

Injection of dopamine type 2 antagonist decreases inverse temperature in reinforcement learning

RLDM 2015, June 7-10 @ University of Alberta, Edmonton (#T42)

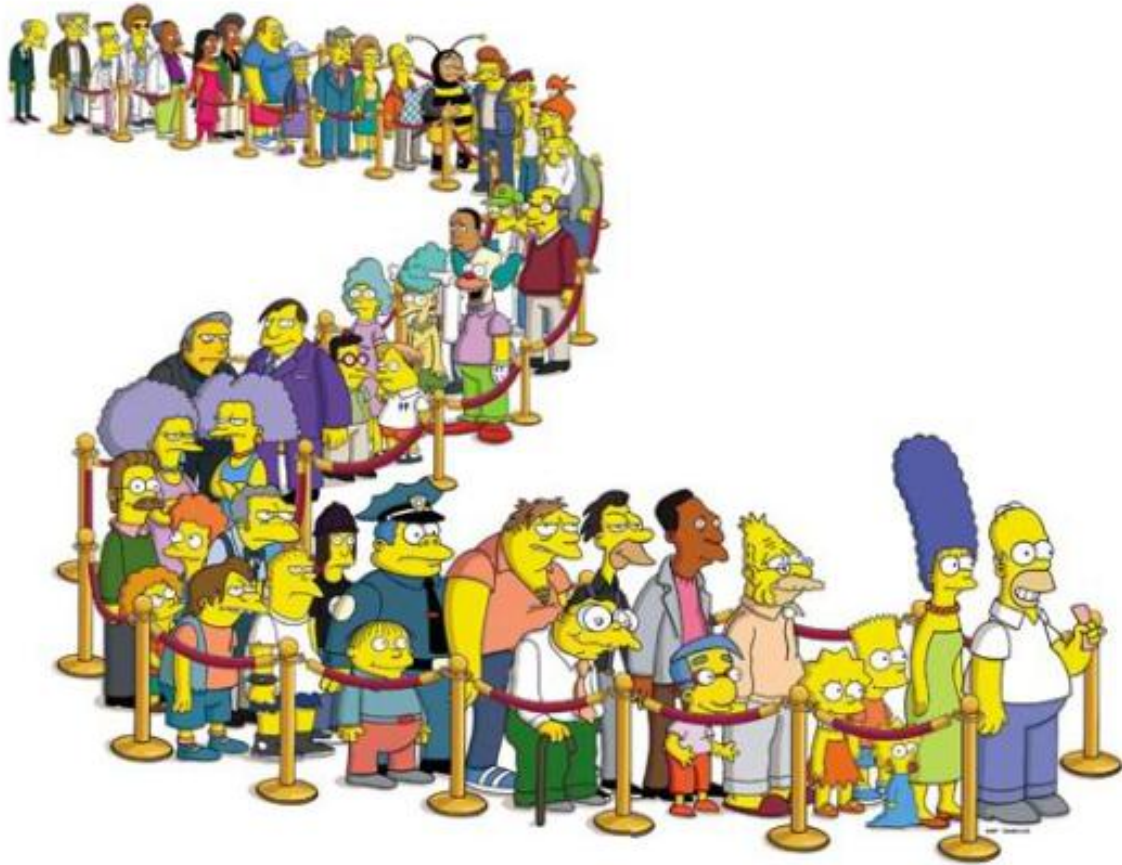
Eunjeong Lee*,
Olga Dal Monte,
Bruno B. Averbeck

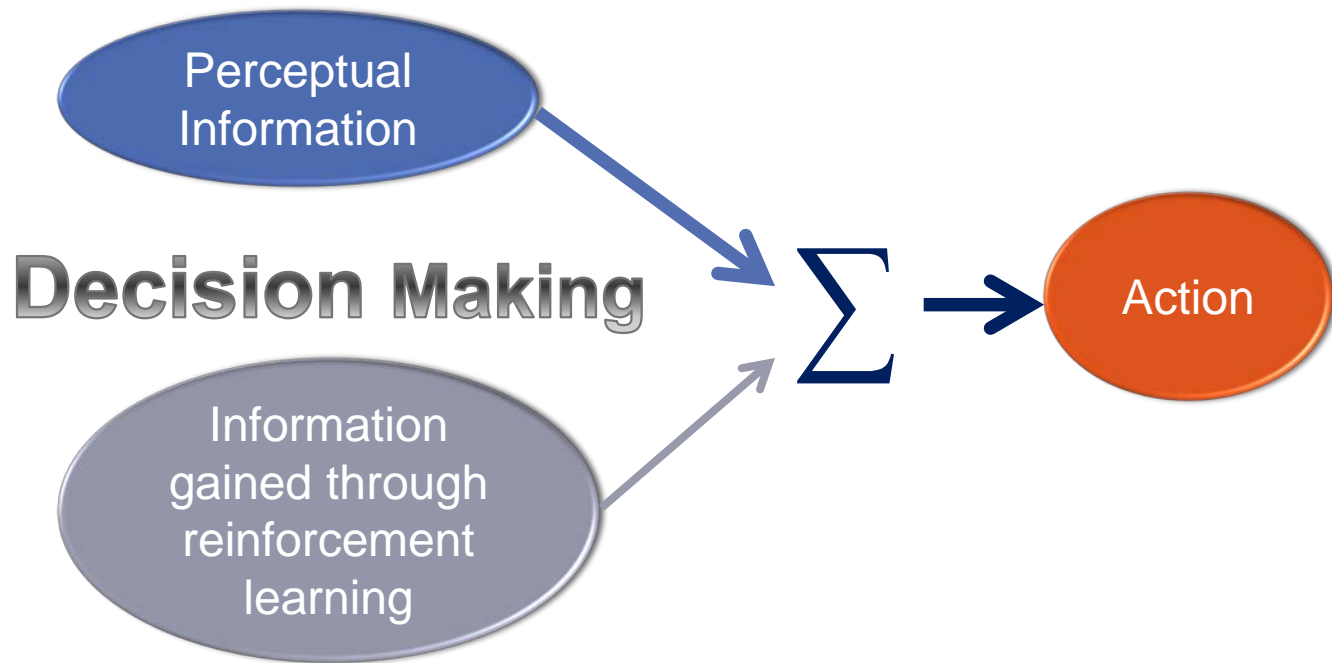


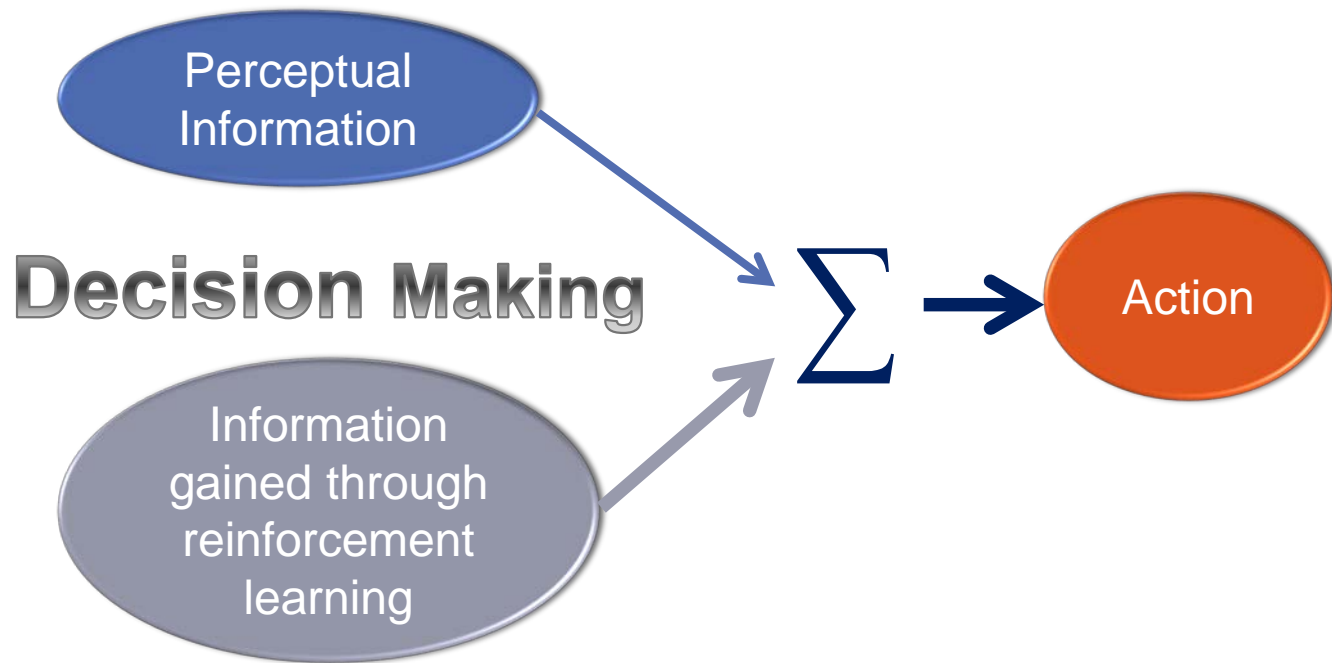
National Institute
of Mental Health

Unit on Learning and Decision Making
Laboratory of Neuropsychology
National Institute of Mental Health/
National Institutes of Health (NIH)

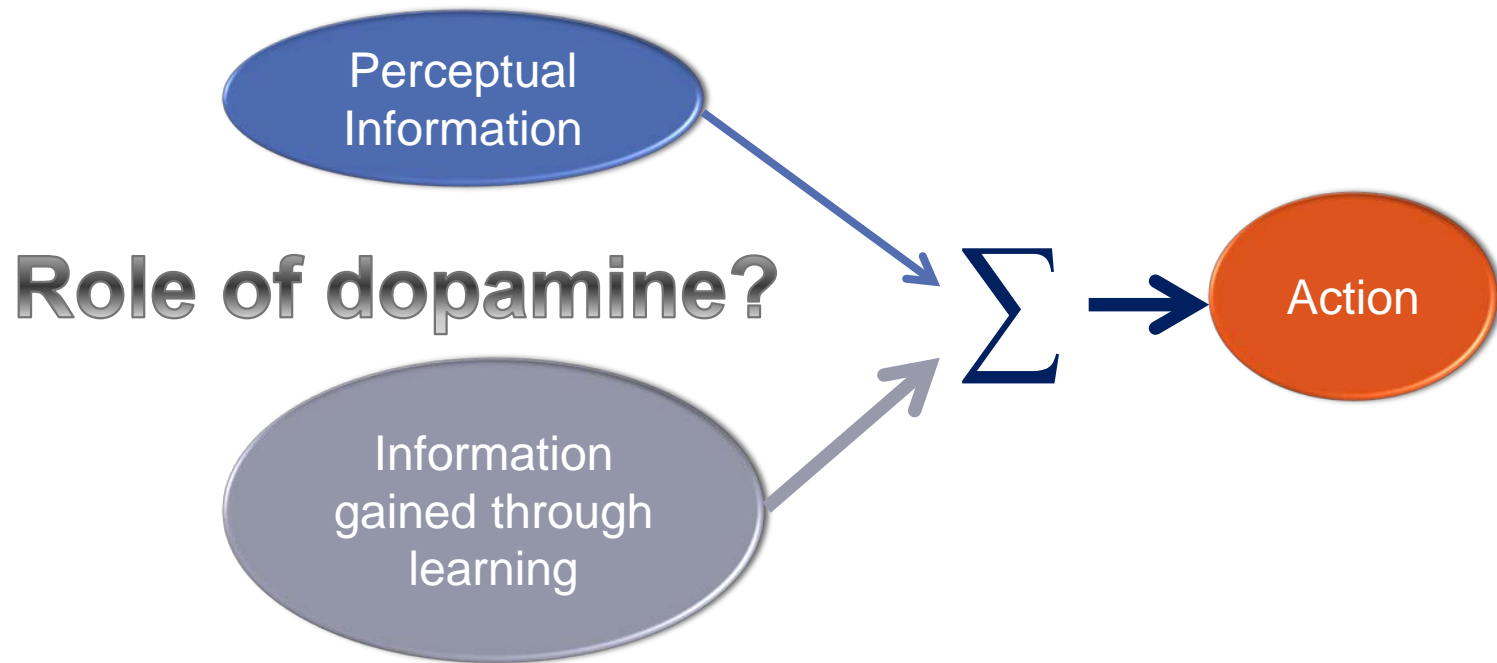








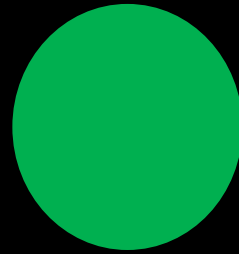
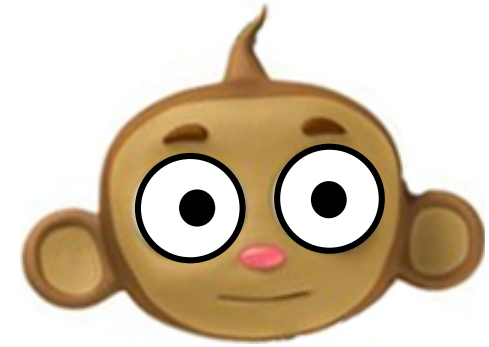
How is dopamine in the dorsal striatum involved in the decision making processes?



