Exploration via disruptions in prefrontal control dynamics
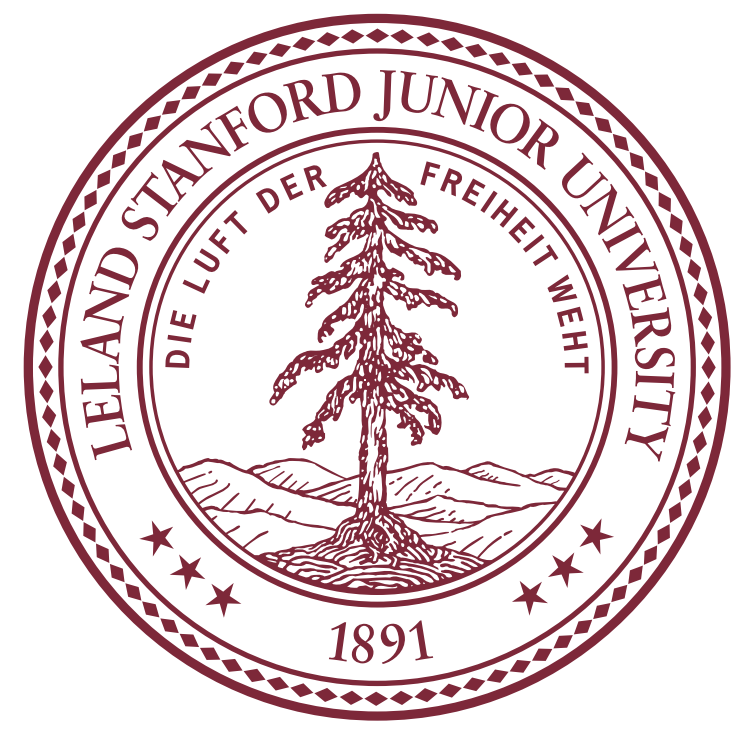
R. Becket Ebitz[1,3], Tim Buschman[1,2], Tirin Moore[3,4]

