

Orbitofrontal cortex as a cognitive map of task space: implications for reversal learning and extinction

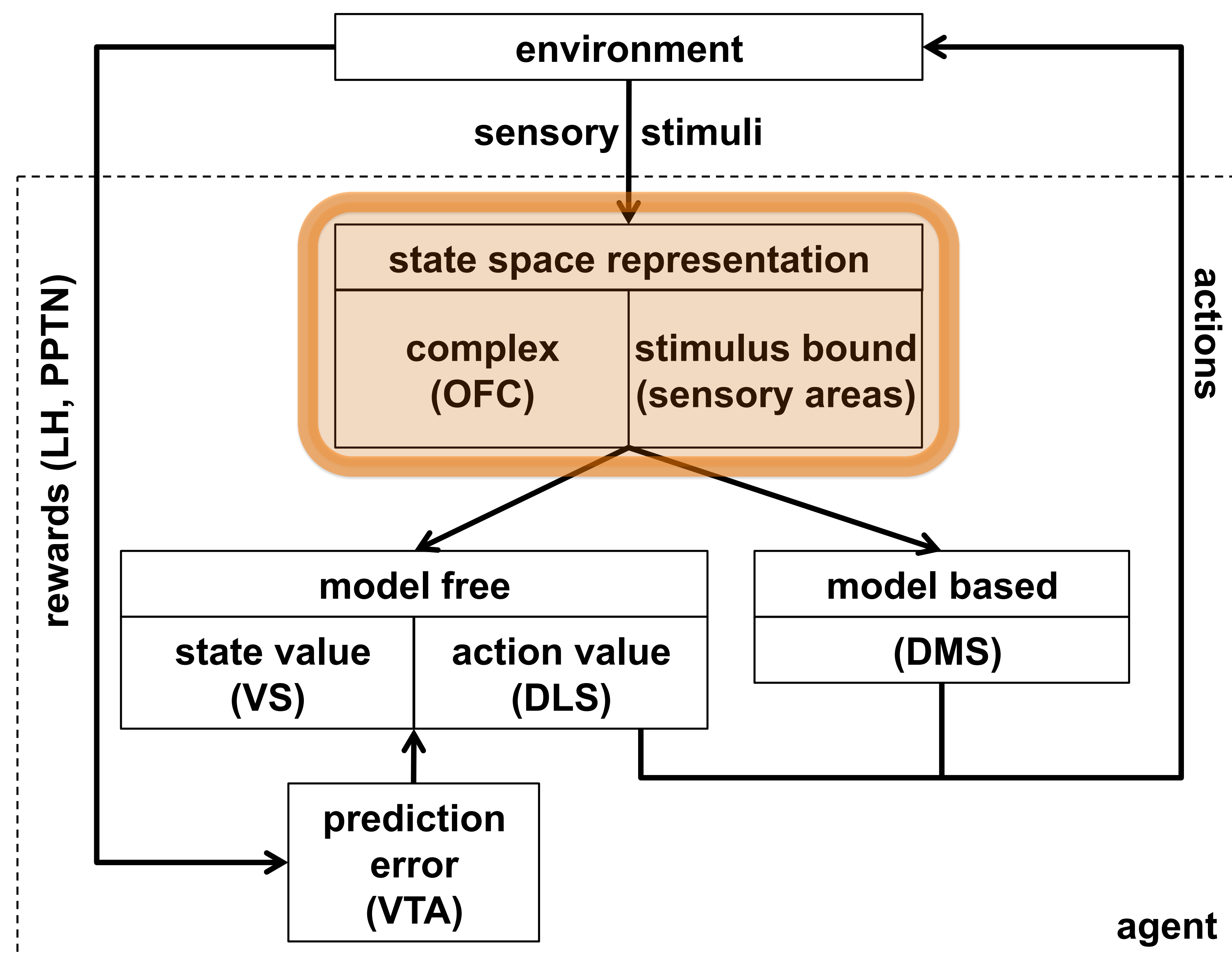
Robert C. Wilson, Yuji K. Takahashi, Geoffrey Schoenbaum and Yael Niv

Summary

- We have previously proposed [1] that OFC encodes a map of task space in much the same way as hippocampus encodes a map of physical space.
- Here we show how this theory can account for classic effects of OFC lesions on reversal learning and extinction.
- We also make testable predictions about the effects of OFC lesions on spontaneous recovery, renewal, rapid reacquisition and reinstatement

The Model

- OFC codes complex states, states that rely on abstract information (such as past actions or rewards)
- Without OFC the state-space is stimulus bound