Oculomotor sequential decision making task



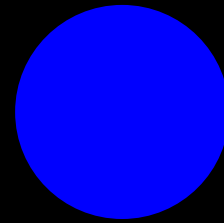
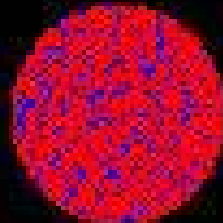
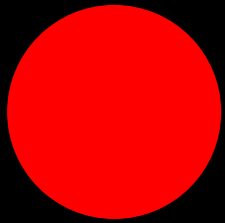
Stimulus and movement



Stimulus and movement



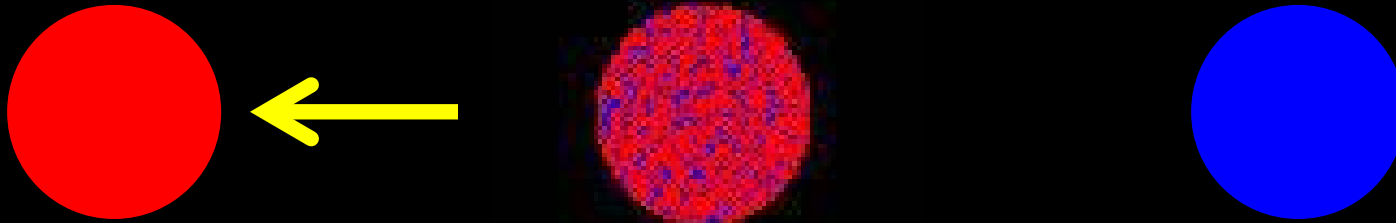
Color bias : more red or more blue pixels



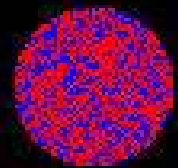
Stimulus and movement



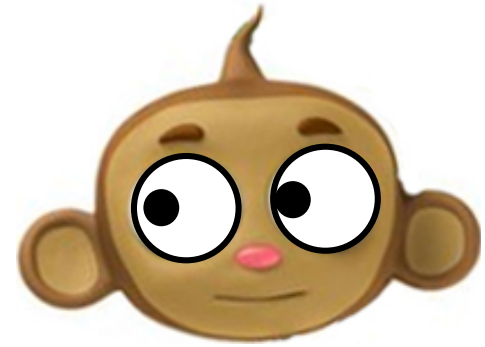
Color bias : more red or more blue pixels



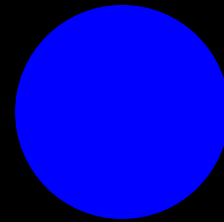
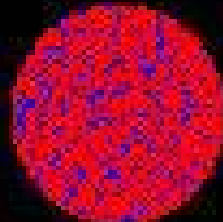
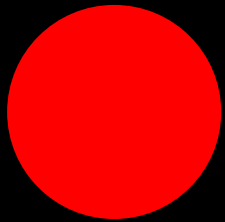
Difficult Choice



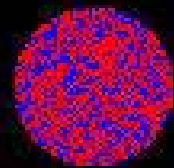
Stimulus and movement



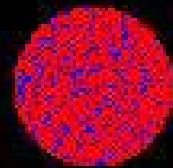
Color bias : more red or more blue pixels



