A Computational Model of the Role of Orbitofrontal Cortex and Ventral Striatum in Signalling Reward Expectancy in Reinforcement Learning

Robert C. Wilson[1], Yuji K. Takahashi[2], Matthew R. Roesch[5], Thomas Stalnaker[2], Geoffrey Schoenbaum[2,3] and Yael Niv[1]

1. Department of Psychology and Neuroscience Institute, Princeton University, 2. Department of Anatomy & Neurobiology, 3. Department of Psychiatry, University of Maryland School of Medicine, 4. University of Maryland School of Medicine 5. Department Psychology, University of Maryland College Park

#404.1/MMM2

Introduction

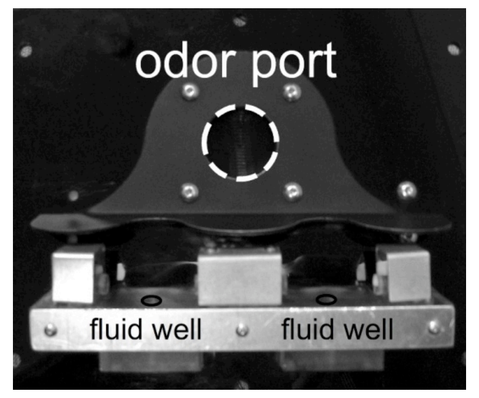
Orbitofrontal cortex (OFC) and ventral striatum (VS) have been implicated in signalling reward expectancies, but their exact roles are unknown.

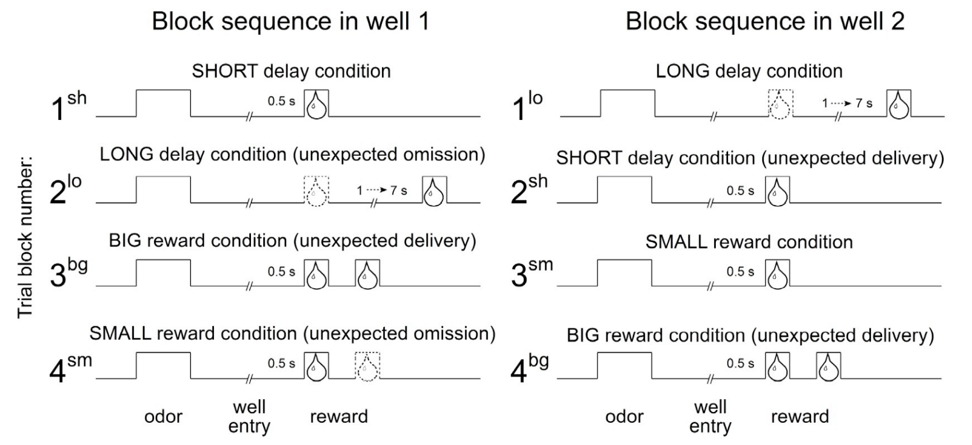
We compare predictions from three different reinforcement learning models to experimental results from Takahashi et al. (presented in the next poster) towards delineating their specific roles.

