

The Role of Orbitofrontal Cortex in Model-Based Planning in the Rat

Kevin Miller

Reinforcement Learning & Decision Making

6/8/2015

What is Planning?

Planning

Press pedal → Car slows down

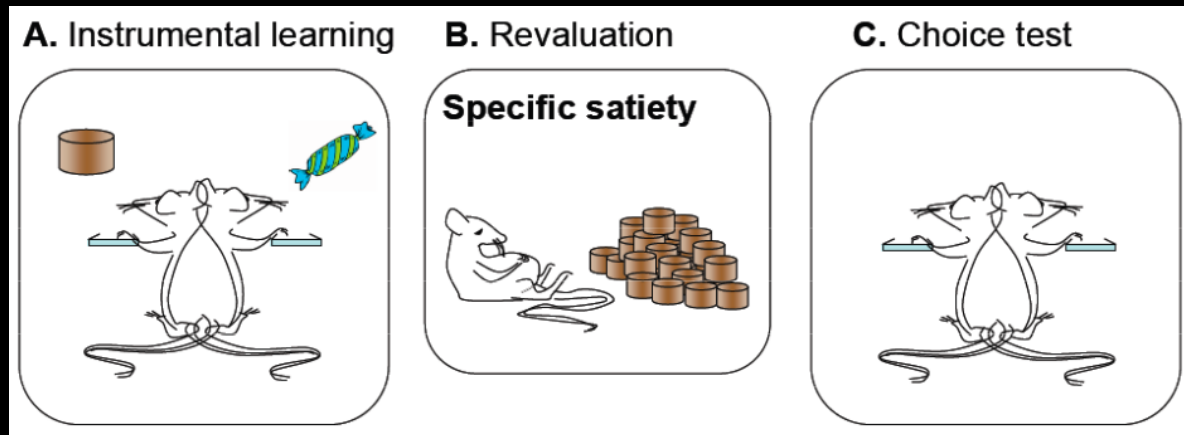
Model-Free

Red light → Press Pedal



Decision-making guided by an internal model of the outcomes of your actions

Existing Planning Tasks



- One-shot manipulations
- Investigating neural correlates requires aggregating over many decisions
- We need new behaviors