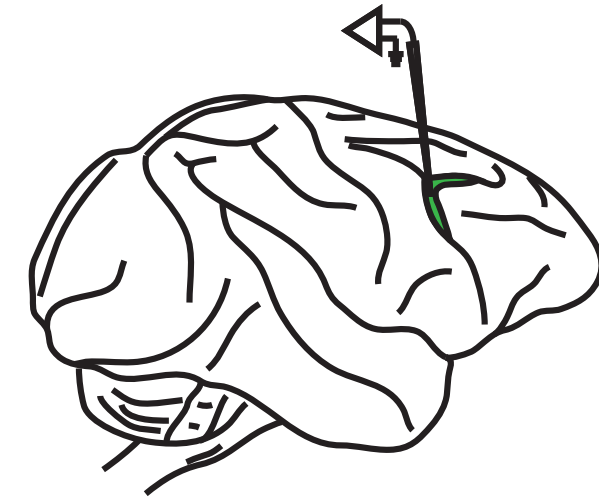[1] Princeton Neuroscience Institute and [2] Department of Pscyhology, Princeton University,
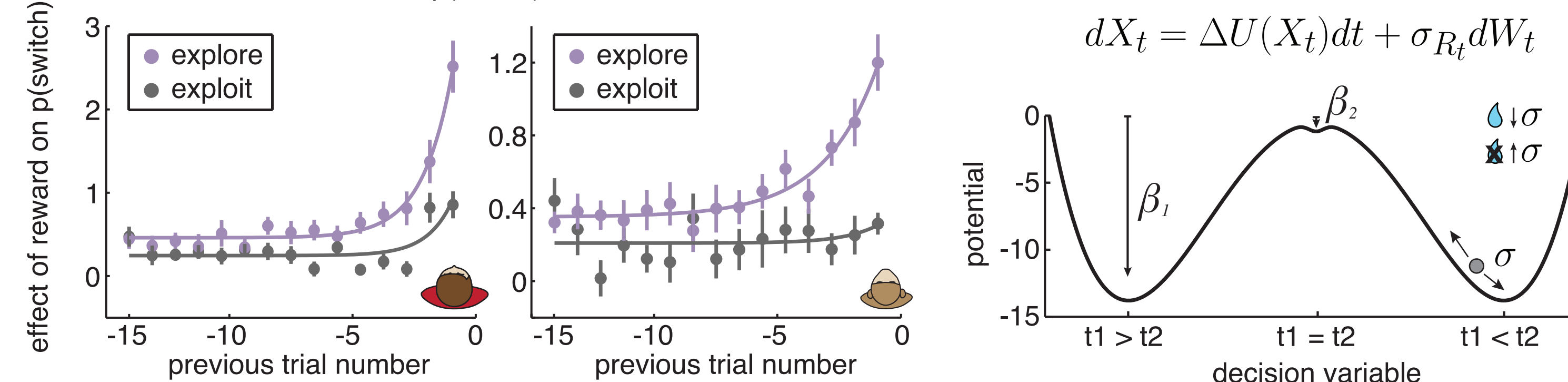