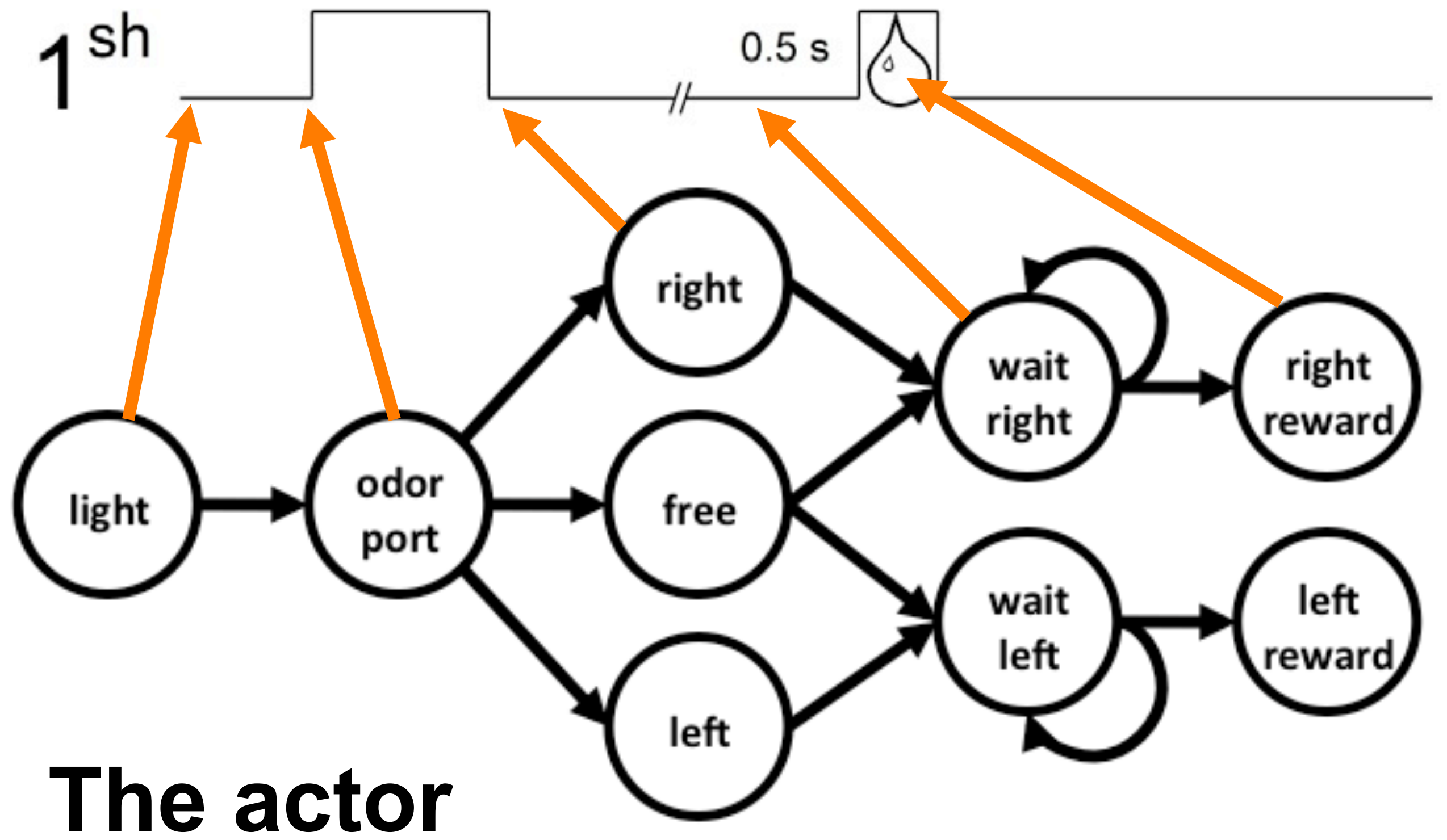
Conclusions

Takahashi et al.'s results are better explained by assuming that ***OFC encodes complex state representations*** (model 3) than by assuming that OFC