Outline



1. A planning task for rats

- Quantifying planning
- Behavior Model

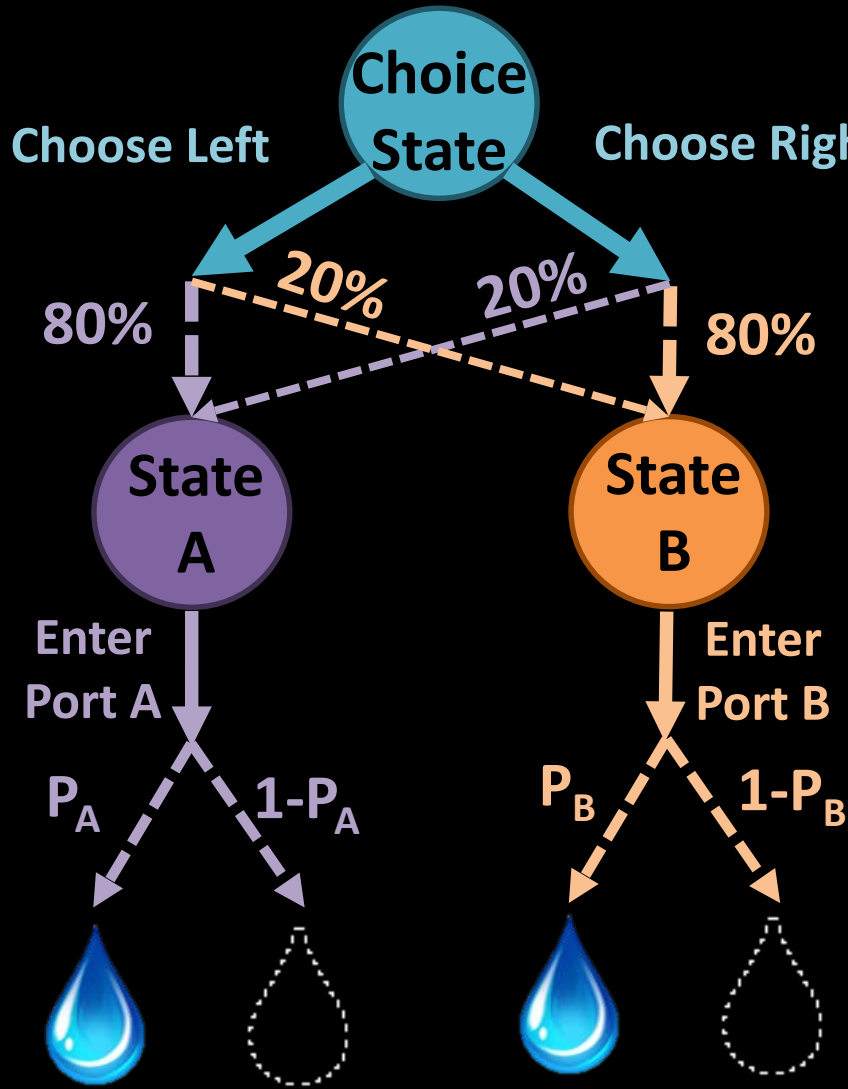
2. Testing causality

- Inactivations in mPFC, OFC, hippocampus

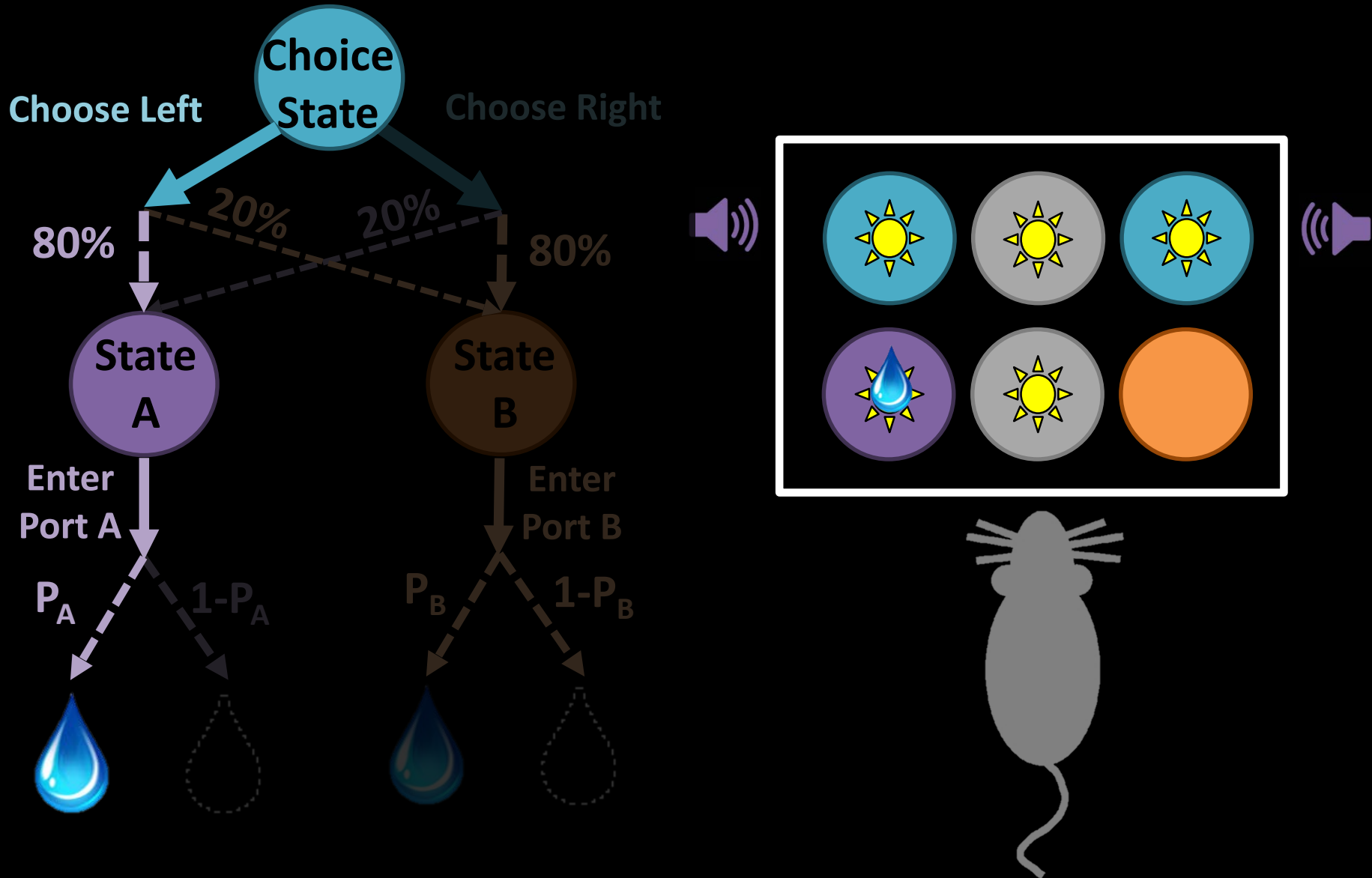
3. Neural correlates of planning

- Electrophysiology in OFC

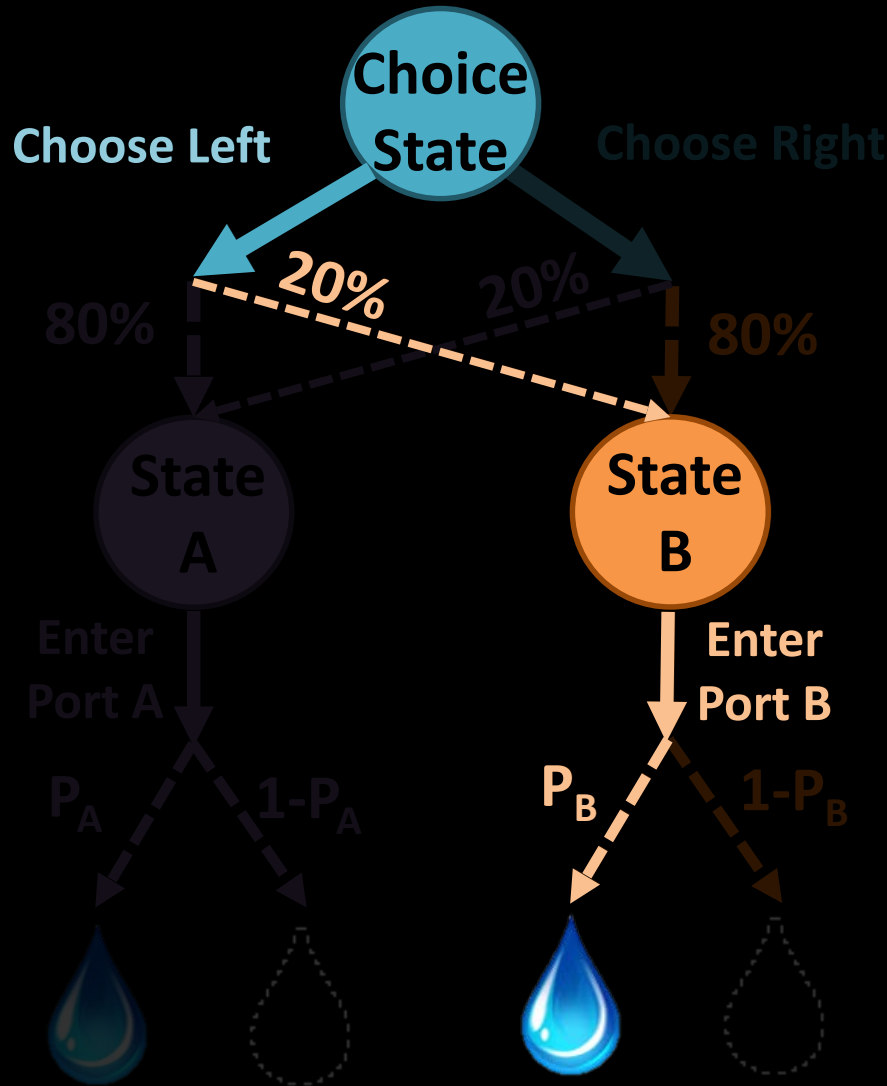
The Rat Two-Step Task



The Rat Two-Step Task



The Rat Two-Step Task



Planning

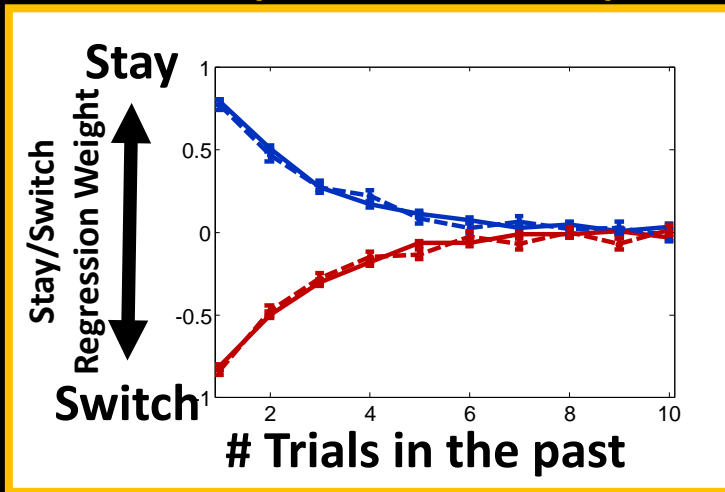
I got rewarded at **Port B**.
I want to get back there.
I'll choose right!