[3] Neurobiology Department and [4] Howard Hughes Medical Institute, Stanford University School of Medicine,
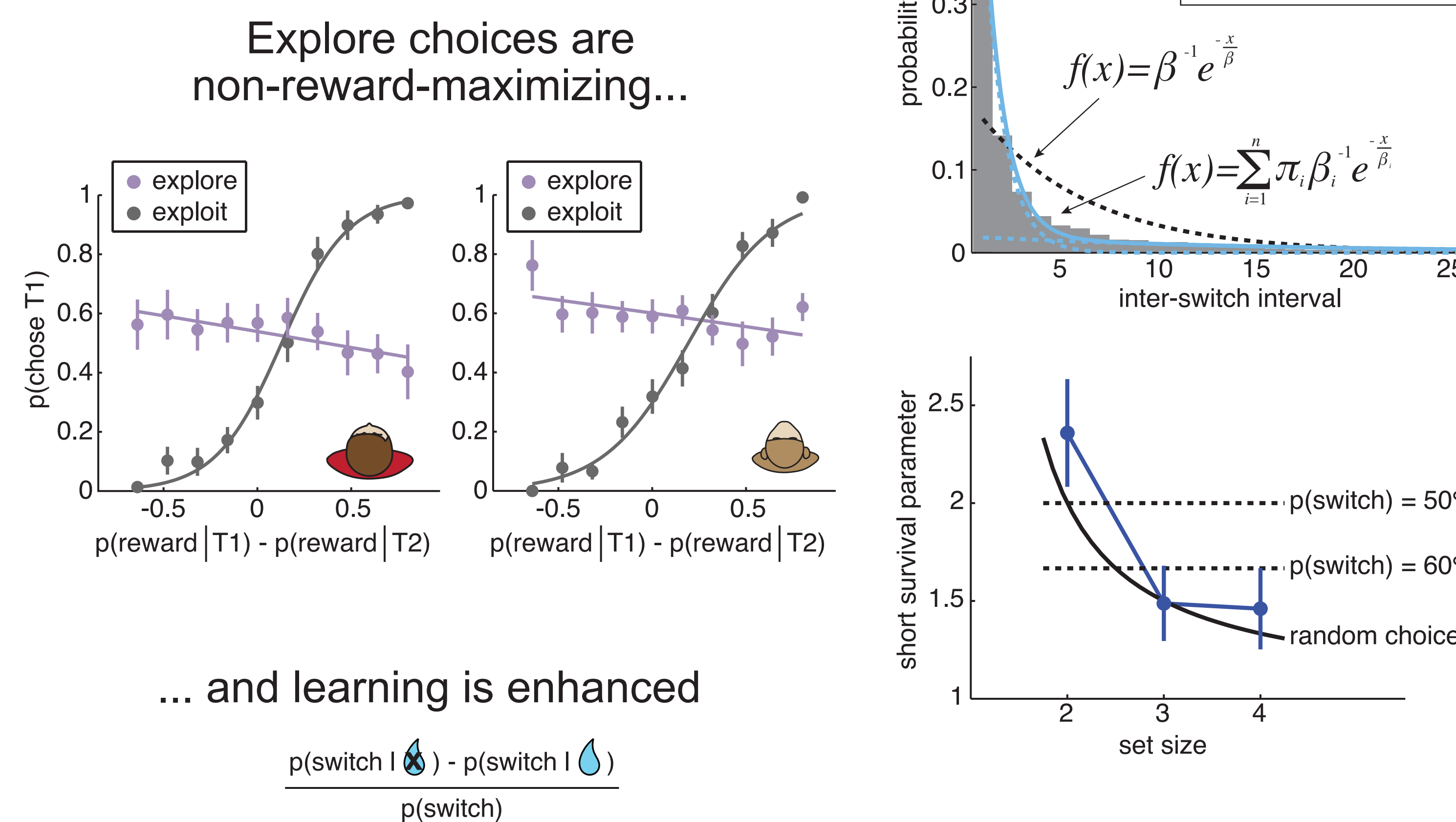