directly encodes expected values (models 1,2).

The task and representation







The actor

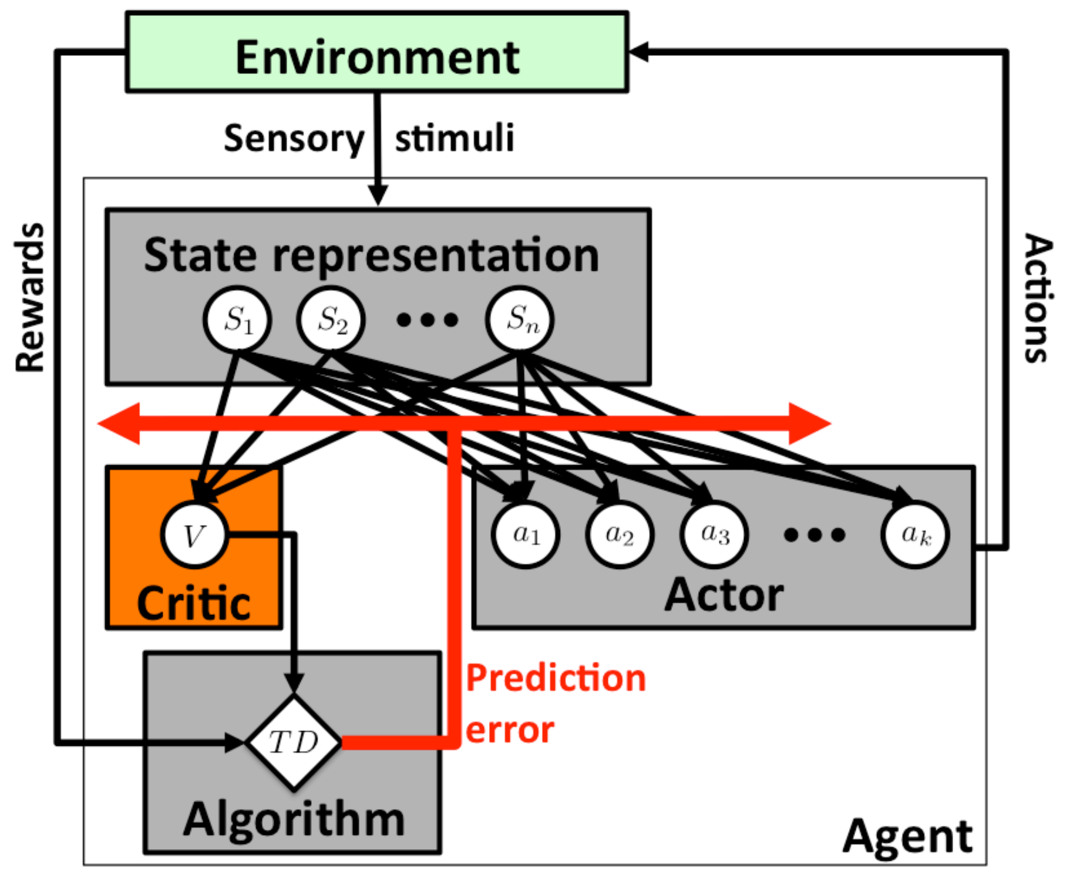
- Same for all models
- Choice probability (on free trials only)