Model-Free

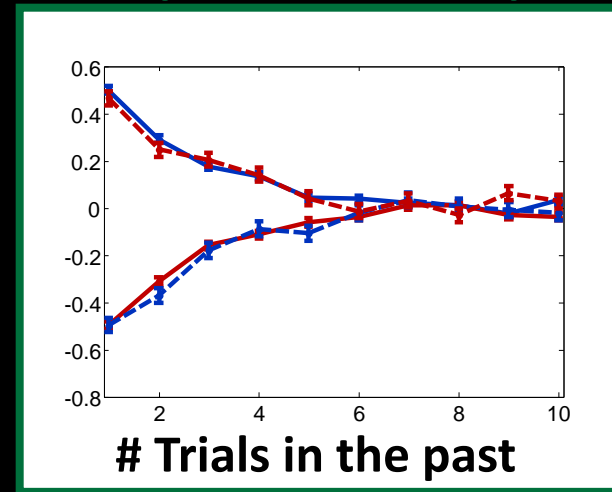
I got rewarded
after **choosing left**.
I'll choose left
again!


Identifying Planning Behavior

**Model-Free Agent
(Simulated Data)**




**Planning Agent
(Simulated Data)**



 80% Transition, Rewarded

 80% Transition, Unrewarded

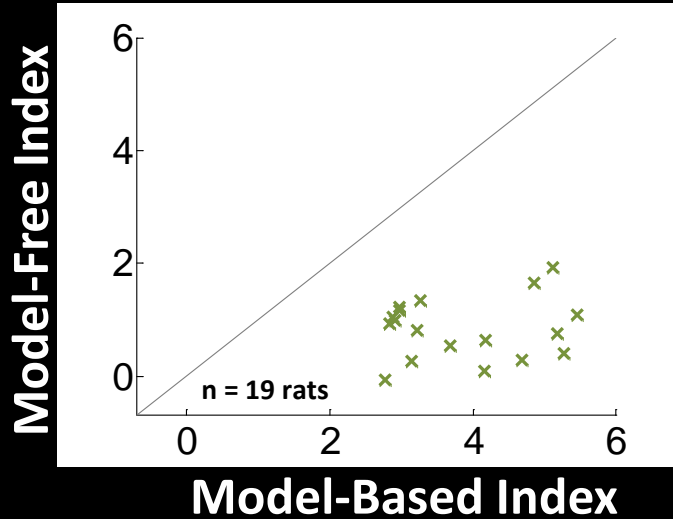
 20% Transition, Rewarded

 20% Transition, Unrewarded

$$ModelFree\ Index = \sum \text{solid blue line} + \sum \text{dashed blue line} - \sum \text{dashed red line} - \sum \text{solid red line}$$