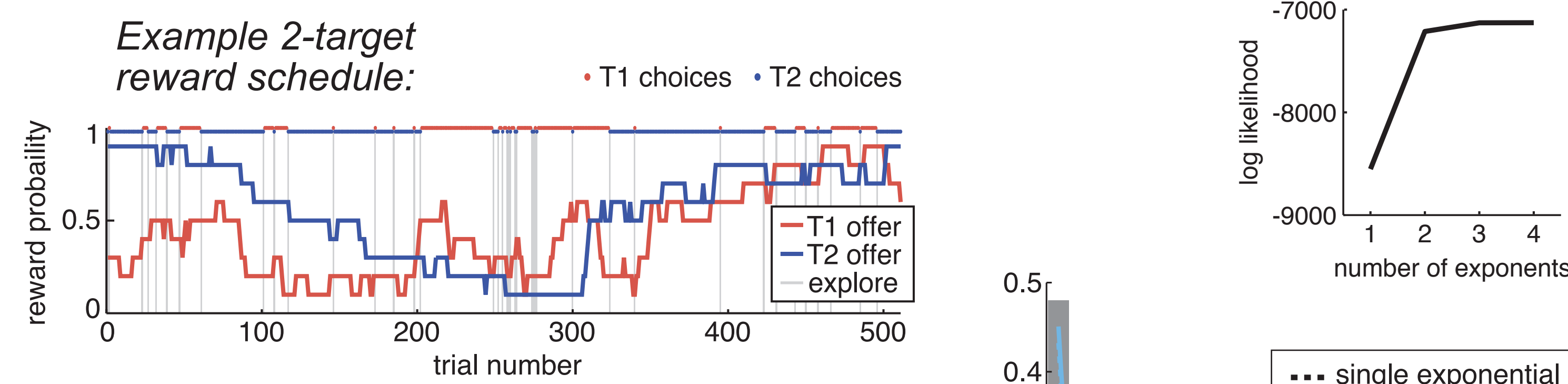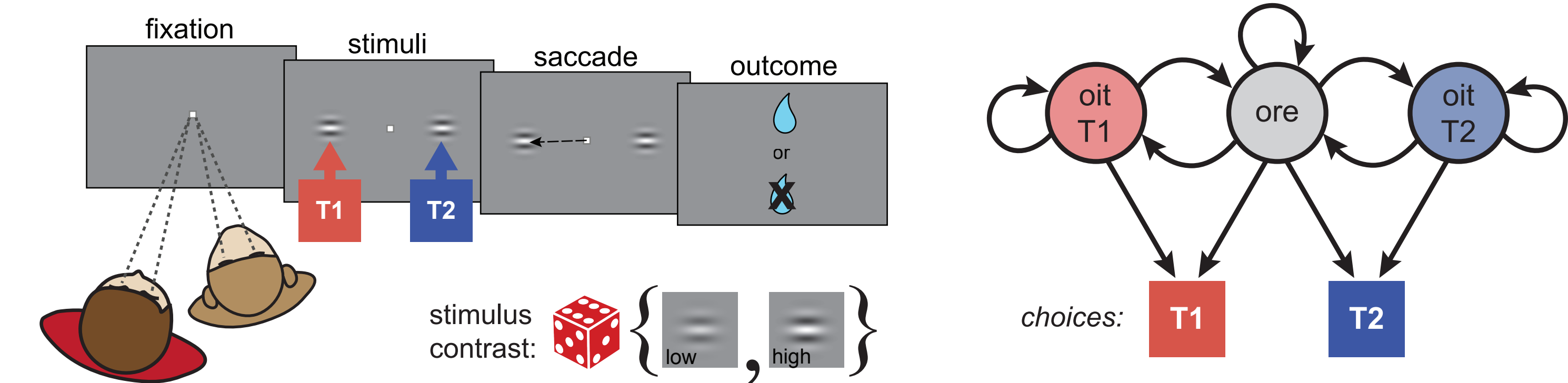
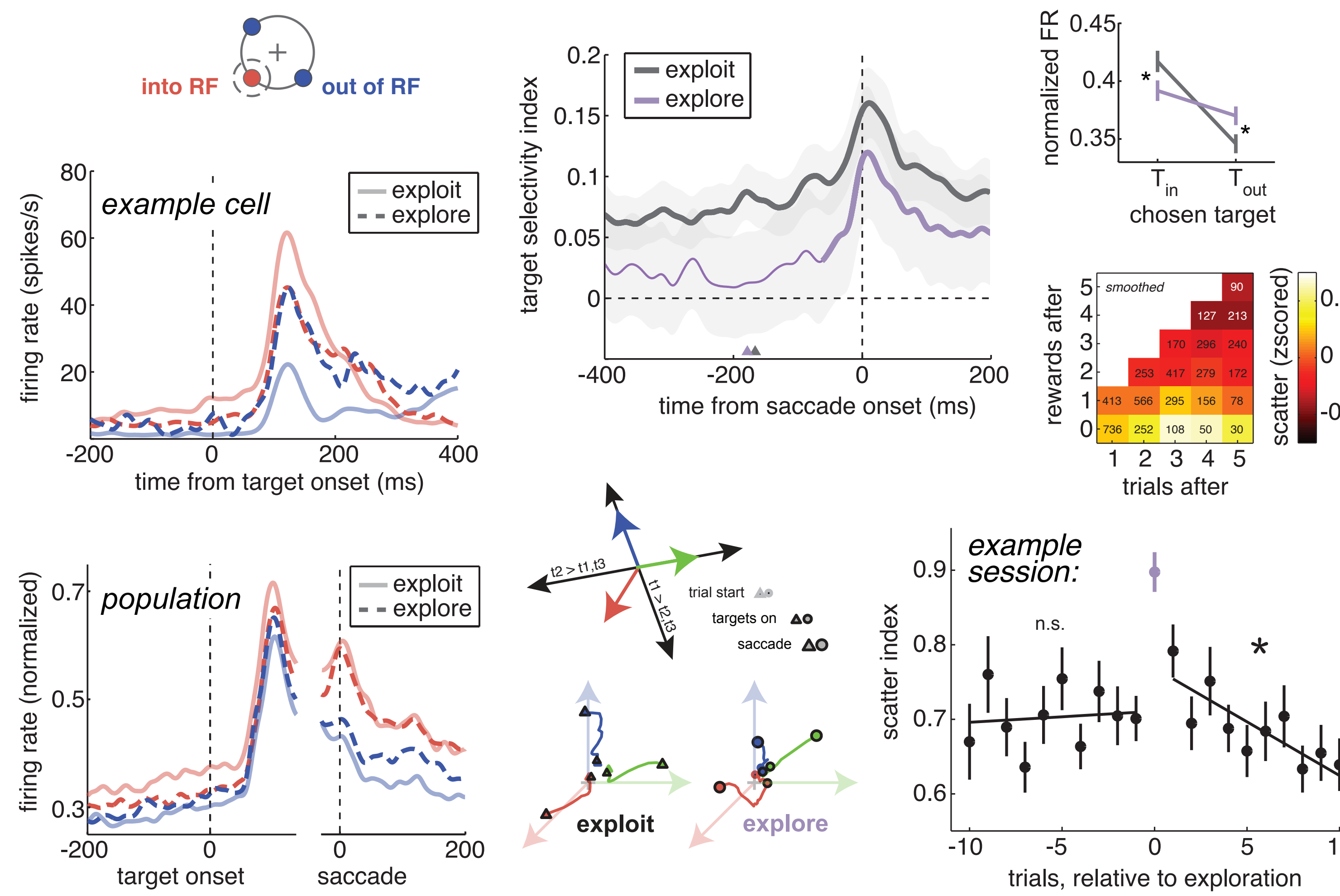## 1. How is exploration implemented in the brain?