$$p(\text{right}) = \frac{\exp(\beta M_{right})}{\exp(\beta M_{right}) + \exp(\beta M_{left})}$$

- Actor learns preferences according to

$$M_{action} \leftarrow M_{action} + \alpha_M \delta$$

Model 1: OFC is the critic (rather than VS)

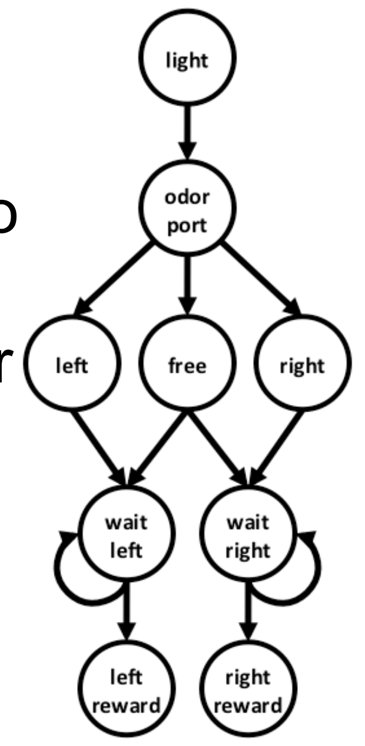


Description

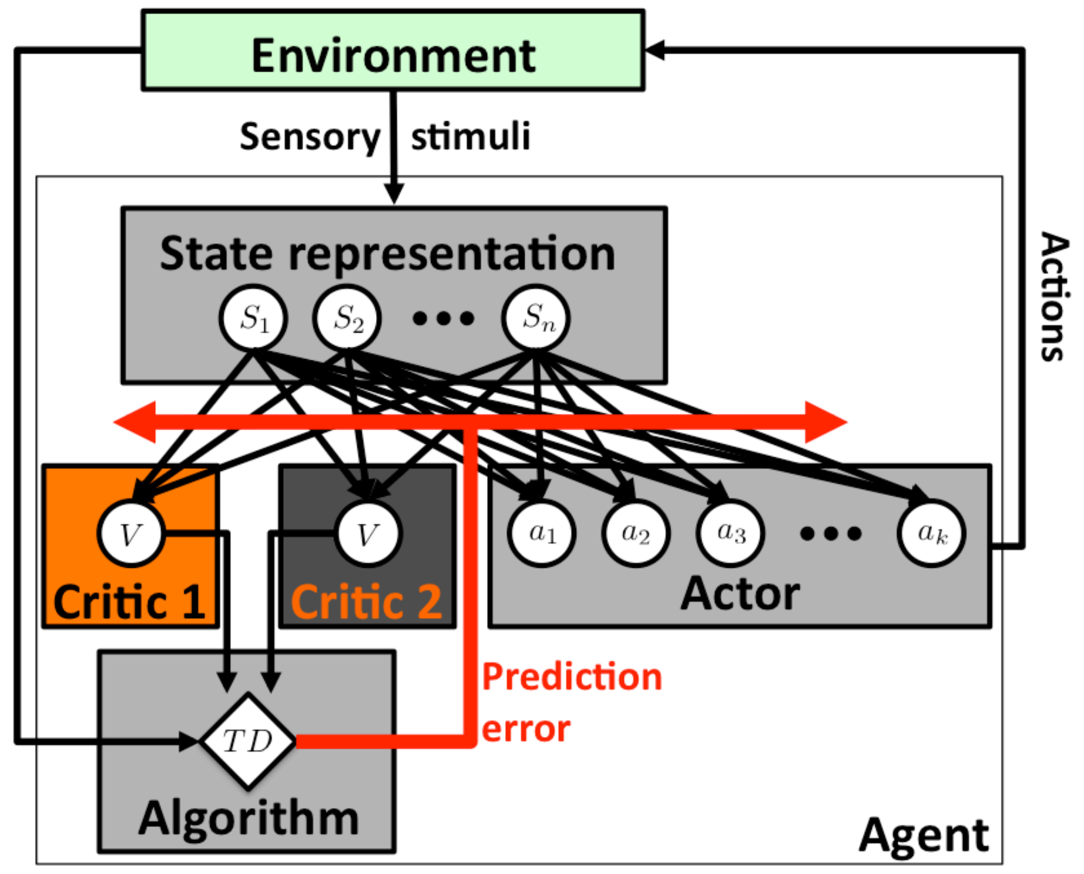
- Traditional 'actor-critic' framework
- OFC as critic, encodes state values
- OFC lesion removes value from prediction error and so rewards are never predicted
- In this model VS does not contribute to prediction error

Learning in the (OFC) critic

- Prediction error: $\delta = r + \gamma V(S') - V(S)$
- Eligibility trace: $e(S) = 1 + \lambda e(S)$
  $e(S') \leftarrow \lambda e(S') \text{ for } S' \neq S$
- Value update: $V(S) \leftarrow V(S) + \alpha e(S)\delta$



Model 2: OFC and VS are competing critics



Description

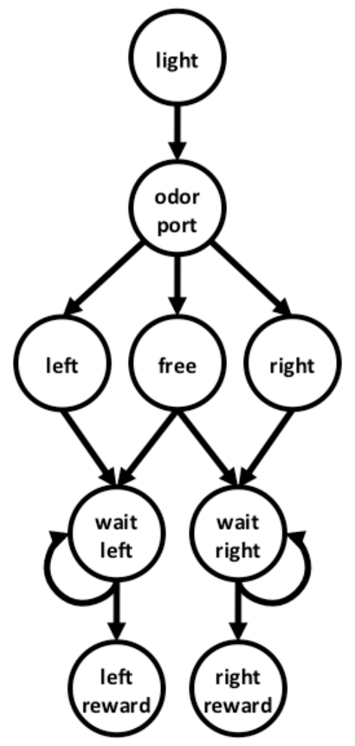
- OFC and VS are both critics. Both encode state values
- OFC has high learning rate, VS low learning rate
- OFC or VS lesion leaves other critic intact