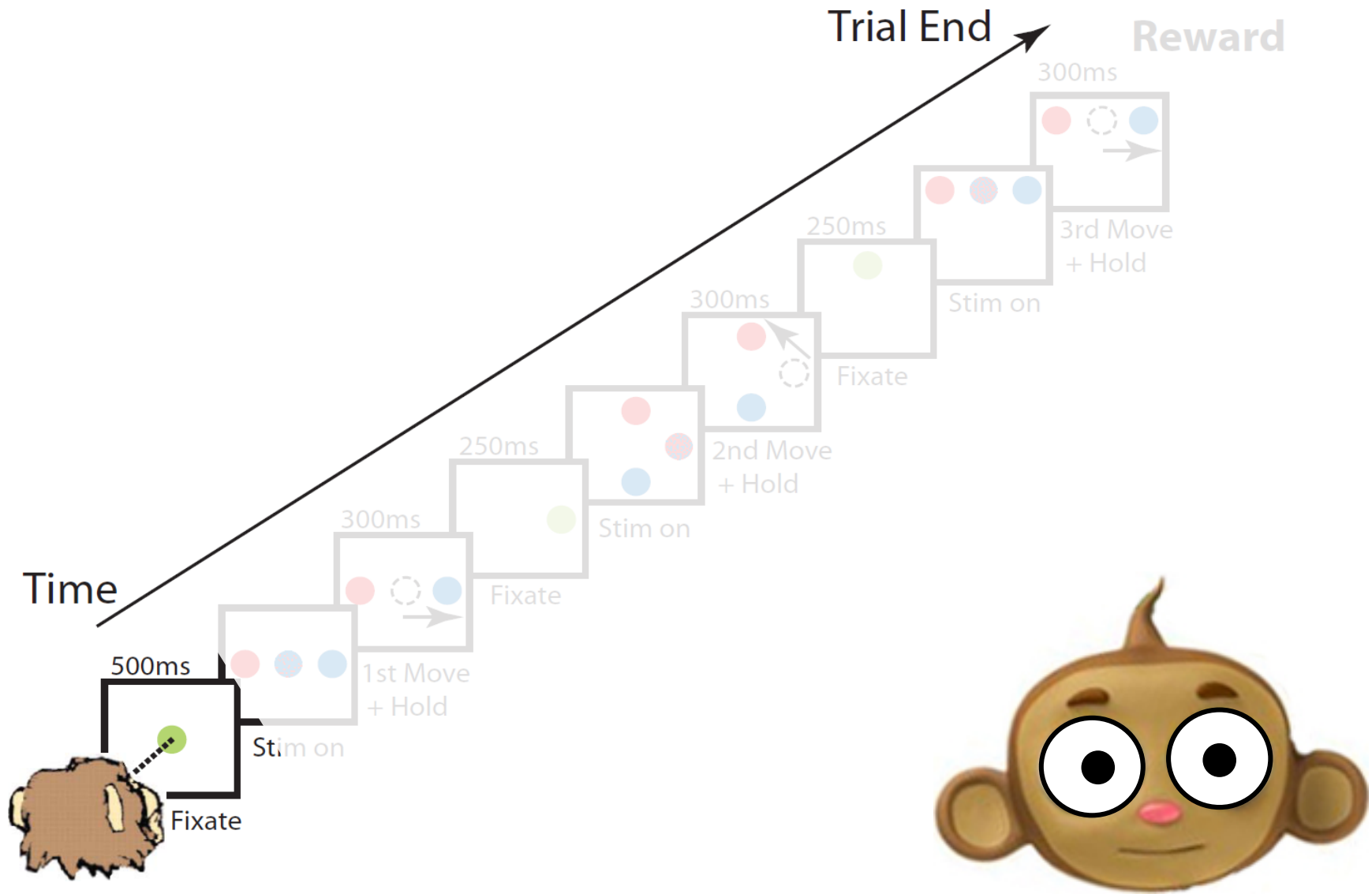
Difficult Choice



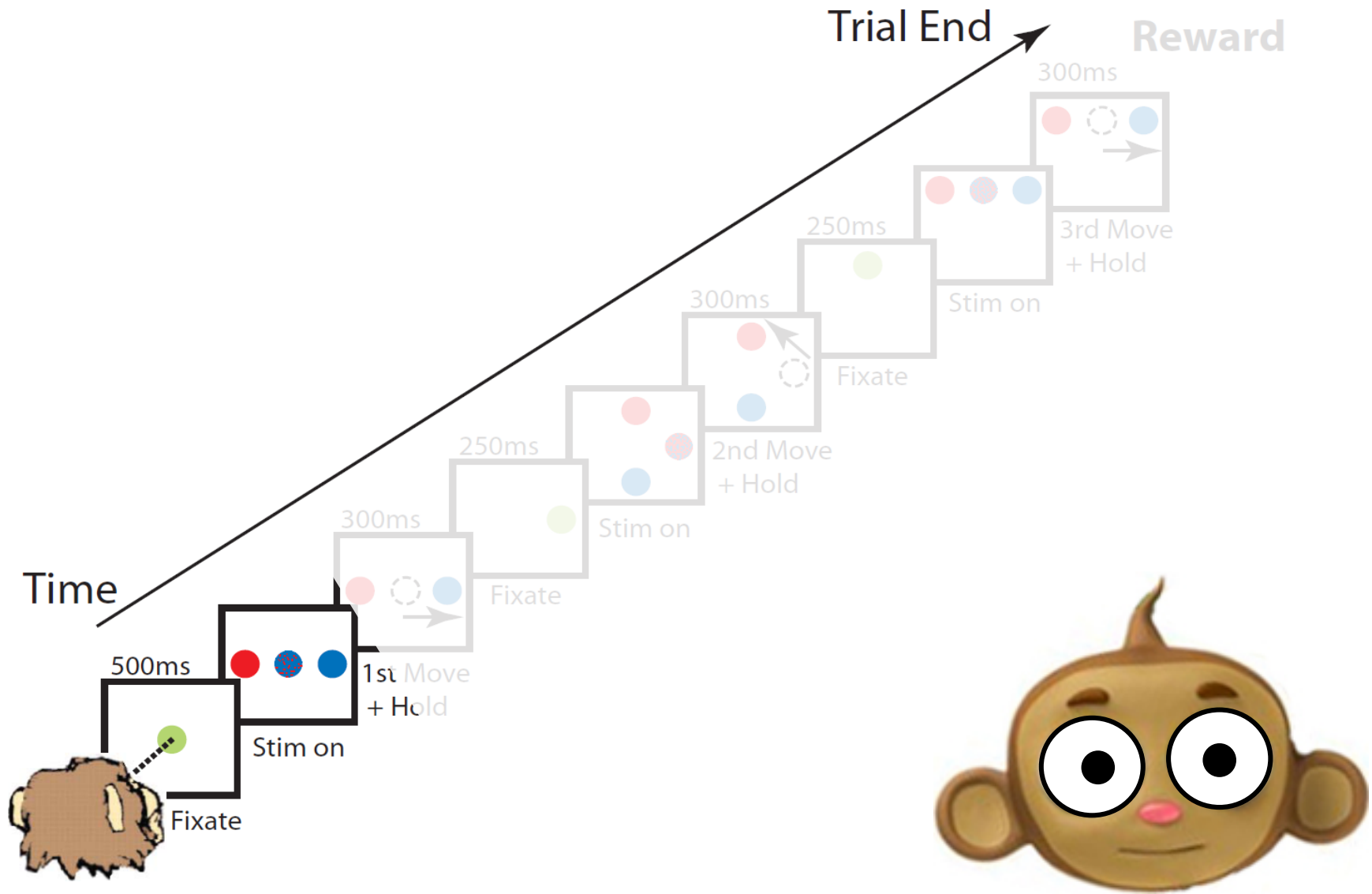
Easy Choice



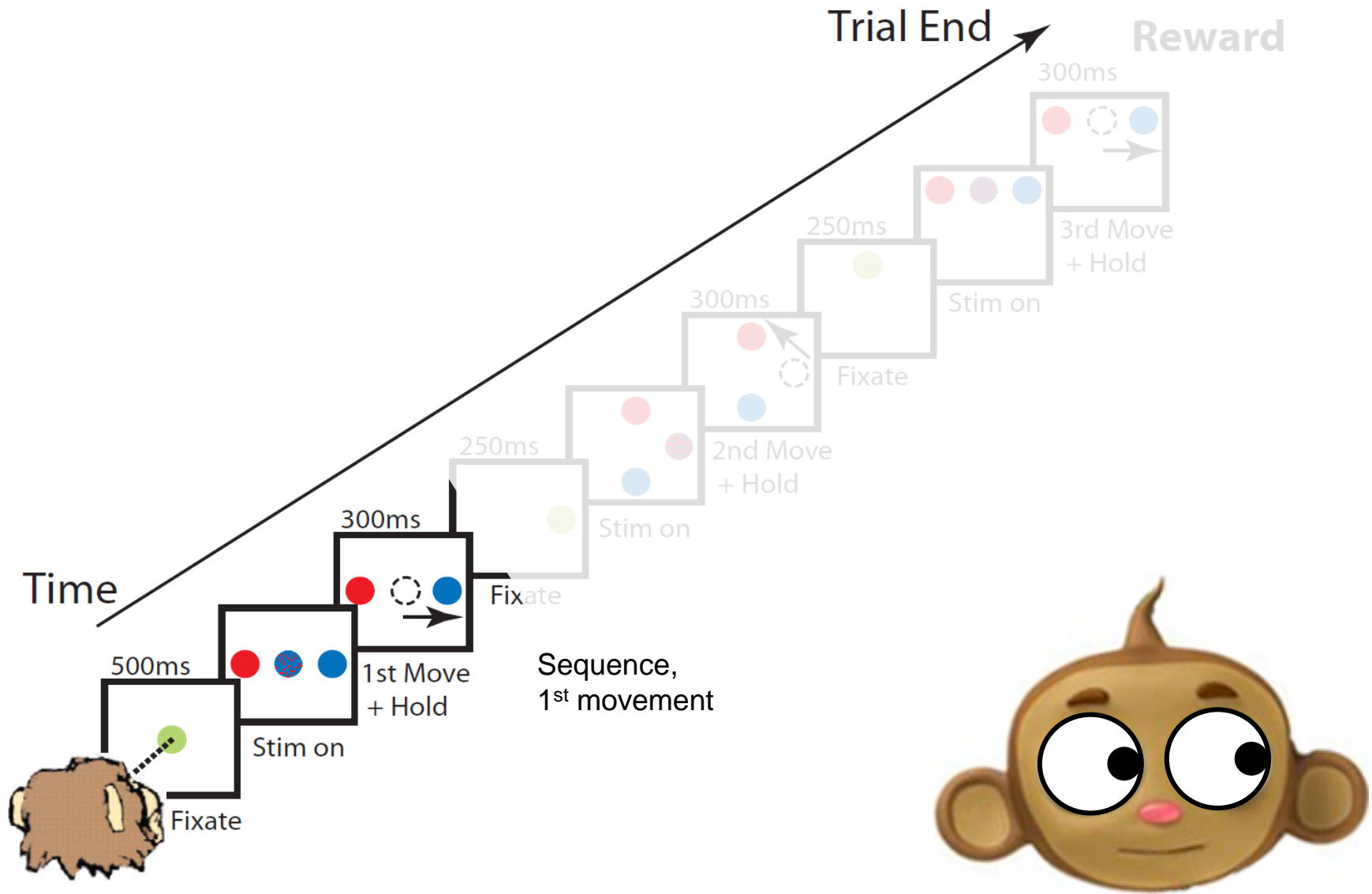
Temporal sequence of choices in a single trial



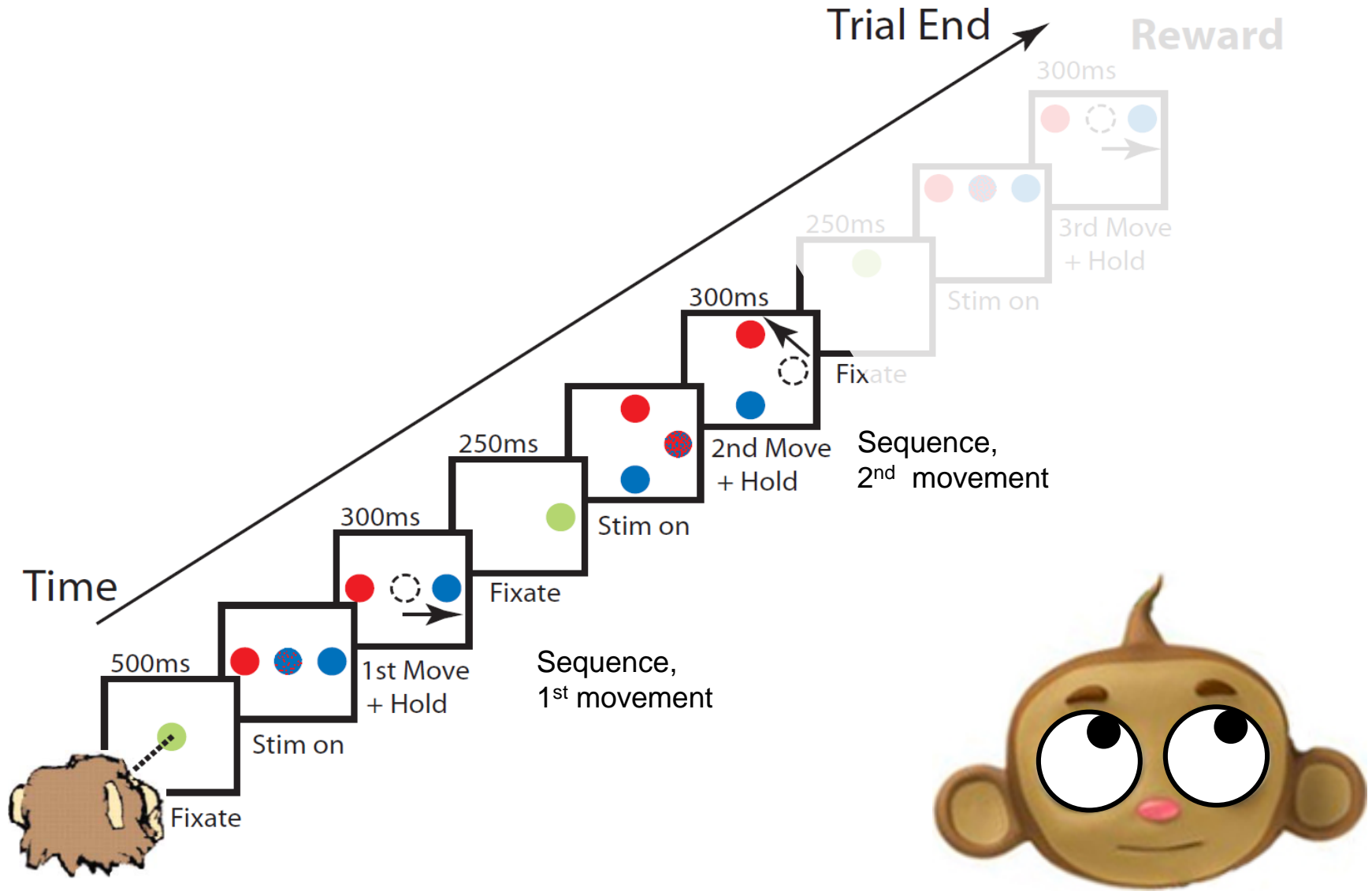
Temporal sequence of choices in a single trial



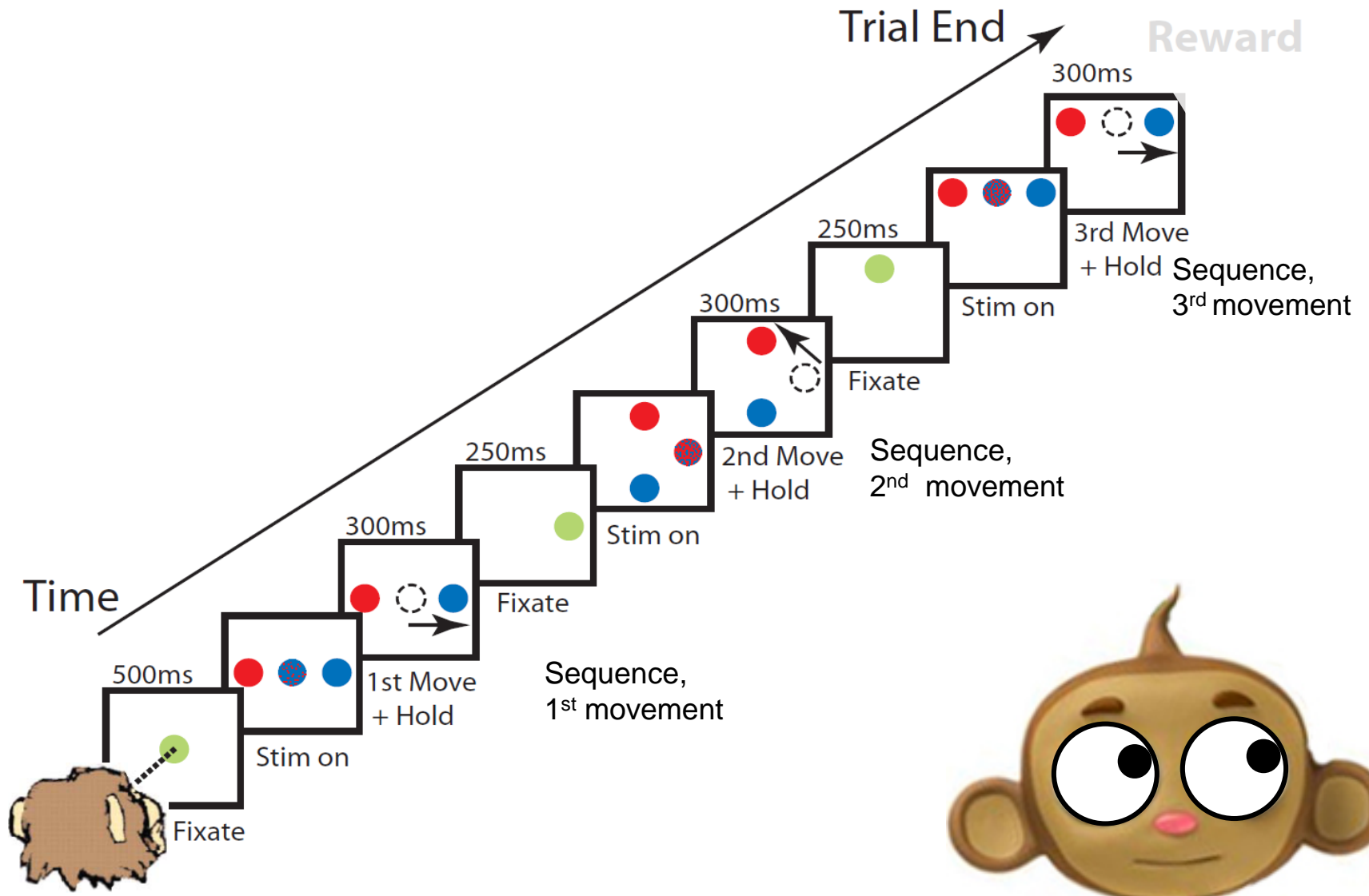
Temporal sequence of choices in a single trial



