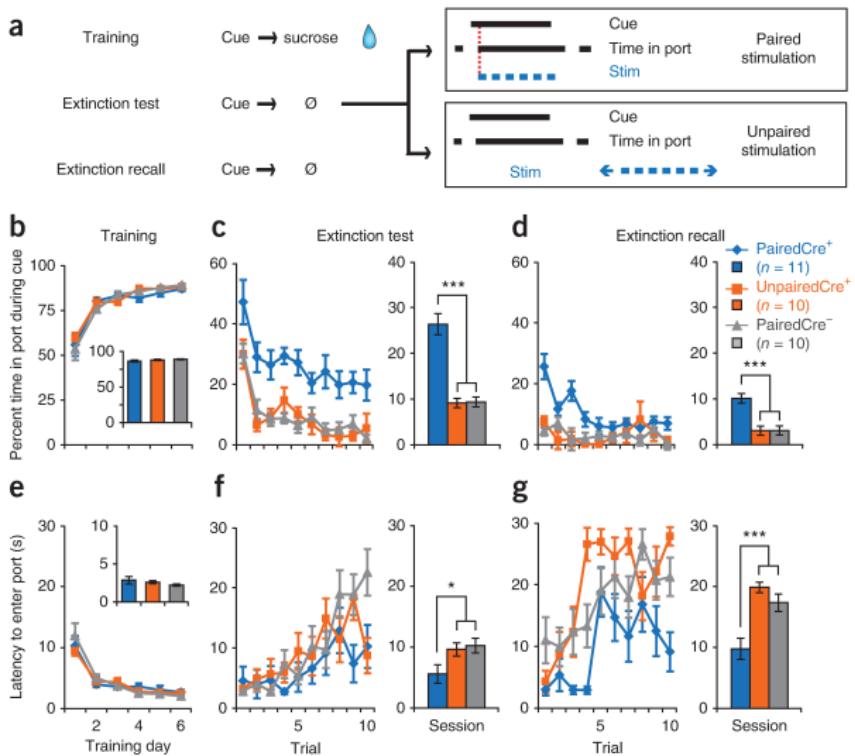


Figure from Steinberg et al. 2013.

Other topics

- ▶ Habitual versus goal-directed behaviour
- ▶ Predictive versus motivational signaling
- ▶ Model-free learning versus model-based learning
- ▶ Disorders: compulsive behaviors, ADHD, schizophrenia
- ▶ Pharmacological hypotheses
- ▶ Substance abuse
- ▶ Changes over life-span: impulsivity during adolescence
- ▶ Tonic dopamine: response vigor, latency and rate
- ▶ ...

References

- ▶ G. Morris et al. Midbrain dopamine neurons encode decisions for future action, *Nature Neuroscience*, 2006.
- ▶ J. O'Doherty et al., Dissociable roles of ventral and dorsal Striatum in instrumental conditioning, *Science*, 2004.
- ▶ M. Roesch et al., Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards, *Nature Neuroscience*, 2007.
- ▶ E. Steinberg et al., A causal link between prediction errors, dopamine neurons and learning, *Nature Neuroscience*, 2013.