We know a lot about how the brain makes choices that maximize reward. However, in variable or uncertain environments, exploiting known-rewarding options is not a sufficient strategy for intelligent decision-making. Instead, actors must also explore: they must sample uncertain alternatives that have the potential to be better. We know a great deal about how the brain generates exploit choices, but much less about the mechanisms underlying explore choices. Here, we ask how exploration and exploitation might be implemented differently in a part of the brain that is responsible for generating decisions: the frontal eye fields.
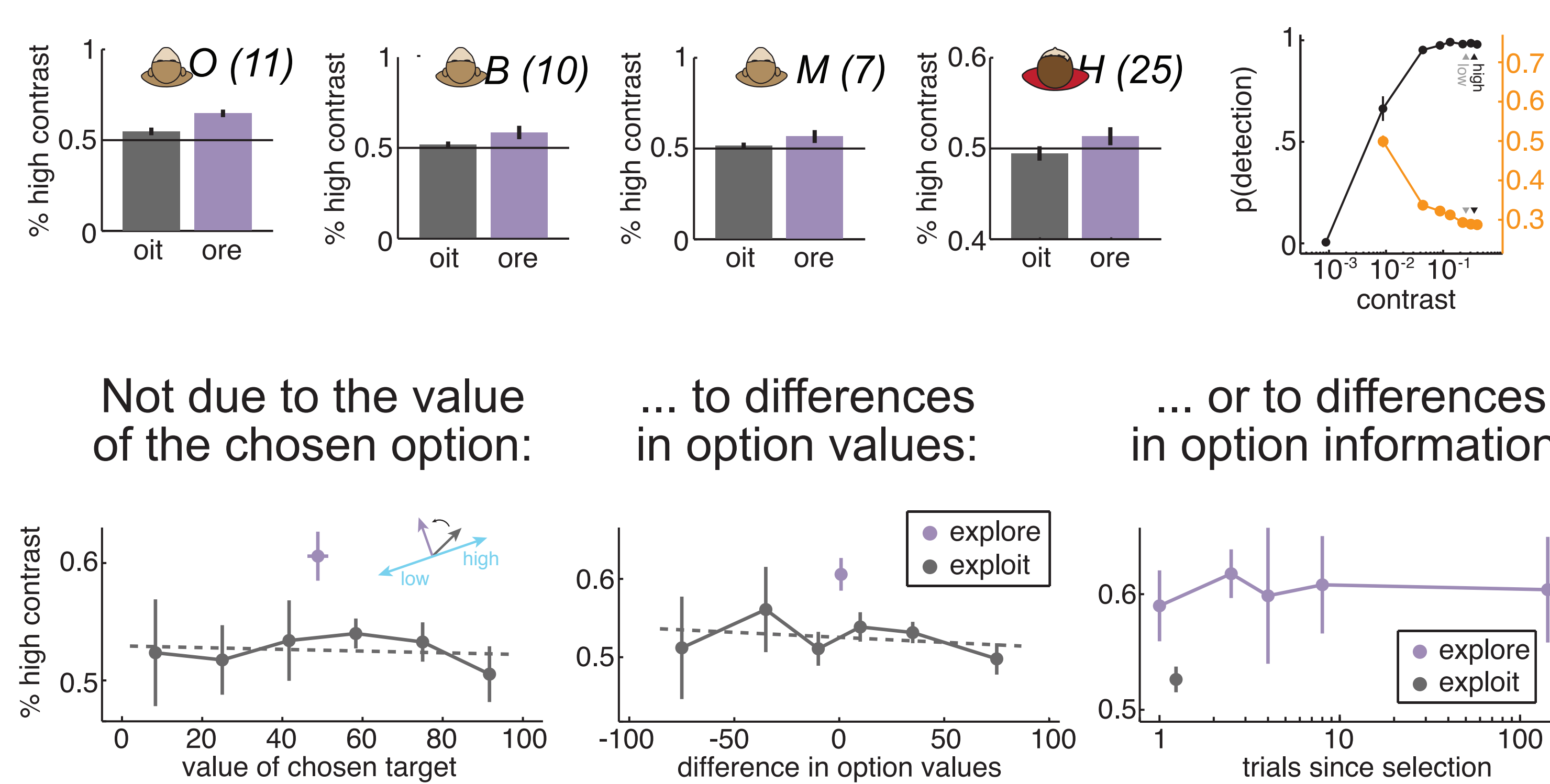
2. Task and explore/exploit choice identification.