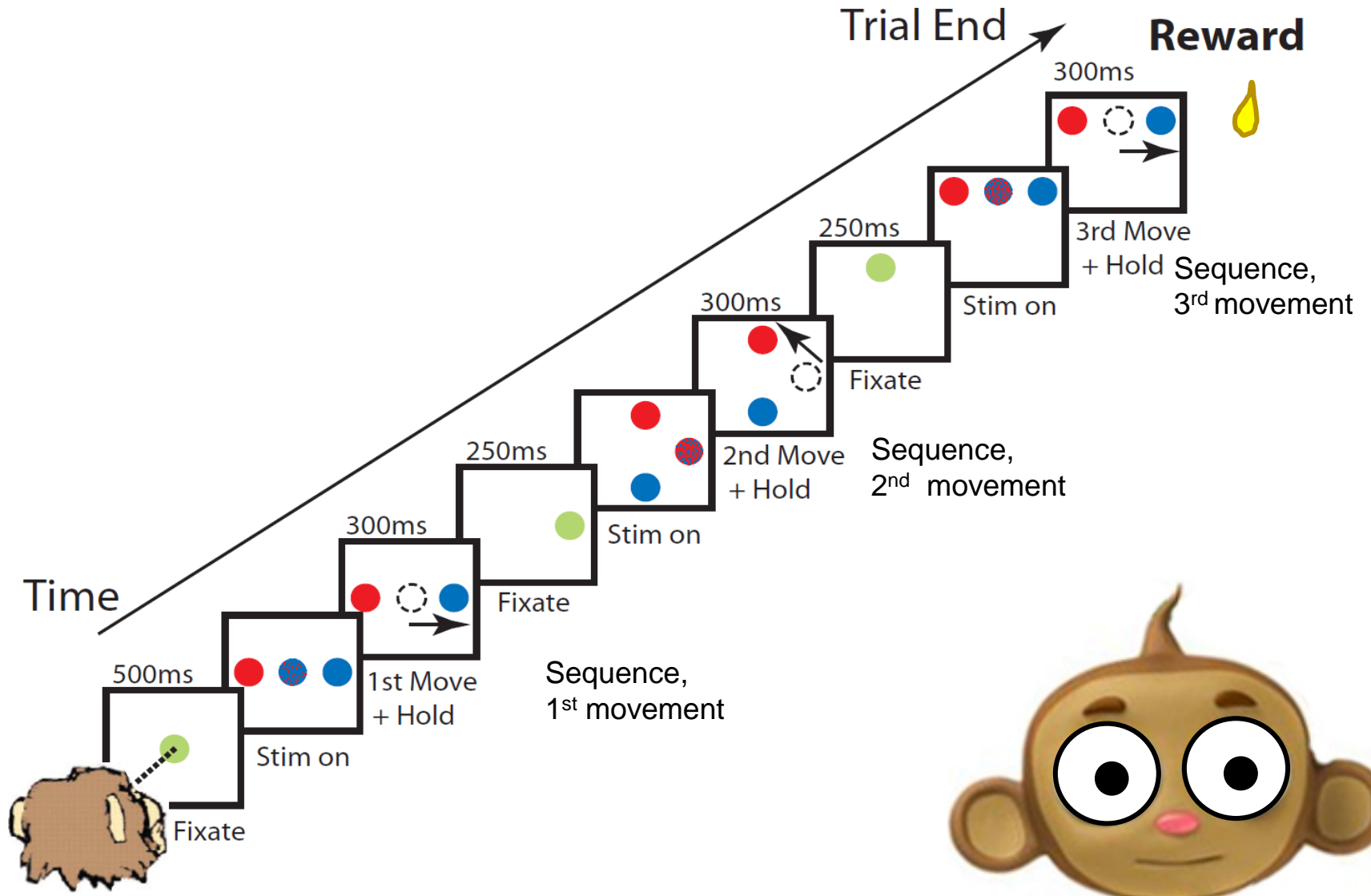
Temporal sequence of choices in a single trial



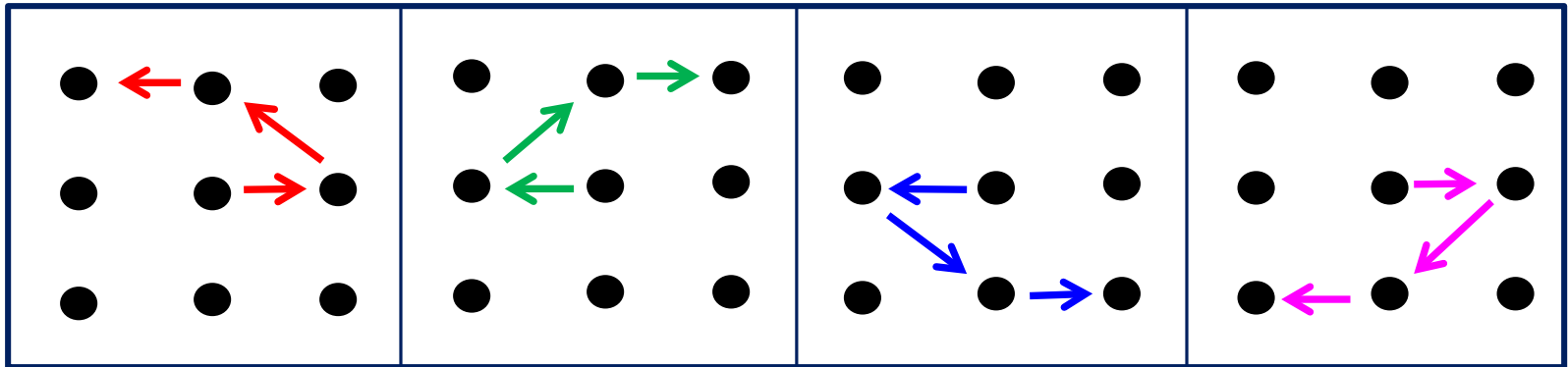
Temporal sequence of choices in a single trial



Temporal sequence of choices in a single trial



Eight possible sequences

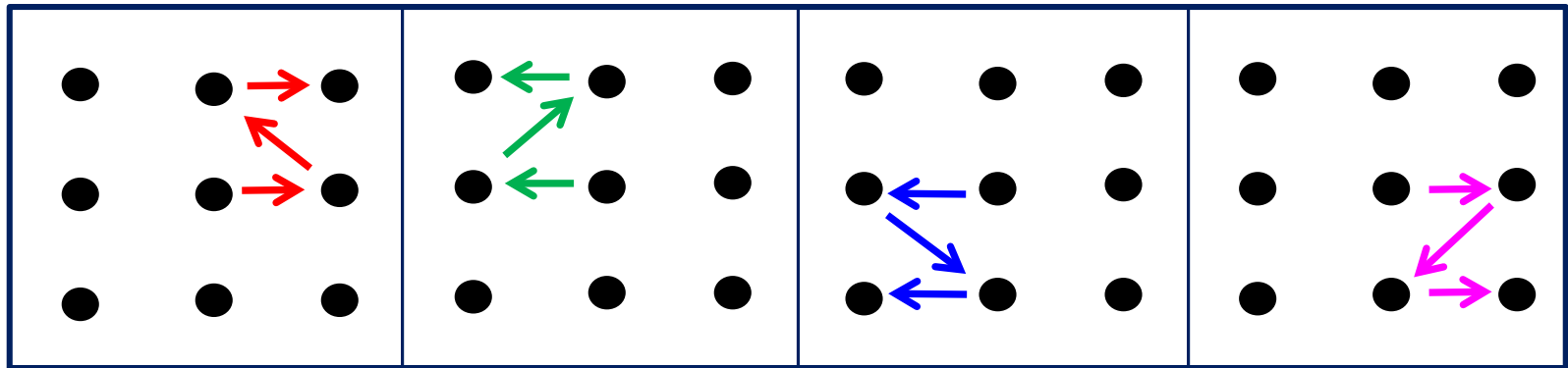


S1

S2

S3

S4



S5

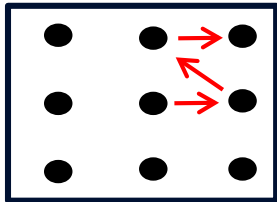
S6

S7

S8

Task conditions

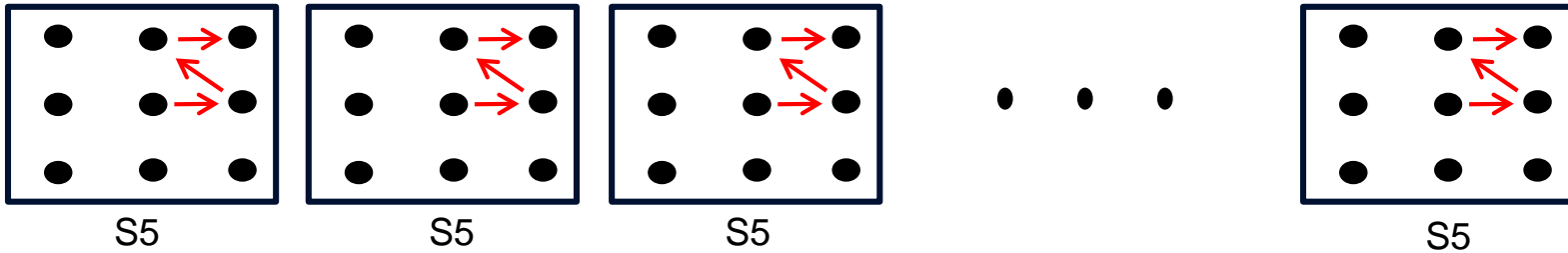
Fixed condition: The sequence of eye movements is fixed for 8 correct trials then switches to a new sequence and remains fixed again.



S5

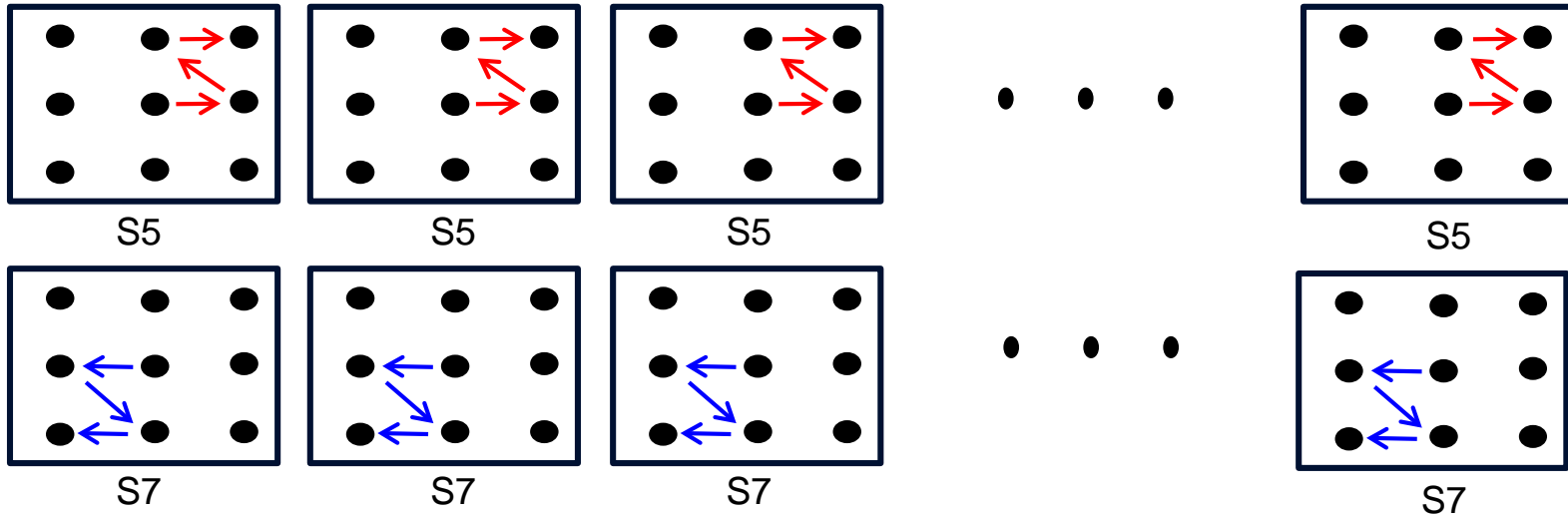
Task conditions

Fixed condition: The sequence of eye movements is fixed for 8 correct trials then switches to a new sequence and remains fixed again.