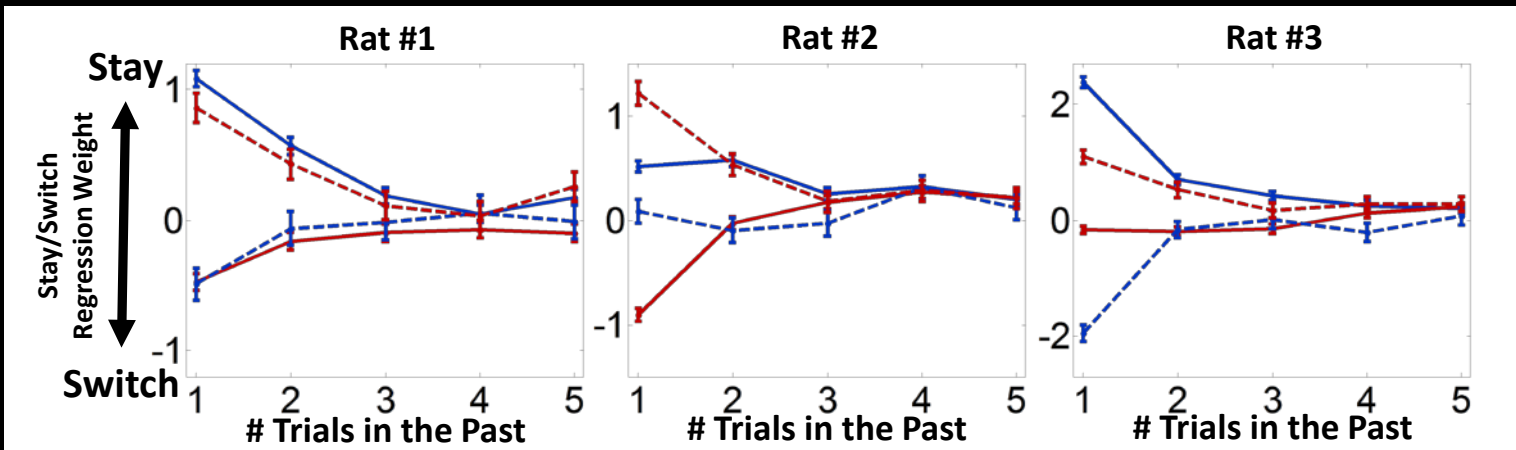
$$ModelBased\ Index = \sum \text{solid blue line} + \sum \text{dashed red line} - \sum \text{dashed blue line} - \sum \text{solid red line}$$

Rat Behavior



- Model-based index \gg model-free index
- All rats show planning-type regressions
- Other patterns present too

Example Rats



80% Transition, Rewarded 20% Transition, Rewarded
80% Transition, Unrewarded 20% Transition, Unrewarded

Trial-by-Trial Model

The Model

1. Model-Based RL

$$Q_{mb}(choicePort) = \sum_{ports} V(rewardPort) * P(rewardPort|choice)$$

$$V(rewardPort) \leftarrow V(rewardPort) + \alpha_{MB}(Reward - V(rewardPort))$$

2. Model-Free RL

$$Q_{MF}(choicePort) \leftarrow V(choicePort) + \alpha_{MF}(Reward - Q_{MF}(choicePort))$$

3. Common-Stay/Uncommon-Switch

4. Perseveration

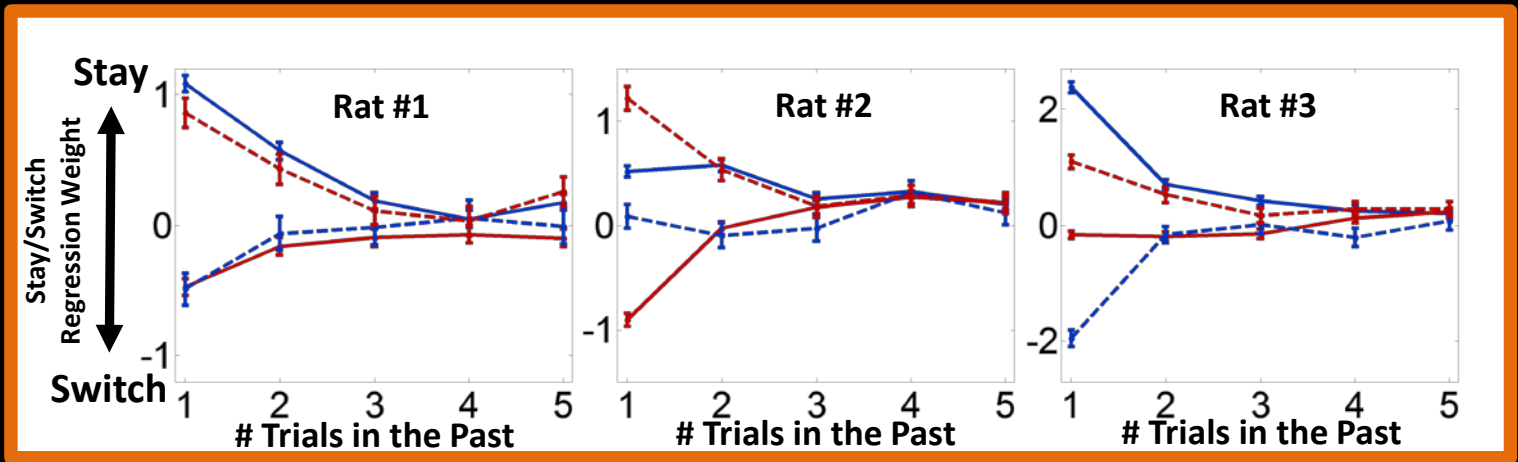
5. Bias

$$Q_{total}(a) = \beta_{MB}Q_{MB}(a) + \beta_{MF}Q_{MF}(a) + \beta_{csus}Q_{csus}(a) + \beta_{persev}Q_{persev}(a) + \beta_{bias}(a)$$