Example 2-target reward schedule:

Explore choices are non-reward-maximizing...

... and learning is enhanced

$$f(x) = \beta^{-1} e^{-\frac{x}{\beta}}$$

$$f(x) = \sum_{i=1}^{n} \pi_i \beta_i^{-1} e^{-\frac{x}{\beta_i}}$$

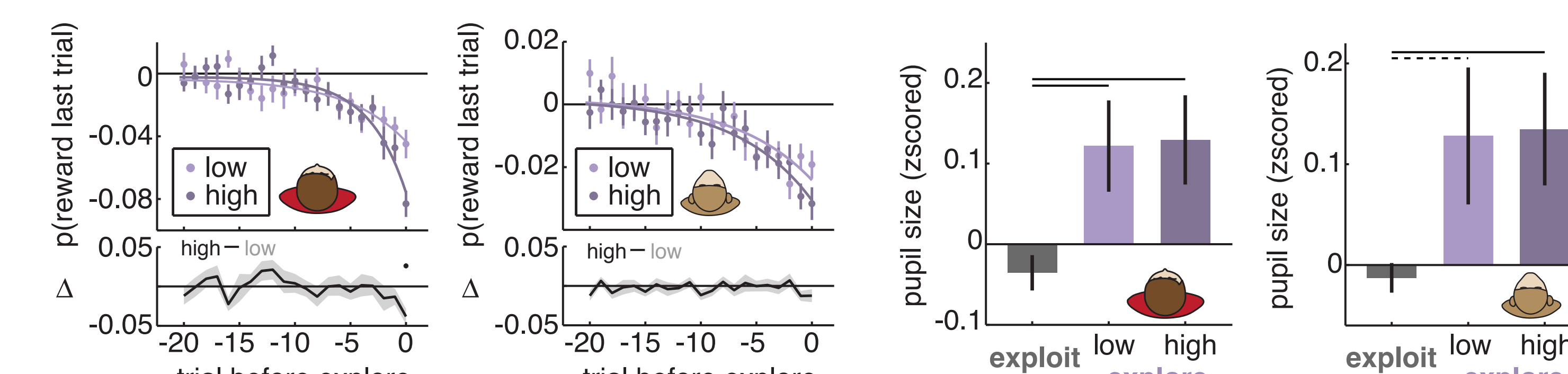$$dX_t = \Delta U(X_t)dt + \sigma_{R_t} dW_t$$