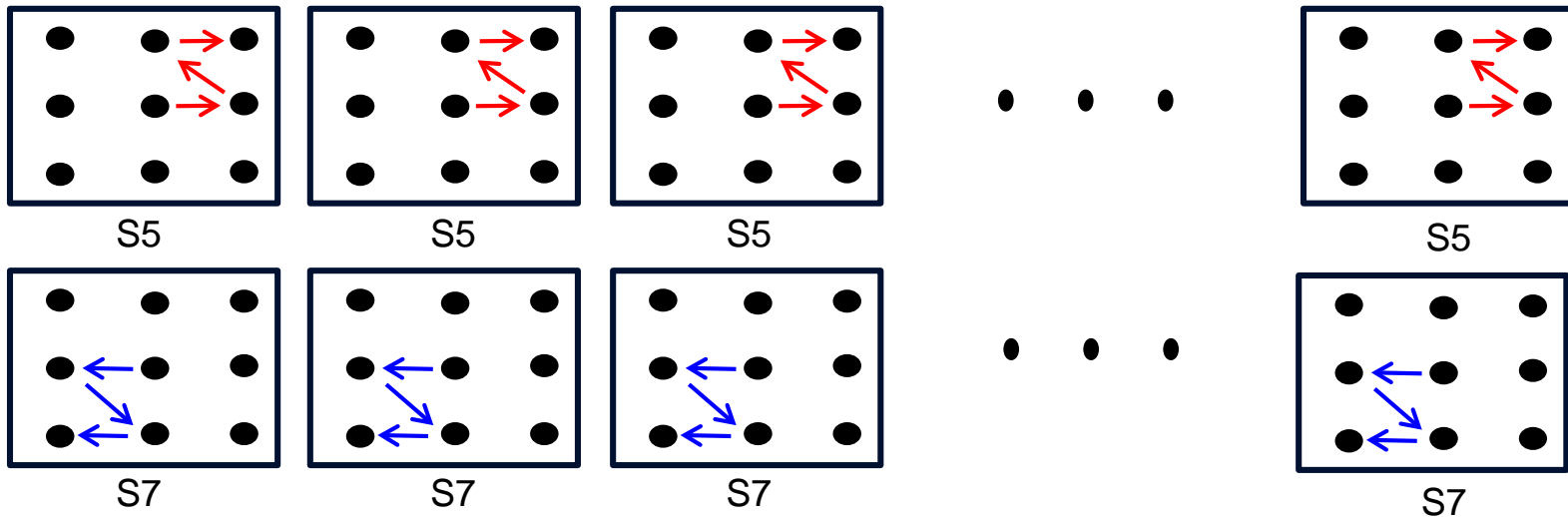
Task conditions

Fixed condition: The sequence of eye movements is fixed for 8 correct trials then switches to a new sequence and remains fixed again.

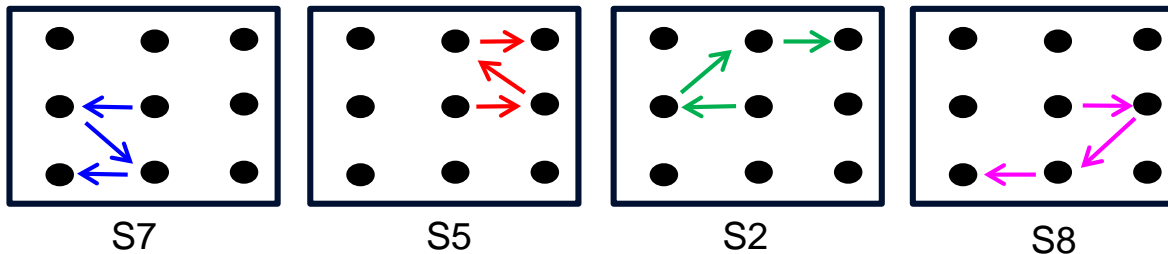


Task conditions

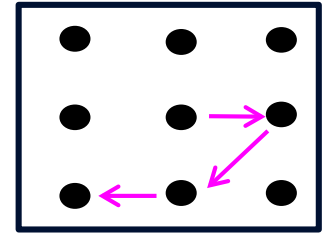
Fixed condition: The sequence of eye movements is fixed for 8 correct trials then switches to a new sequence and remains fixed again.



Random condition: The sequence changes every trial.



Fixed vs. random condition



S8

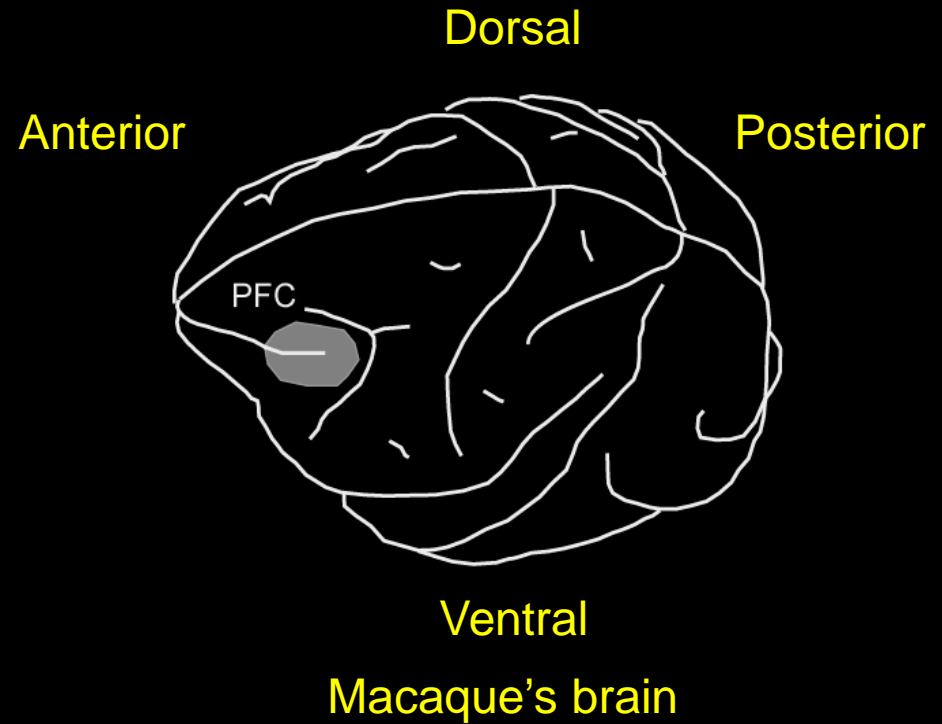
In the random condition the animal is forced to rely on the fixation stimulus (mixture of two colors) to determine movement direction.

(Decision is based on immediately available perceptual information)

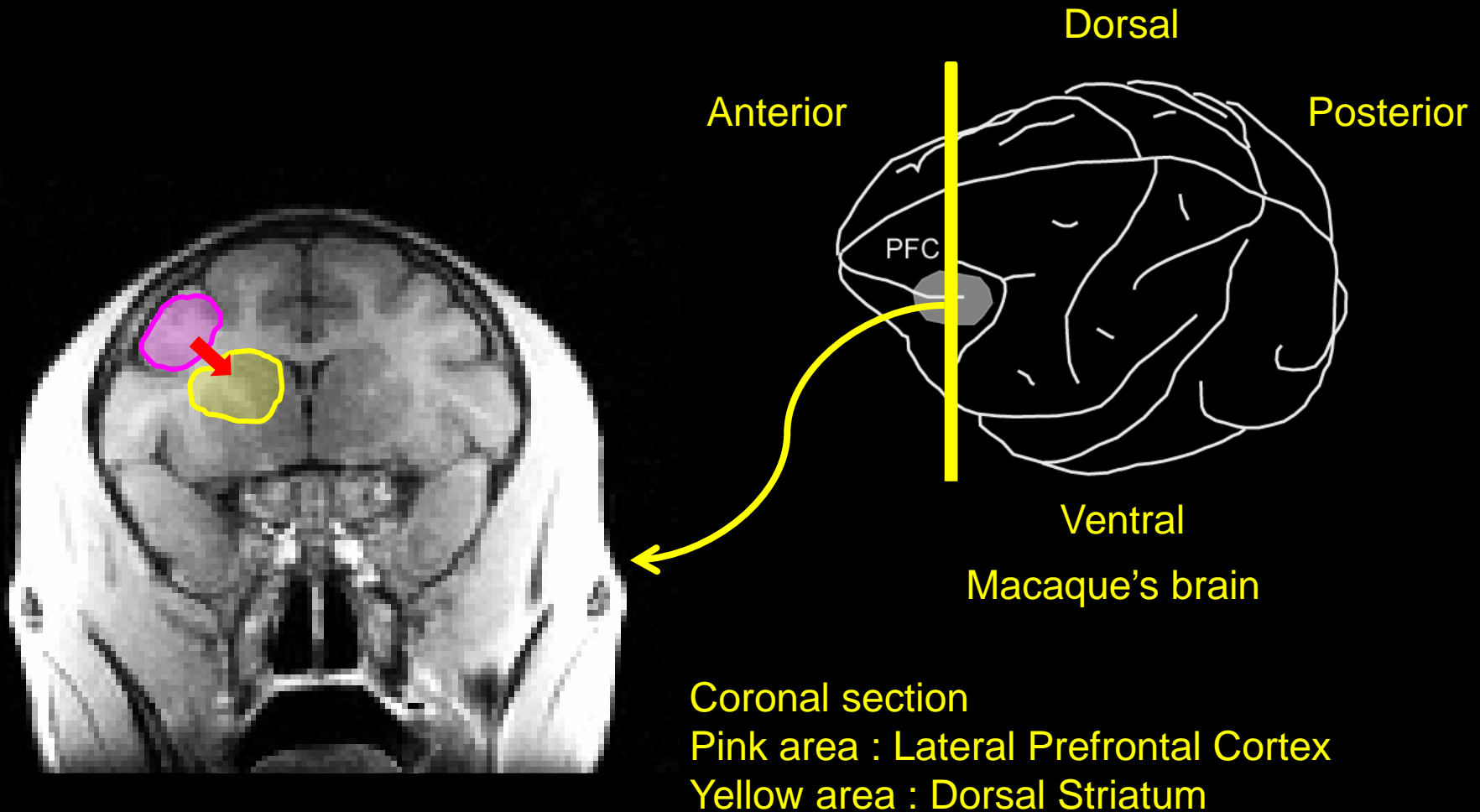
In the fixed condition, after a few consecutive trials of a given sequence the animal can utilize knowledge of the sequence to determine the correct movement direction.