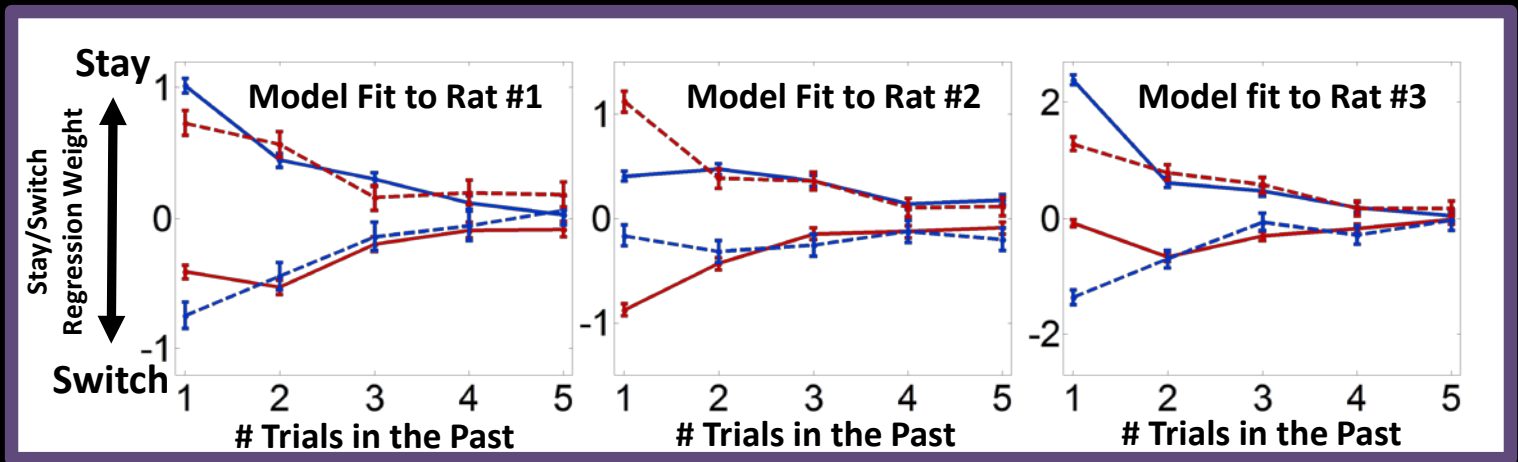
$$\pi(a) = \frac{e^{Q_{total}(a)}}{\sum_{a'} e^{Q_{total}(a')}}$$

Model Matches Data Qualitatively

Data:



Model:



- 80% Transition, Rewarded
 - - - - 20% Transition, Rewarded
- 80% Transition, Unrewarded
 - - - - 20% Transition, Unrewarded

Outline



1. A planning task for rats

- Quantifying planning
- Behavior Model

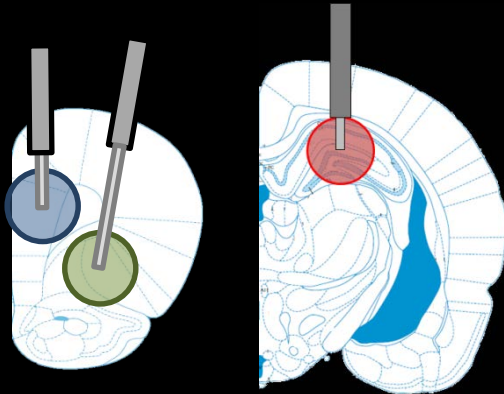
2. Testing causality

- Inactivations in mPFC, OFC, Hippocampus

3. Neural correlates of planning

- Electrophysiology in OFC

Where does planning happen in the brain?



1. Prelimbic PFC

- Causal to simpler planning tasks
- Homologous to primate dlPFC?

2. Orbitofrontal Cortex

- Encodes economic value?
- Model-based values only?
- State space for RL?

3. Dorsal Hippocampus

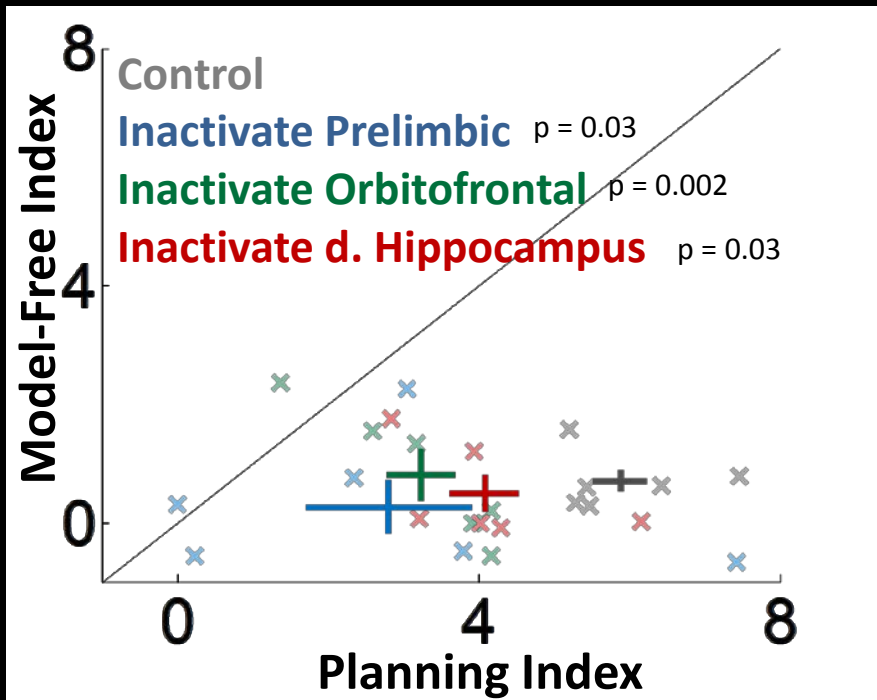
- “Cognitive map” of space
- Preplay at choice points
- But: not causal in some simpler tasks