Learning in the two critics

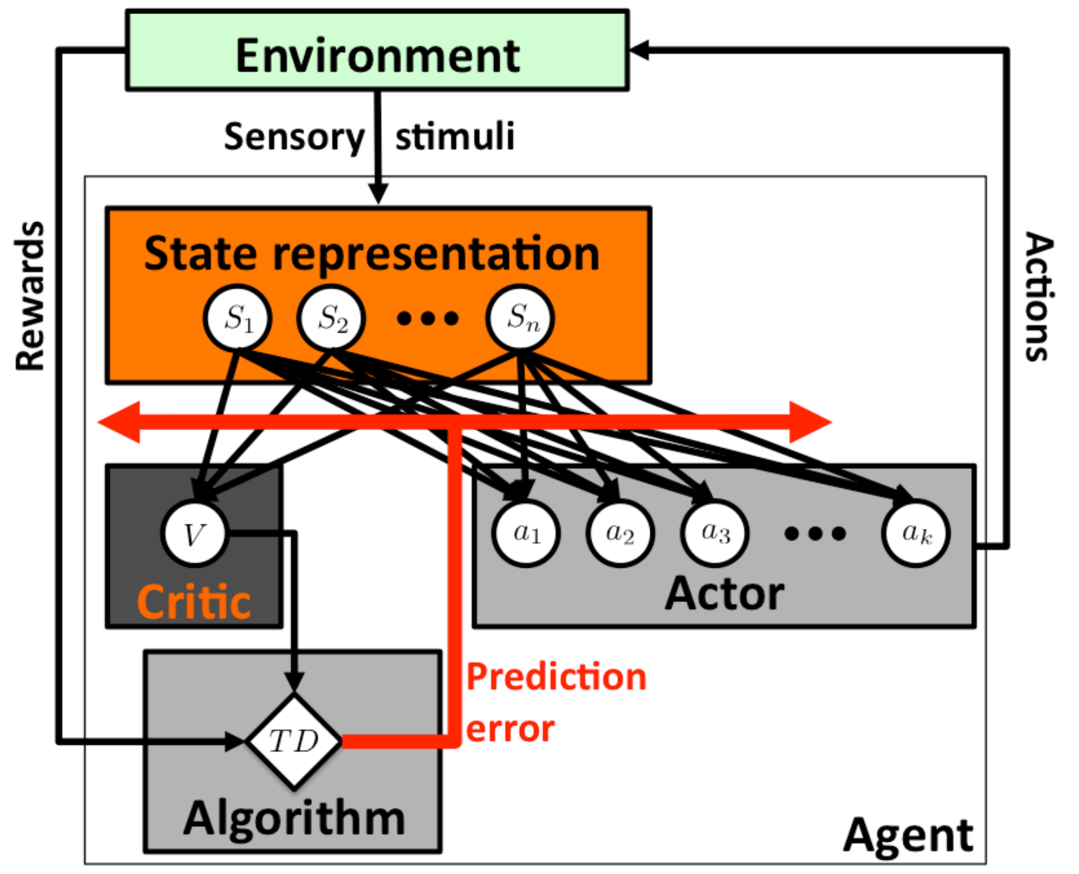
- Composite prediction error:

$$\delta = r + \gamma(V_1(S') + V_2(S')) - (V_1(S) + V_2(S))$$

- Two eligibility traces: $e_1(S), e_2(S)$
- Value update in OFC: $V_1(S) \leftarrow V_1(S) + \alpha_1 e_1(S)\delta$
- Value update in VS: $V_2(S) \leftarrow V_2(S) + \alpha_2 e_2(S)\delta$



Model 3: OFC builds task representation, VS is the critic

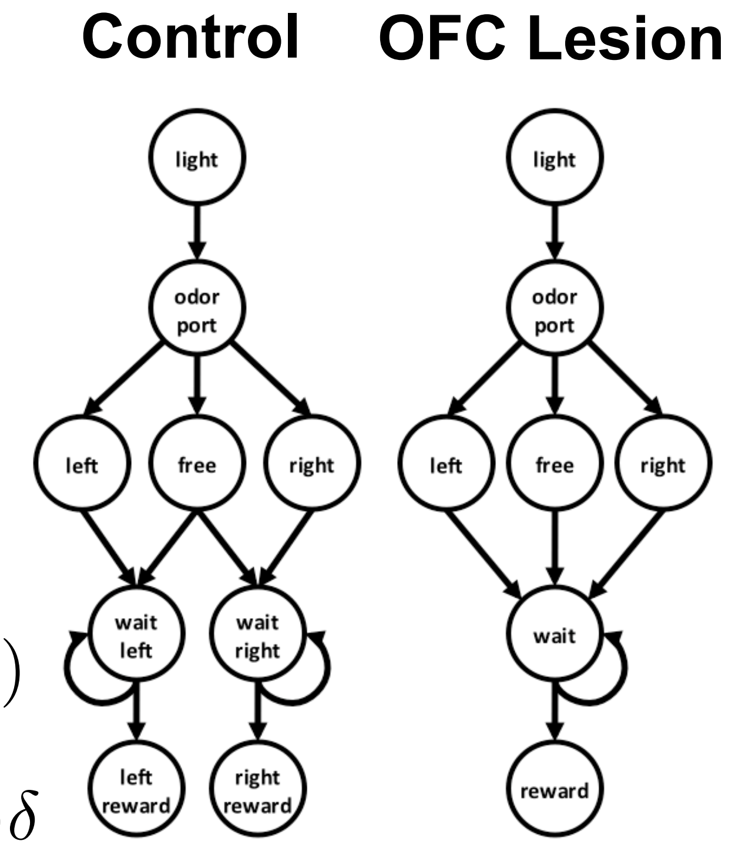


Description

- OFC constructs elaborate state representation that feeds into VS and enhances learning
- OFC lesion changes state representation to a simpler, stimulus-bound one
- VS lesion removes critic