SARSA learning rule

Prediction error : $\delta_t = r_t + \gamma Q(s', a') - Q(s, a)$

Action value update : $Q(s, a) \rightarrow Q(s, a) + \alpha \delta_t$

Action selection : $\pi(a|s) = \frac{\exp(\beta Q(s, a))}{\sum_{a'} \exp(\beta Q(s, a'))}$

References

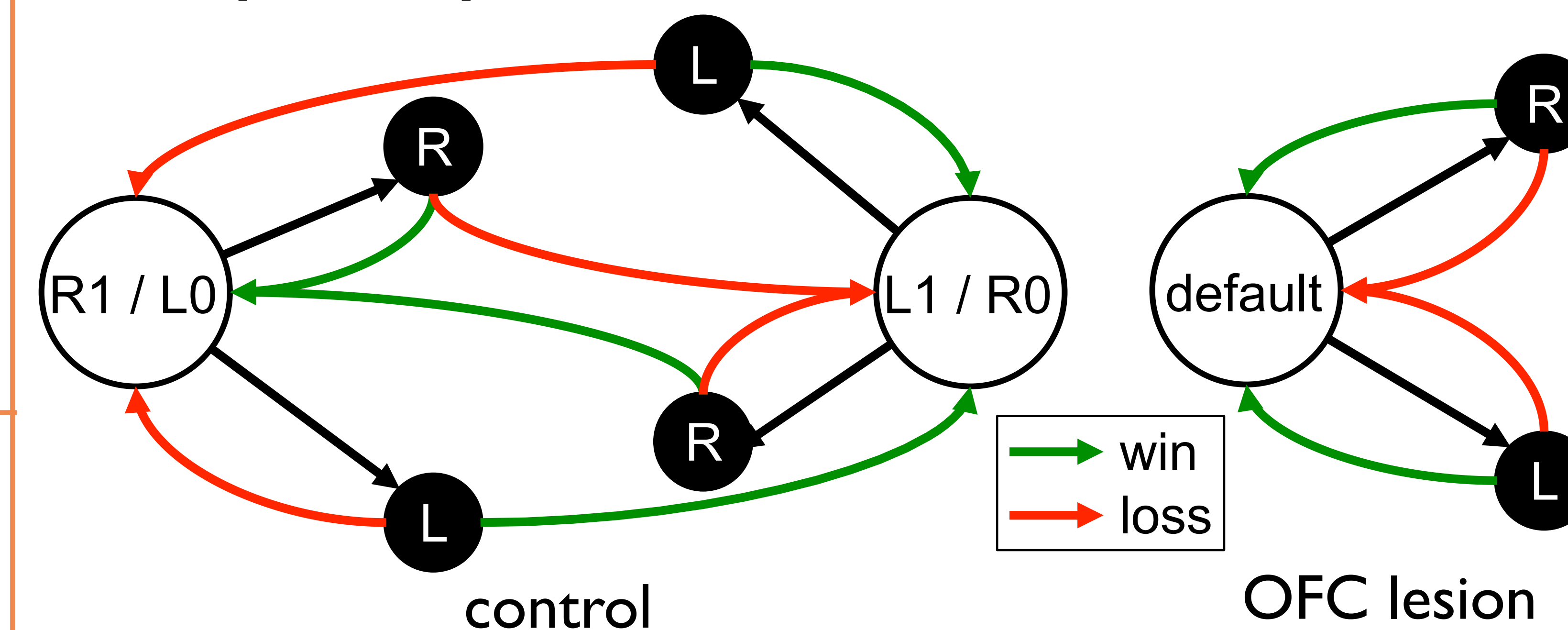
- [1] Y. K. Takahashi, et al. (2011) *Nature Neuroscience*, 14, 1590–1597
 [2] C.M. Butter (1969) *Physiology and Behavior*, 4, 163-171
 [3] C.M. Butter, et al. (1963) *Exp Neurol*, 7, 65–75.