We can do it all in
the same animal!

Inactivations Impair Planning

Mean Parameter Values



	Control	Δ PL	Δ OFC	Δ dH
β_{MB}	3.2	-2.2*	-1.7*	-0.94*

- Implant rats ($n=6$) with cannula in each region
- Infuse muscimol prior to training to silence neural activity

Outline

1. A planning task for rats

- Quantifying planning
- Behavior Model



2. Testing causality

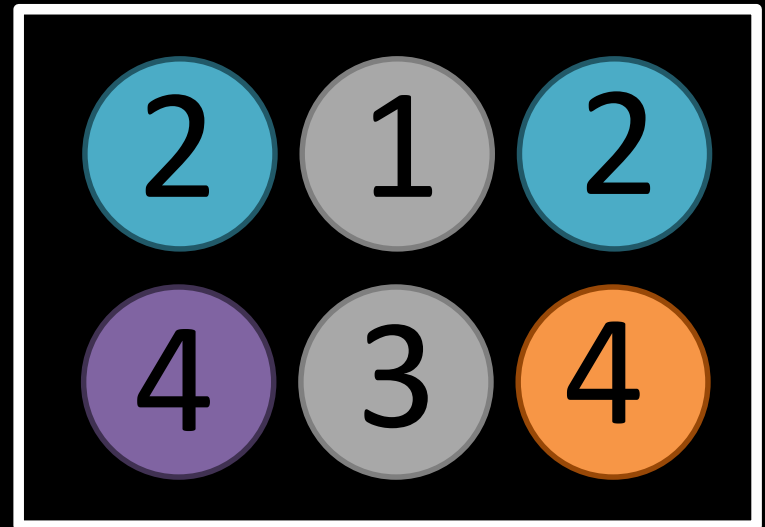
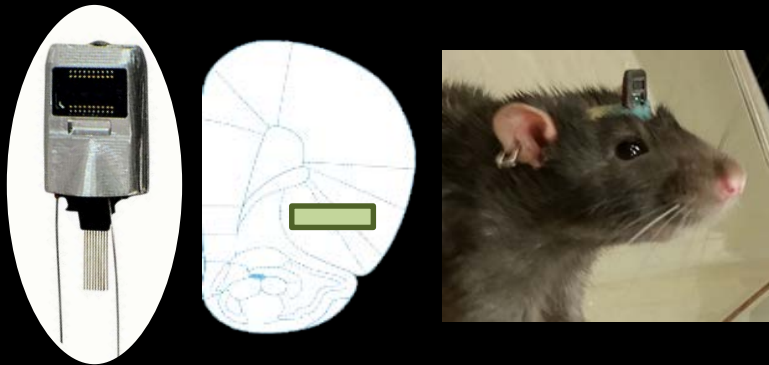
- Inactivations in mPFC, hippocampus, OFC

3. Neural correlates of planning

- Electrophysiology in OFC

Array Recordings in OFC

- Implant rats with multielectrode array in OFC
- Currently: Pilot data from one rat
- Time-lock recordings to the four pokes of each trial