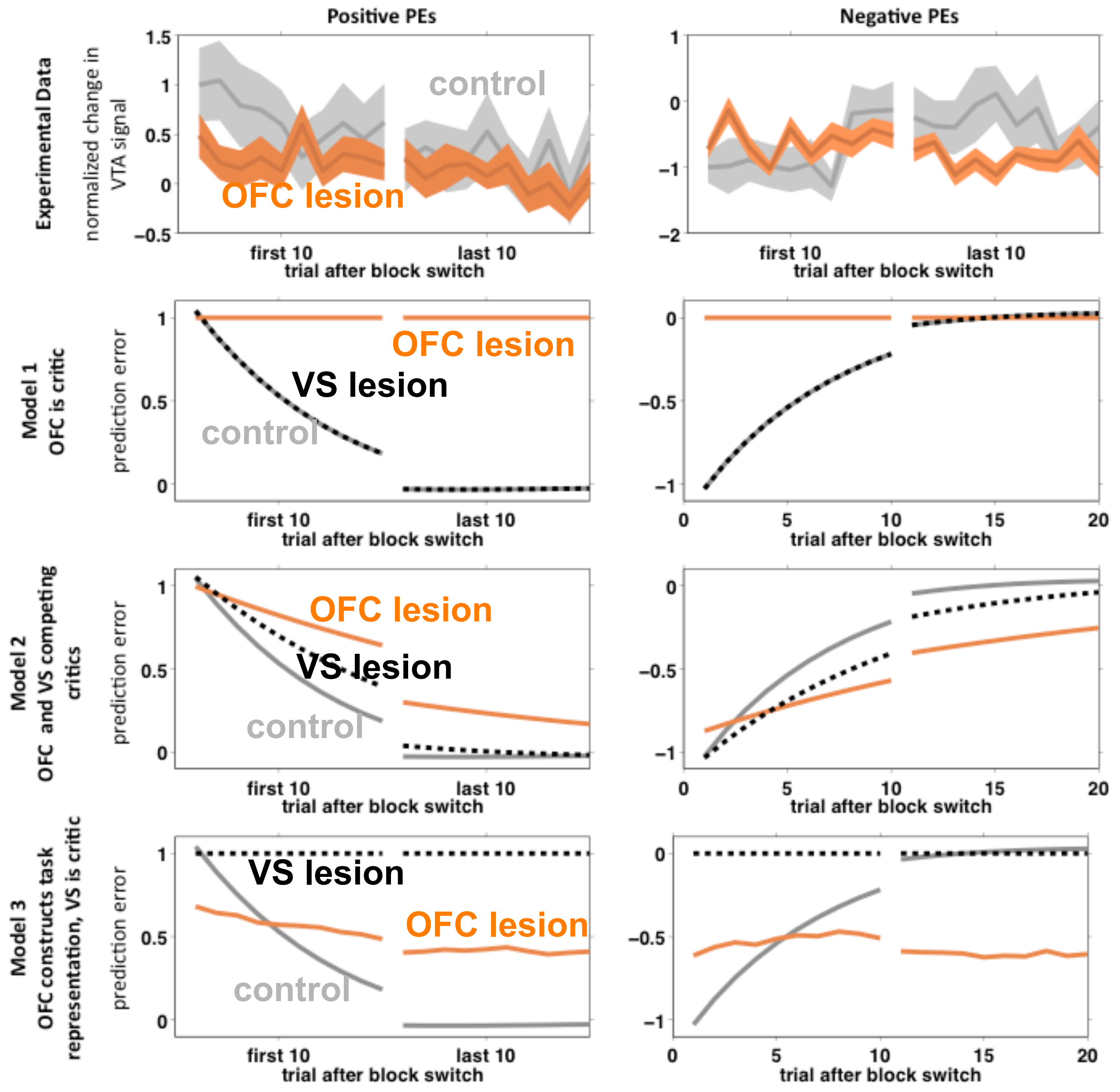
Learning in the (VS) critic

- Prediction error: $\delta = r + \gamma V(S') - V(S)$
- Eligibility trace: $e(S)$
- Value update: $V(S) \leftarrow V(S) + \alpha e(S)\delta$