(Decision is based on learned sequence information)

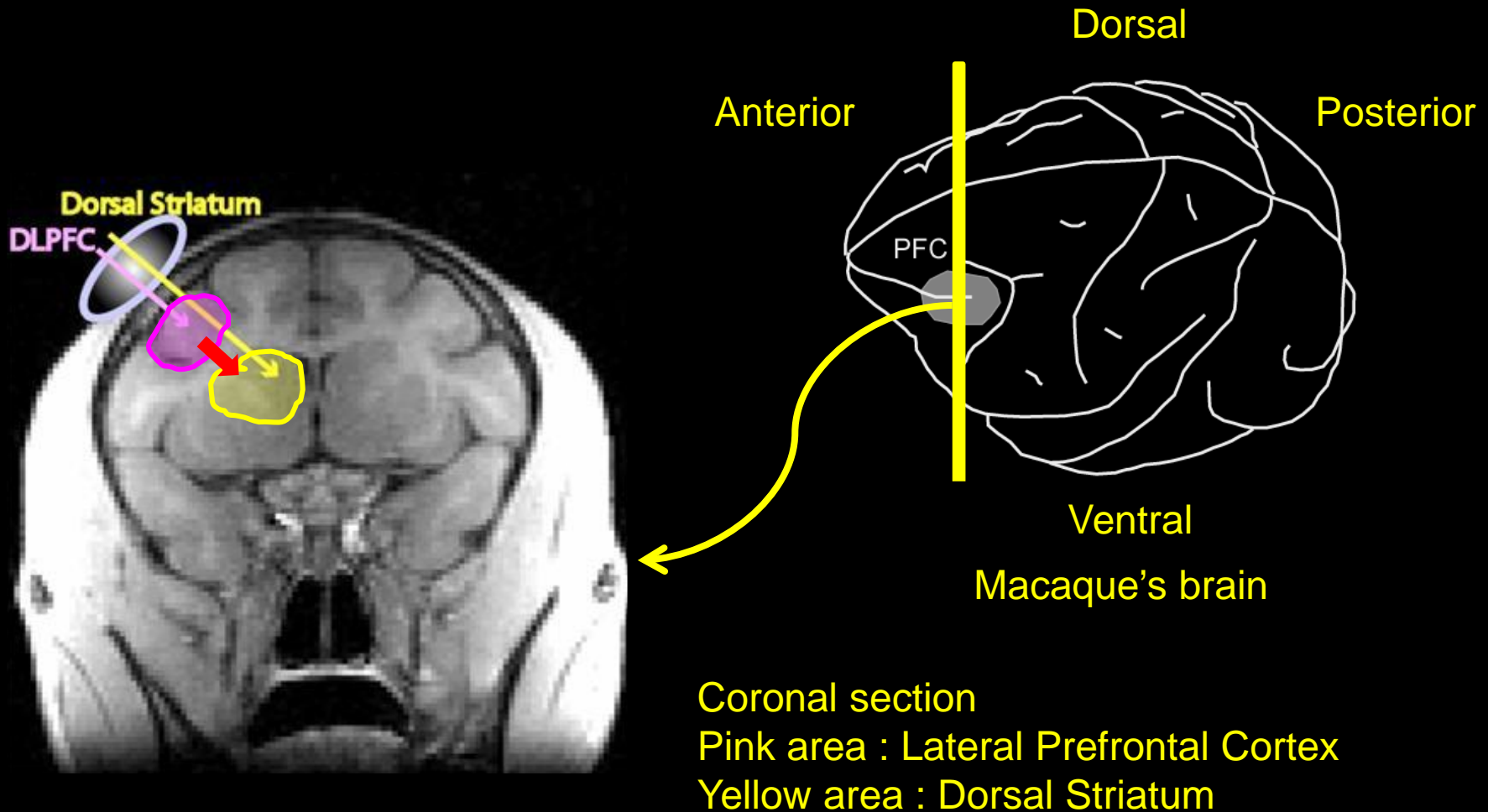
Prefrontal Cortex and Striatum



Prefrontal Cortex and Striatum



Prefrontal Cortex and Striatum



Reinforcement Learning

Action Value (current) =
Action Value (previous) +
learning rate * (reward(current) – Action Value (previous))

$$V_i(t) = V_i(t - 1) + \rho_f * RPE,$$
$$RPE = (r(t) - V_i(t - 1))$$

Rewards $r(t)$: correct actions = 1,
 incorrect actions = 0

Selecting an action:
Dependency on past outcomes and color bias in the fixation stimulus.

$$d_i(t) = \frac{e^{(\beta v_i(t) - \gamma CB)}}{\sum_{i=1}^2 e^{(\beta v_i(t) - \gamma CB)}}$$

$$\arg \min(-\sum_{t=1}^T \log(d_i(t)c_i(t) + (1 - d_i(t))(1 - c_i(t))))$$

β : inverse temperature parameter = choice consistency

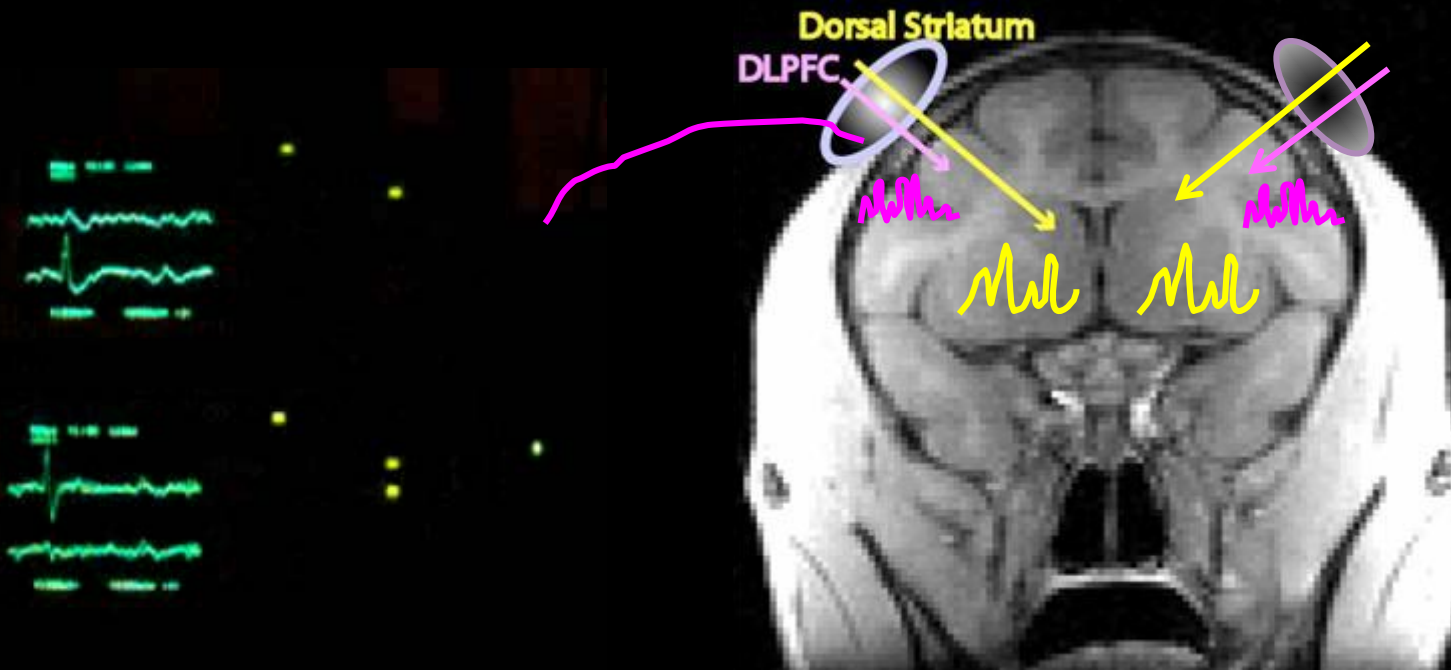
$\beta \uparrow$: noise (response variability) decrease

$\beta \downarrow$: noise (response variability) increase

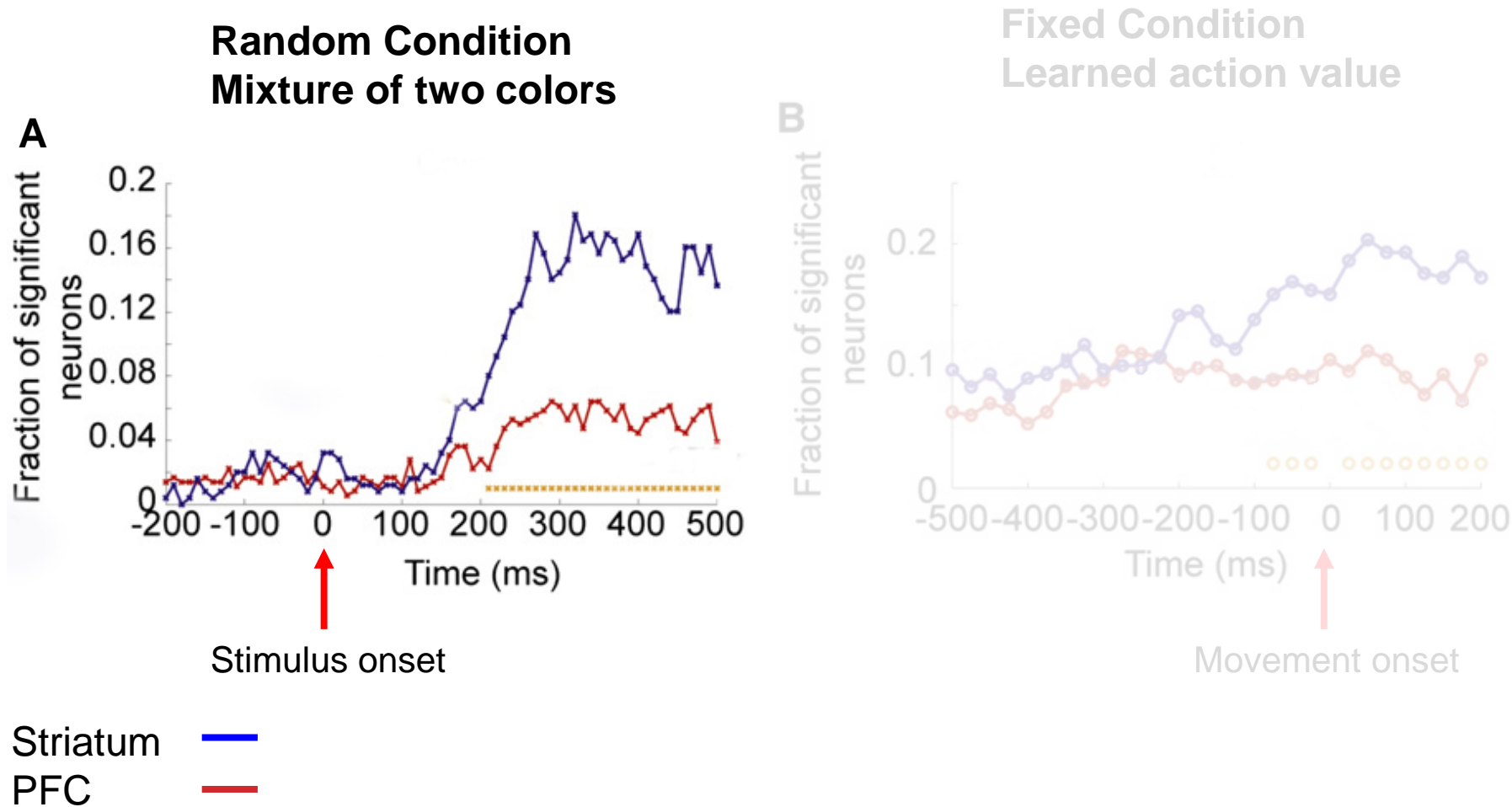
γ : weight of the fixation stimulus in the choice process