Values for Choice and Learning

Model-Based RL

Choice: $Q_{mb}(choicePort) = \sum_{ports} V(rewardPort) * P(rewardPort|choice)$

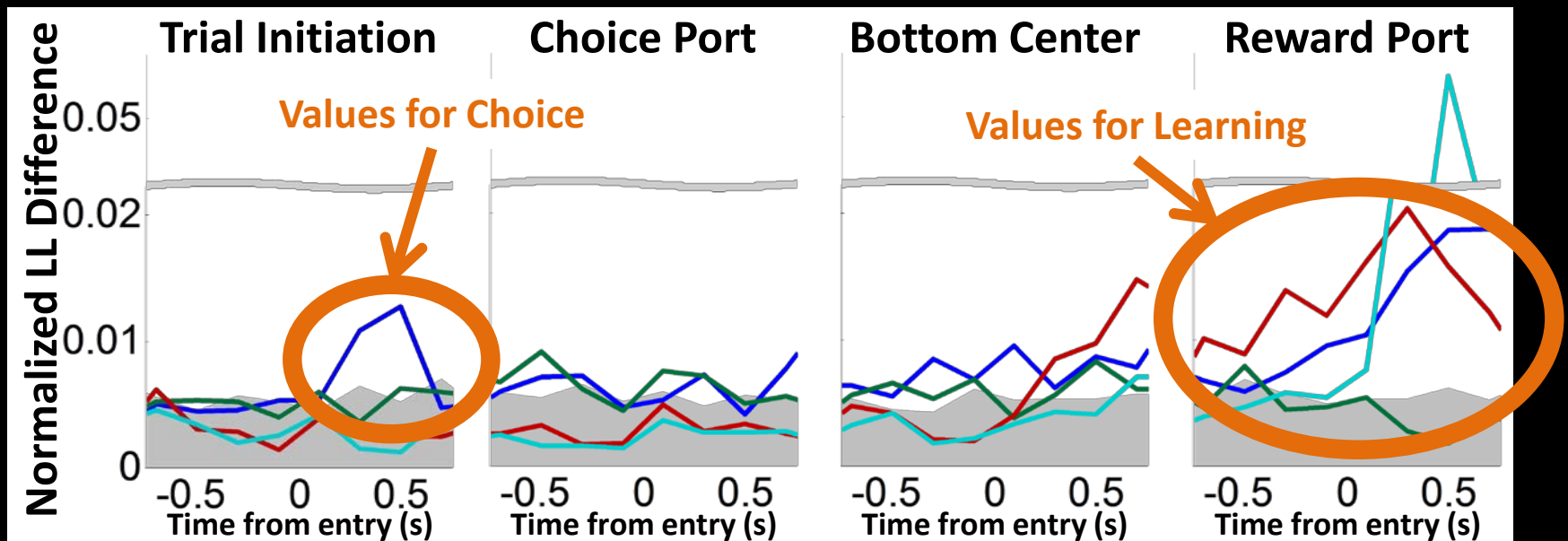
Learning: $V(rewardPort) \leftarrow V(rewardPort) + \alpha(Reward - V(rewardPort))$

Choice Value: $Q_{mb}(\text{left choice port}) - Q_{mb}(\text{right choice port})$

Chosen Value: $Q_{mb}(\text{chosen port})$

Outcome Value: $V(\text{reward port})$

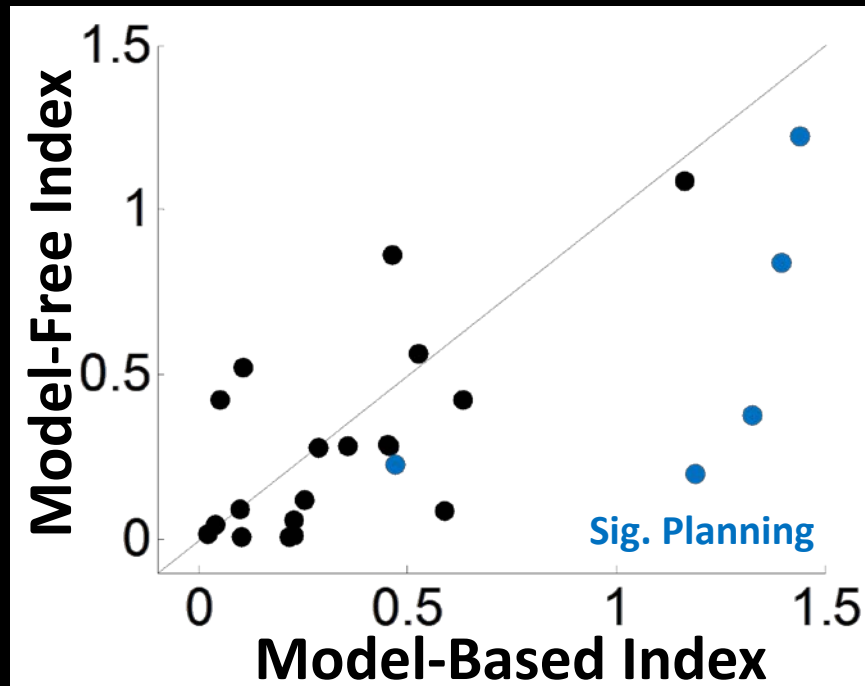
Reward



Choice Values are Model-Based

For each cell:

- Run Model-Free/Model-Based regression on firing rate at decision time
- Compute model-free and model-based index

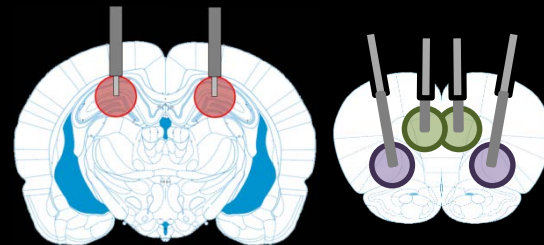
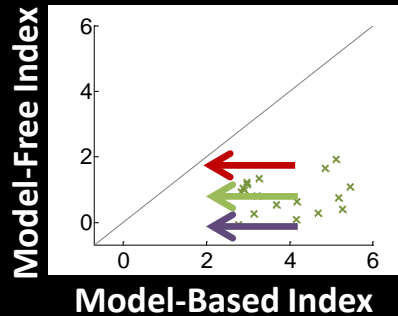
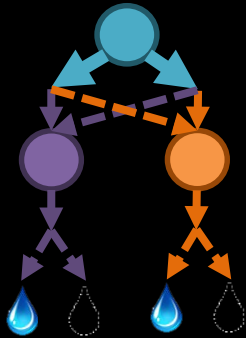


Planning Index > Model-Free Index ($p=0.03$)