Reversal learning

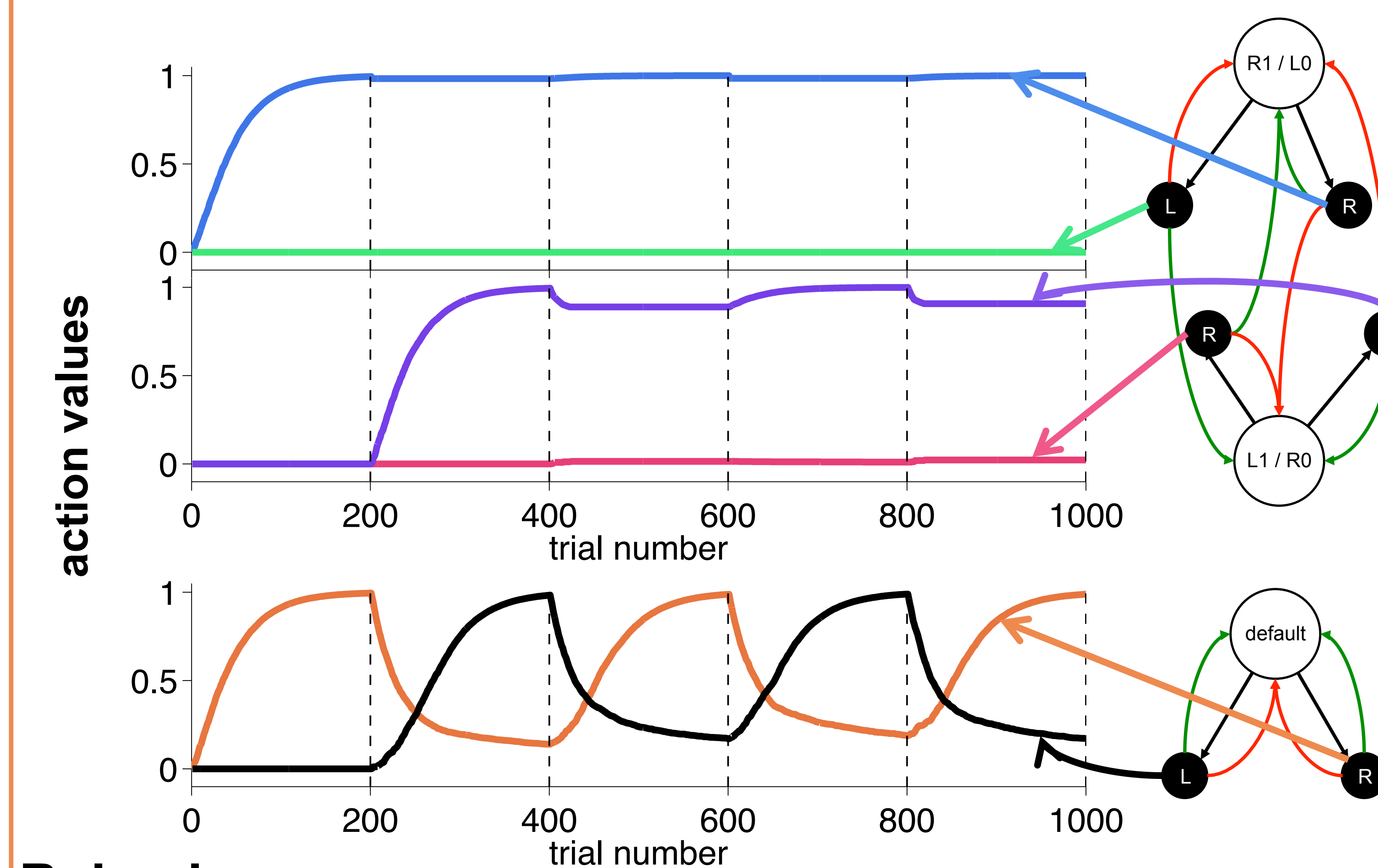
Task



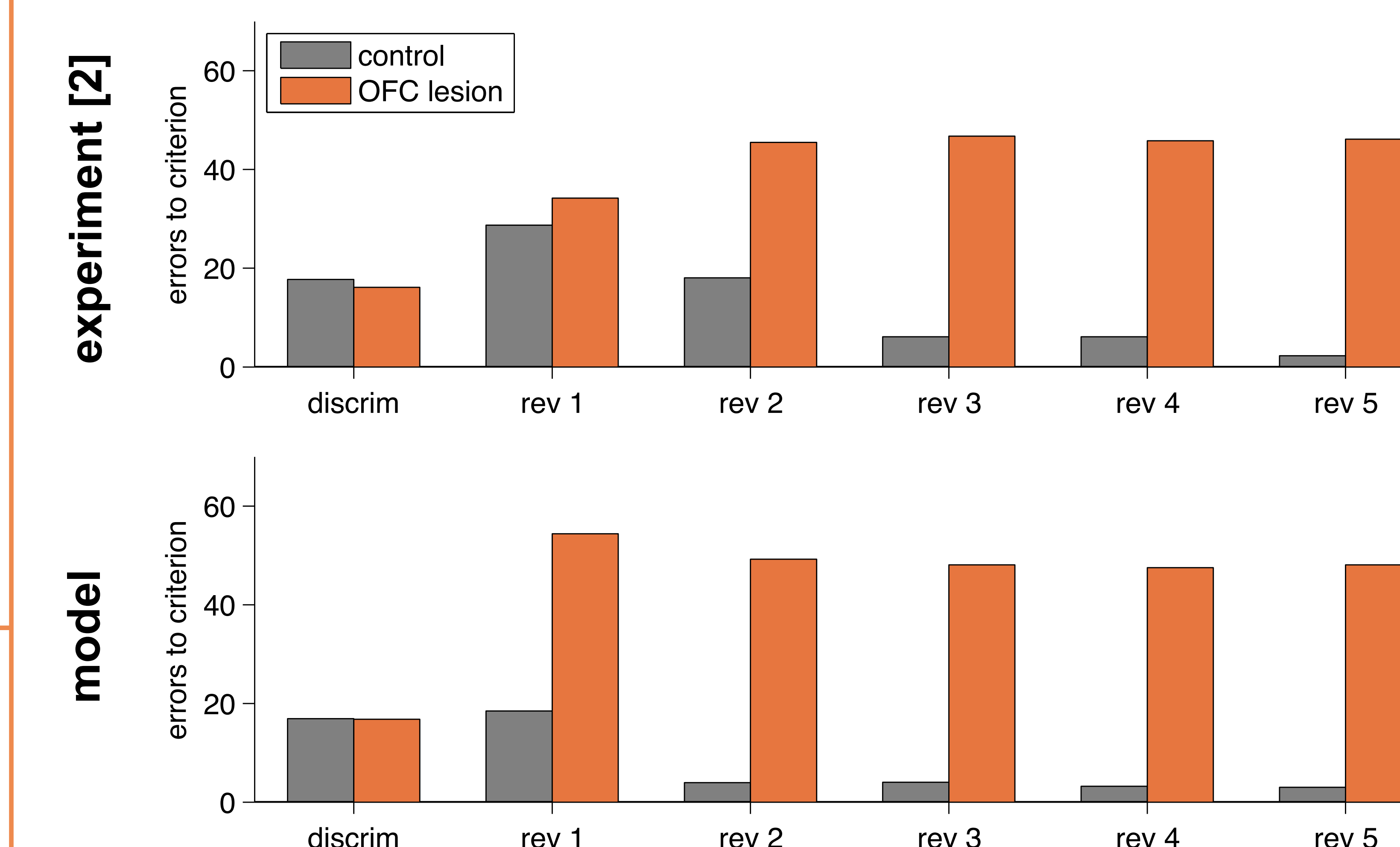