c_i : chosen action i , action 1 = 1, action 2 = 0

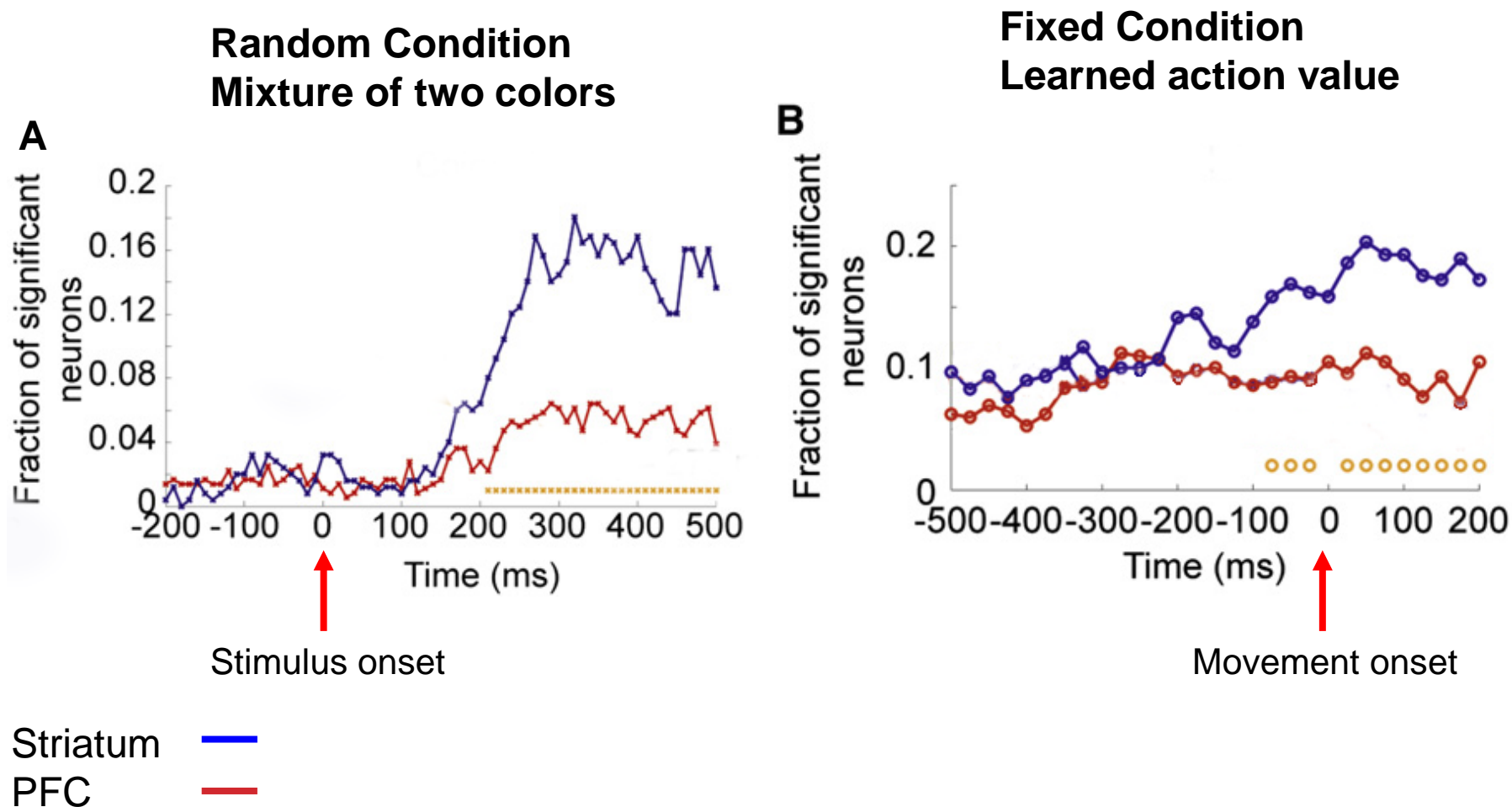
How do striatal neurons represent the decision making process?



The dStr represents the value of actions driven by immediately available perceptual information or past reinforcement.



The dStr represents the value of actions driven by immediately available perceptual information or past reinforcement.

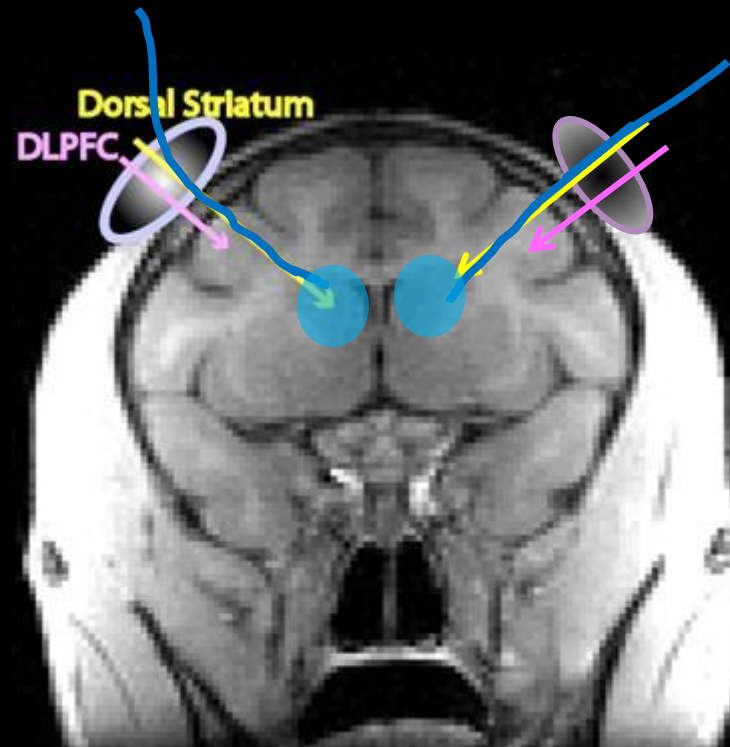


Does dopamine mediate the transition from perceptual inference to reinforcement learning?

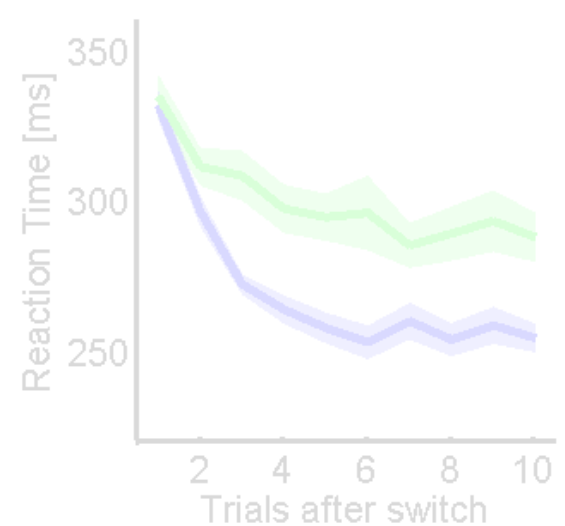
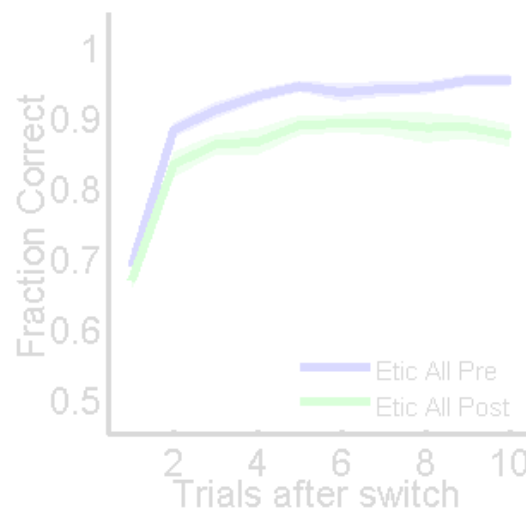
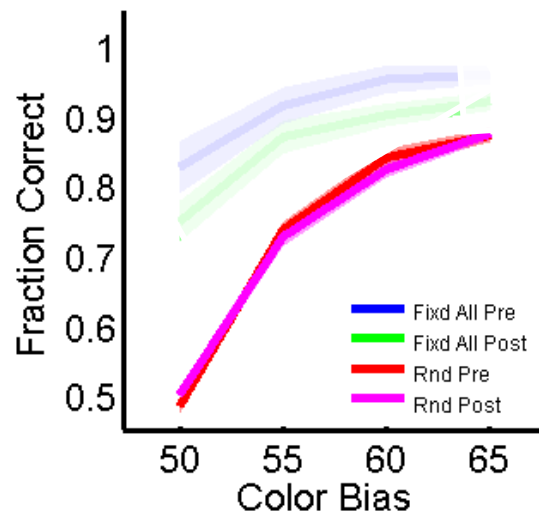
Experiment steps

1. Pre-injection
2. Drug injection
3. Post-injection

* side A - 16 channels
side B - 7 channels
(total 23 channels max,
each 4 channel used for microinjection)



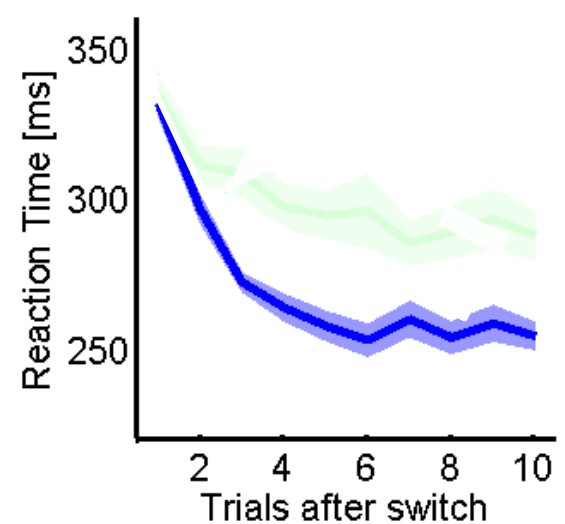
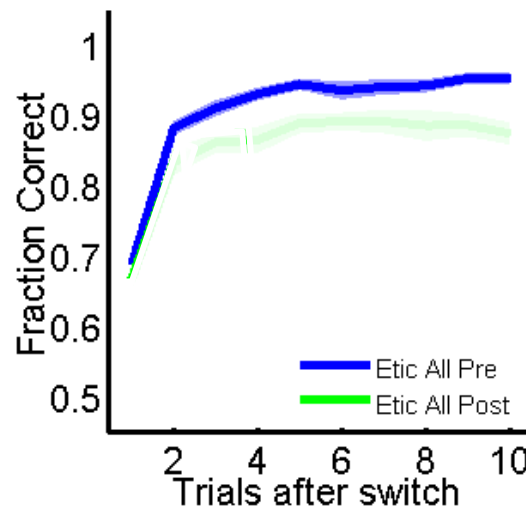
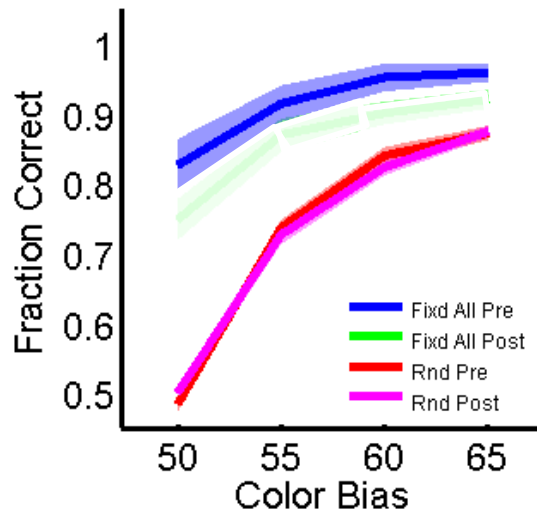
Use of reinforcement learning from previous trials and immediately available information in the fixation stimulus