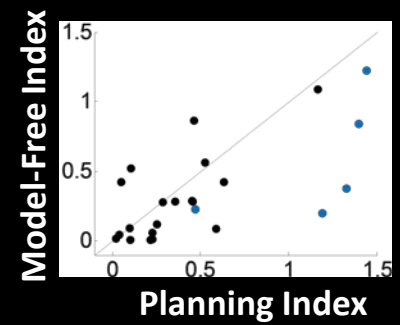
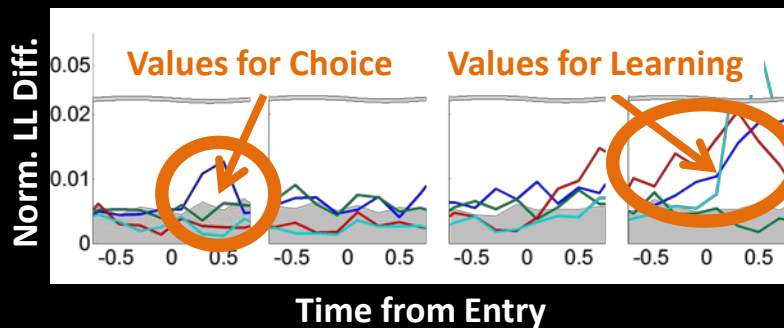
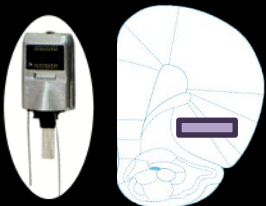
5/25 cells individually significant ($p<0.05$; expected by chance at $p=0.007$)

Conclusions

- Novel task gives behavioral readout of planning
 - Rats adopt planning strategy
 - PL, dHippo, OFC may be involved in planning



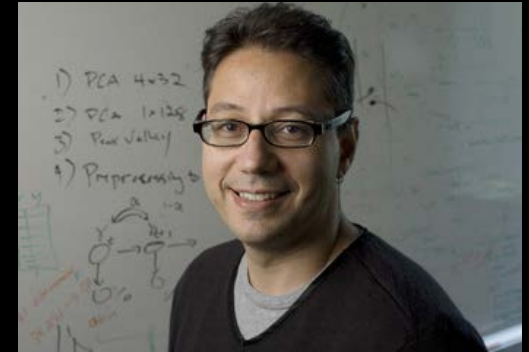
- OFC encodes values for both choice and learning
 - Choice values model-based



Thanks!



Matt Botvinick



Carlos Brody



Nathaniel Daw



Yael Niv



Chuck Kopec



Jeff Erlich



Ann Duan



Bob Wilson



Thomas Akam

Jovanna Teran
Klaus Osorio
Ricki Latourette
Scott Baptista
Tegan James
Samantha Stein
Adrian Sirko



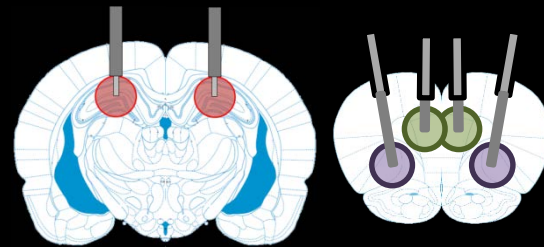
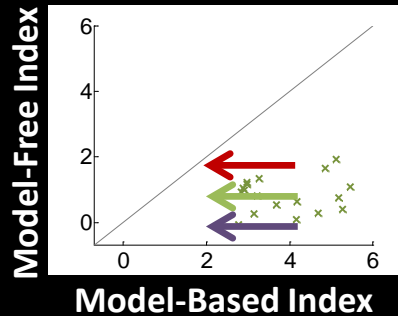
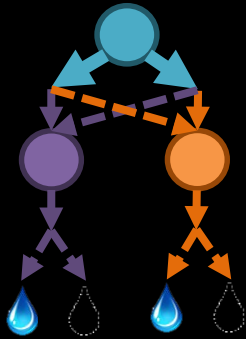
Athena Akrami



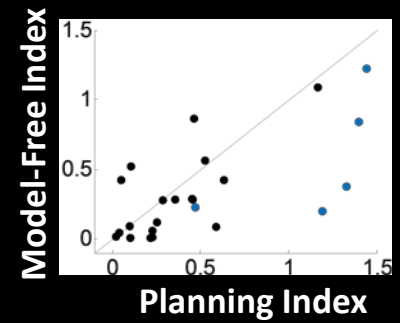
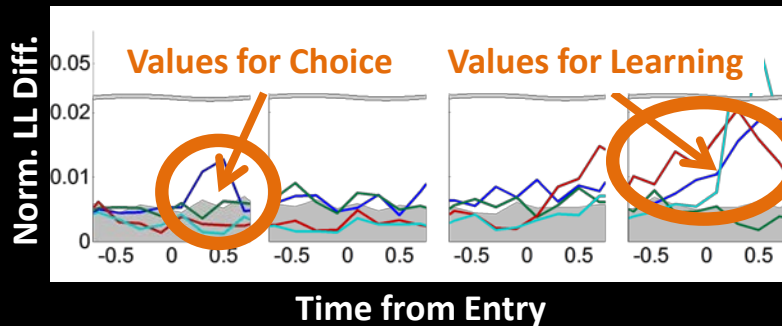
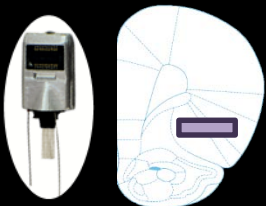
Tim Hanks

Thanks!

- Novel task gives behavioral readout of planning
 - Rats adopt planning strategy
 - PL, dHippo, OFC may be involved in planning



- OFC encodes values for both choice and learning
 - Choice values model-based



Thanks!