State-space representation



Action values

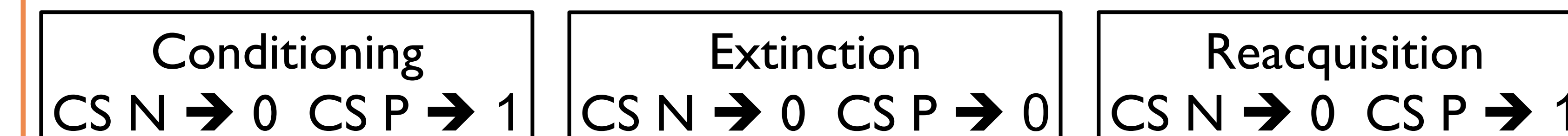


Behaviour

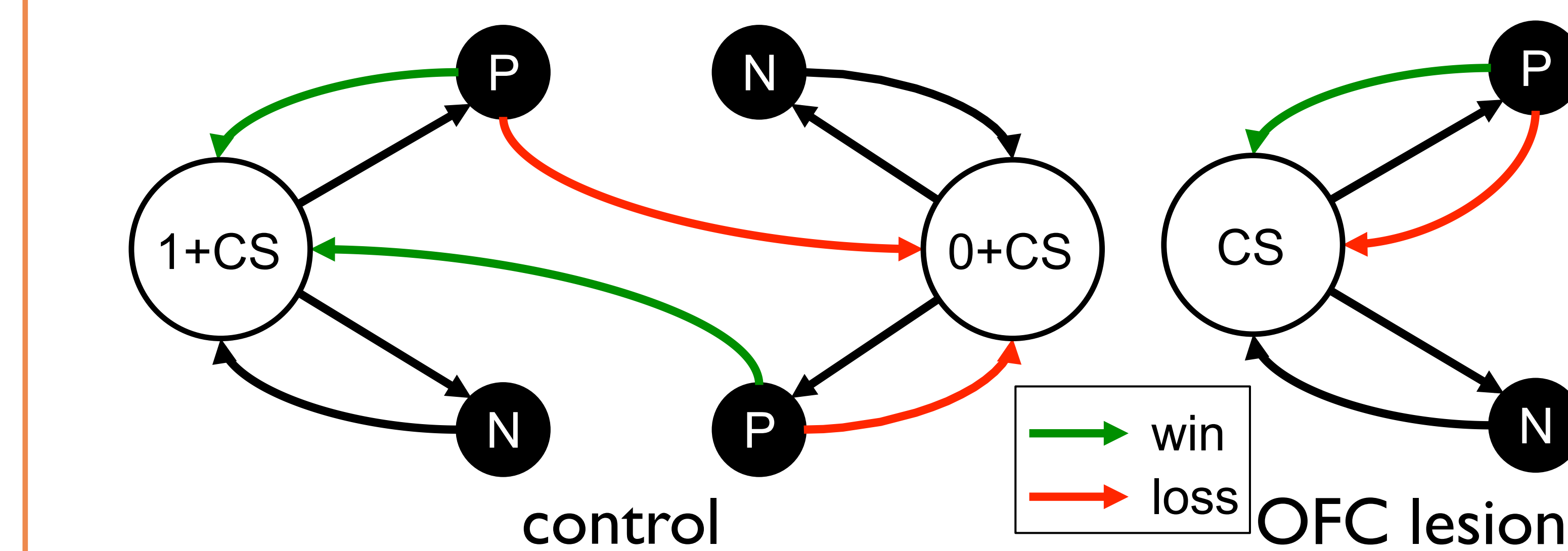


Extinction

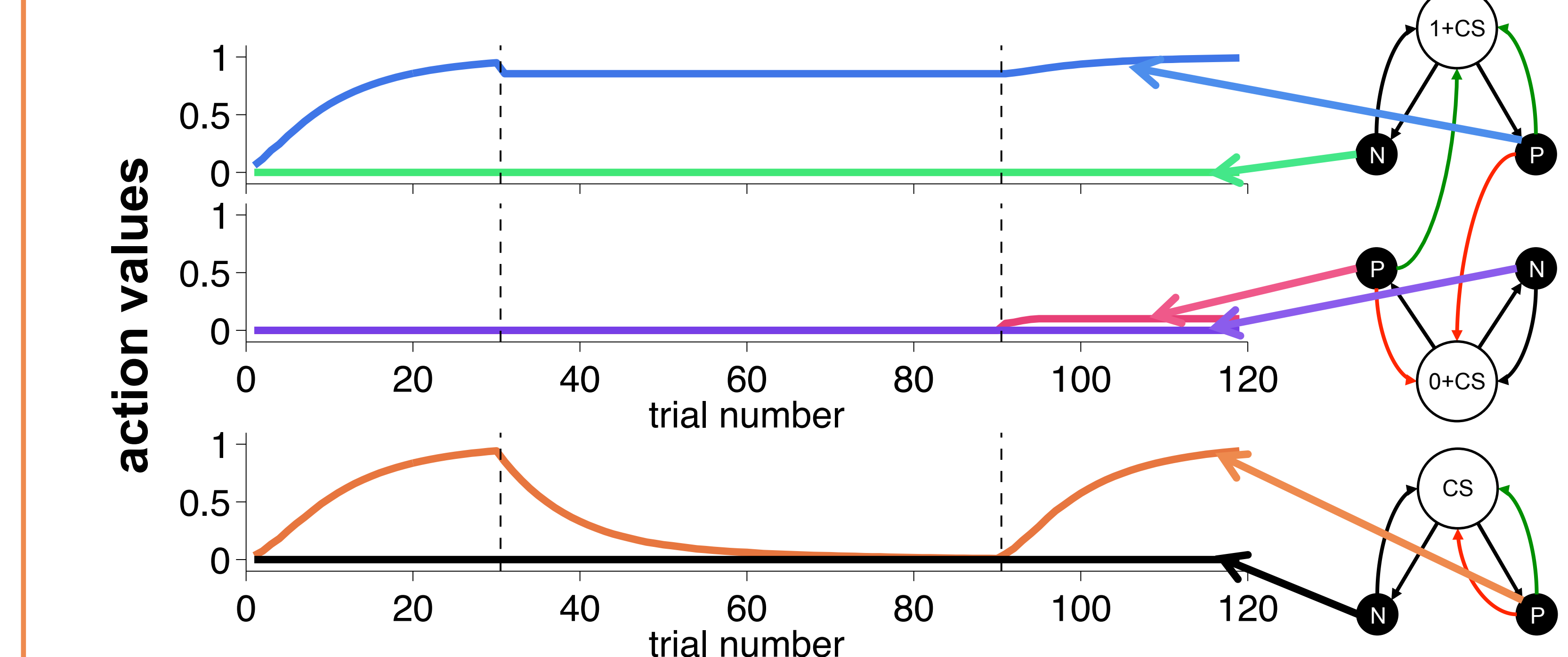
Task