3. Classic choice-predictive signals in prefrontal cortex are substantially reduced during exploration.

into RF   out of RF
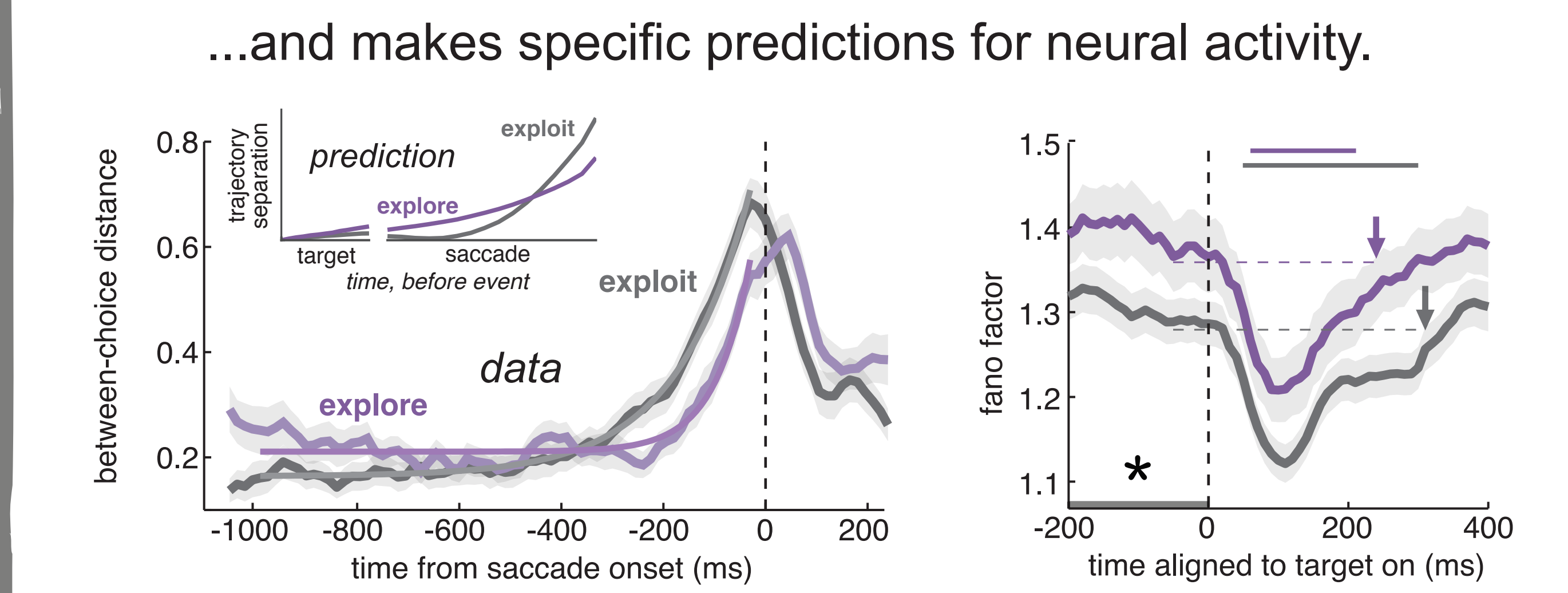
example cell

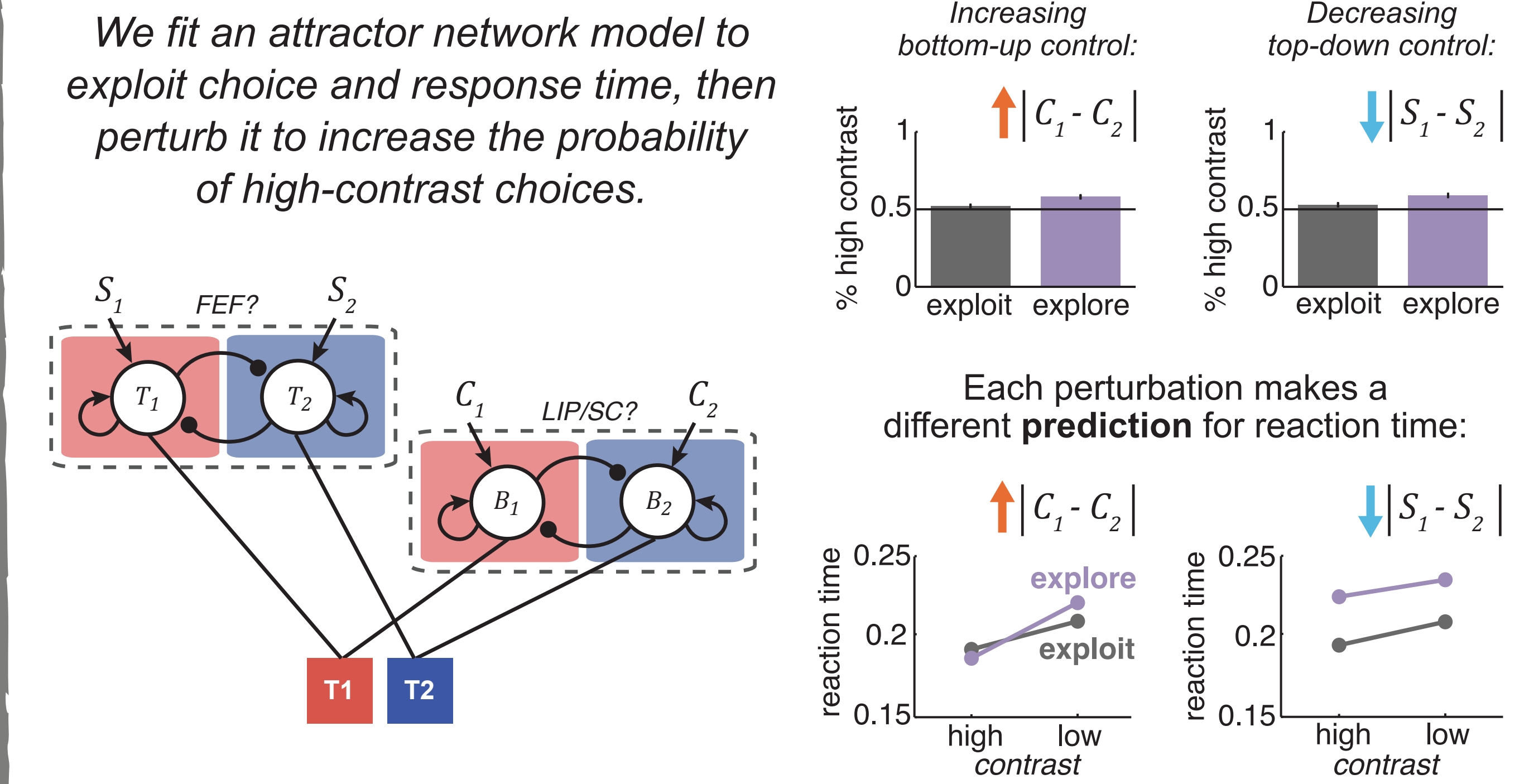*population*

example session:

4. Humans and monkeys are more influenced by option salience during exploration.

Not due to the value of the chosen option:

... to differences in option values:

... or to differences in option information:

High-contrast explore choices do not look any less explore-like than low-contrast explore choices:

5. A disruption in prefrontal attractor dynamics predicts these changes in brain and behavior.

We fit an attractor network model to exploit choice and response time, then perturb it to increase the probability of high-contrast choices.

Increasing bottom-up control: ↑$|C_1 - C_2|$

Decreasing top-down control: ↓$|S_1 - S_2|$

Each perturbation makes a different **prediction** for reaction time:

The data is consistent with ↓$|S_1 - S_2|$

Different contrast: *Same contrast:*

This perturbation changes model dynamics

Model fit to exploit

fast trajectory separation — insensitive to initial condition

slow trajectory separation — sensitive to initial condition

...and makes specific predictions for neural activity.

prediction

data

6. During exploration,

○ FEF activity is less predictive of choice.

○ Decisions are less determined by top down control.

○ A model in which attractor dynamics are disrupted can replicate behavior and predicts counter-intuitive features of neural activity.