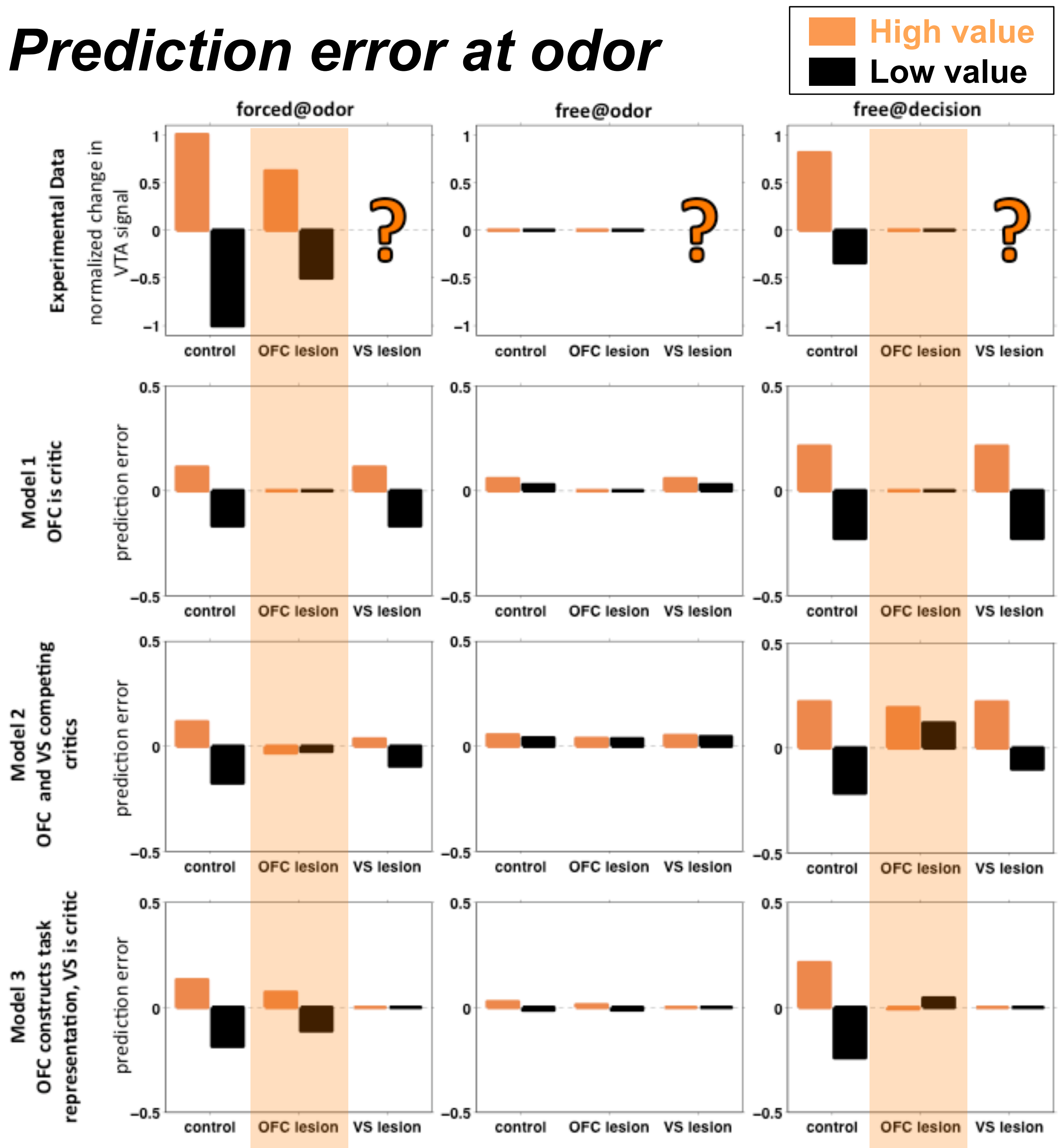


Results

Prediction error at reward



Prediction error at odor


High value
Low value
forced@odor
free@odor
free@decision
Experimental Data
normalized change in VTA signal
Model 1 OFC is critic
Model 2 OFC and VS competing critics
Model 3 OFC constructs task representation, VS is critic
prediction error
control
OFC lesion
VS lesion
?