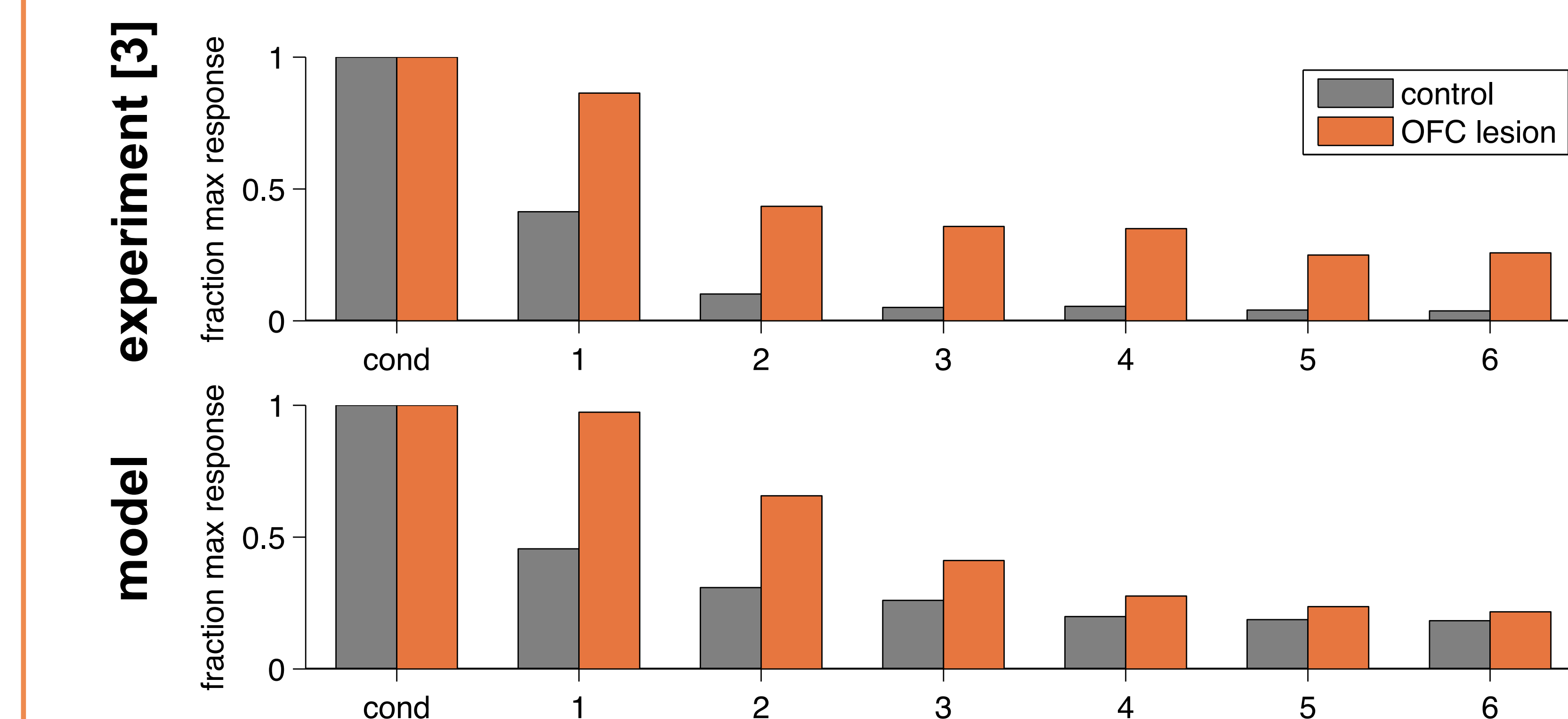
State-space representation



Action values



Behaviour



Experimental predictions