Fixed condition
(knowledge of sequence)
Pre-Fixed —
Post-Fixed —

Random condition
(pay attention to stimulus)
Pre-Rand —
Post-Rand —

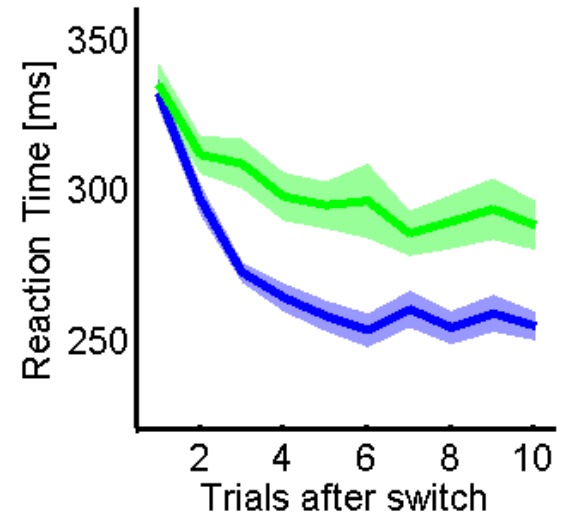
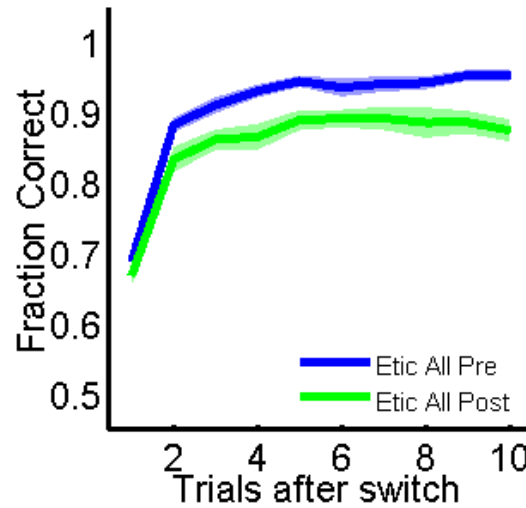
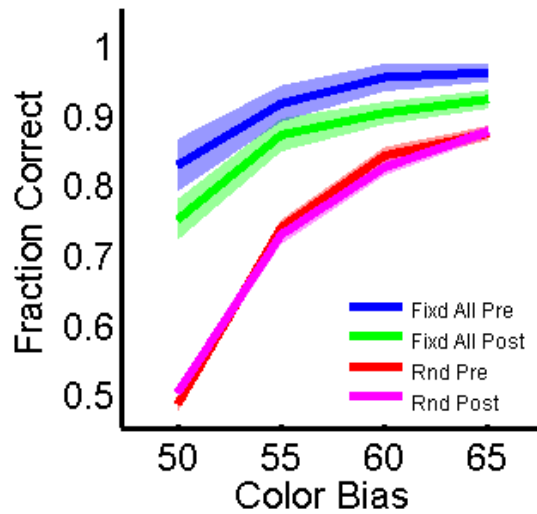
Use of reinforcement learning from previous trials and immediately available information in the fixation stimulus



Fixed condition
(knowledge of sequence)
Pre-Fixed ——— (blue)
Post-Fixed ——— (green)

Random condition
(pay attention to stimulus)
Pre-Rand ——— (red)
Post-Rand ——— (magenta)

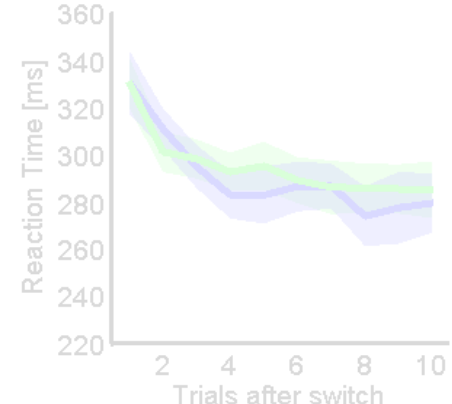
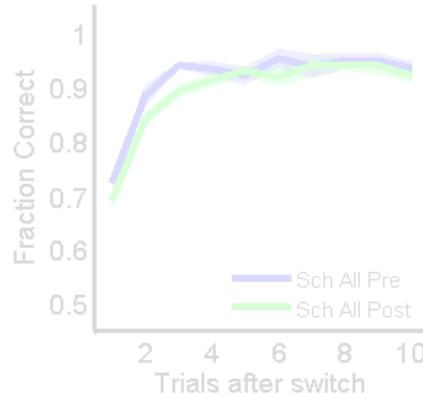
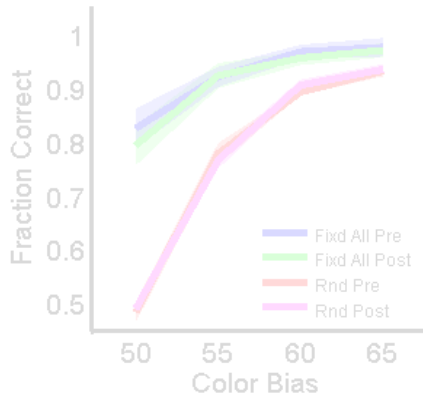
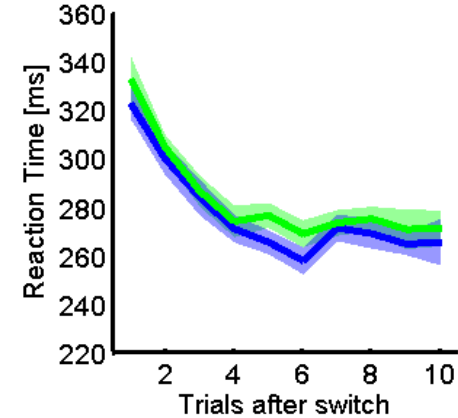
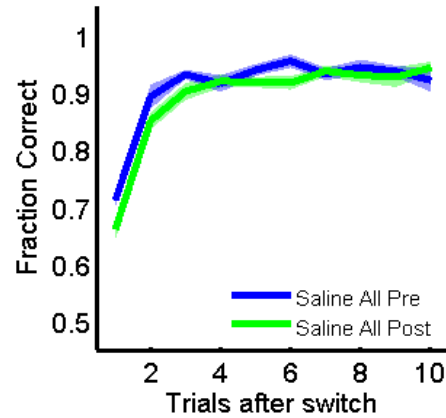
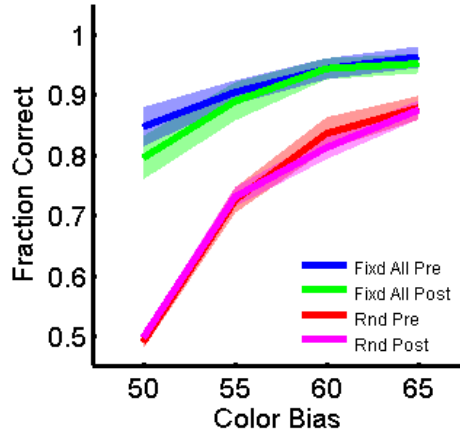
Effects of injection of D2R antagonist : Disruption of ability to use reinforcement learning



Fixed condition
(knowledge of sequence)
Pre-Fixed ——— (blue line)
Post-Fixed ——— (green line)

Random condition
(pay attention to stimulus)
Pre-Rand ——— (red line)
Post-Rand ——— (magenta line)

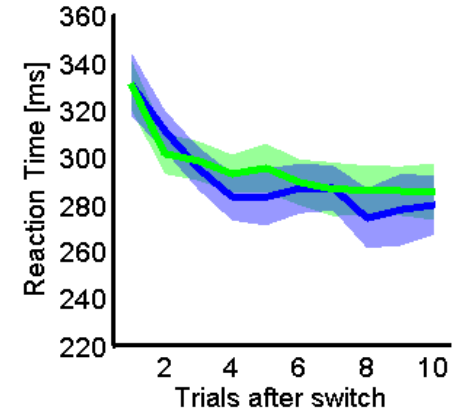
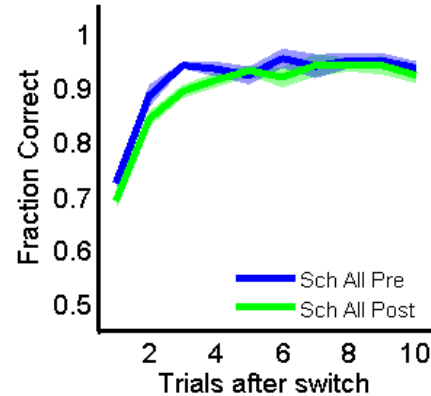
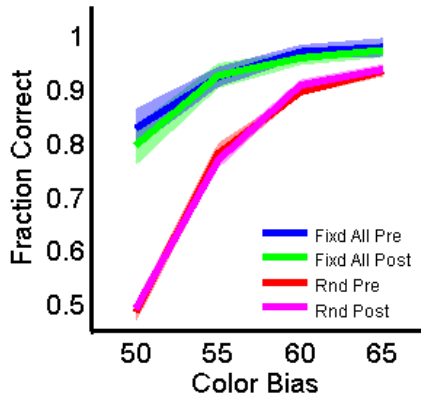
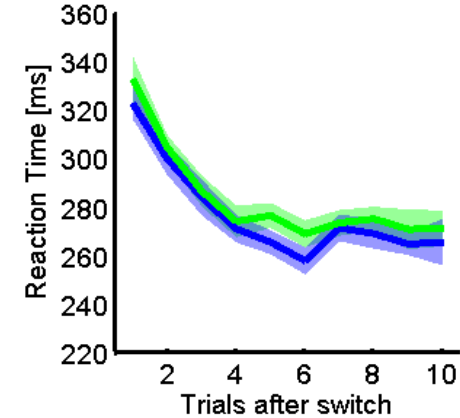
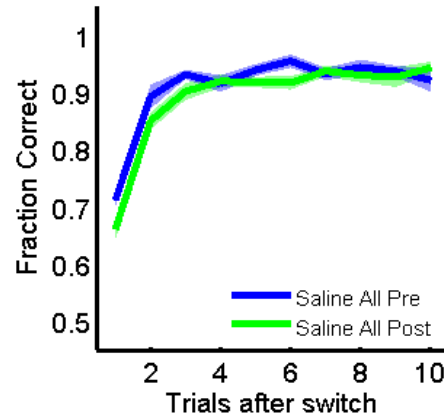
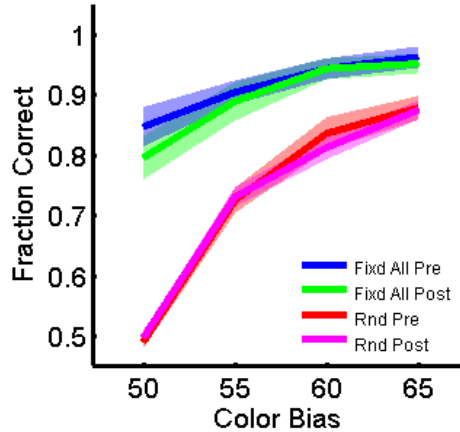
Effects of injection of saline and D1R antagonist : No behavioral effects



Fixed condition
(knowledge of sequence)
Pre-Fixed —
Post-Fixed —

Random condition
(pay attention to stimulus)
Pre-Rand —
Post-Rand —

Effects of injection of saline and D1R antagonist : No behavioral effects

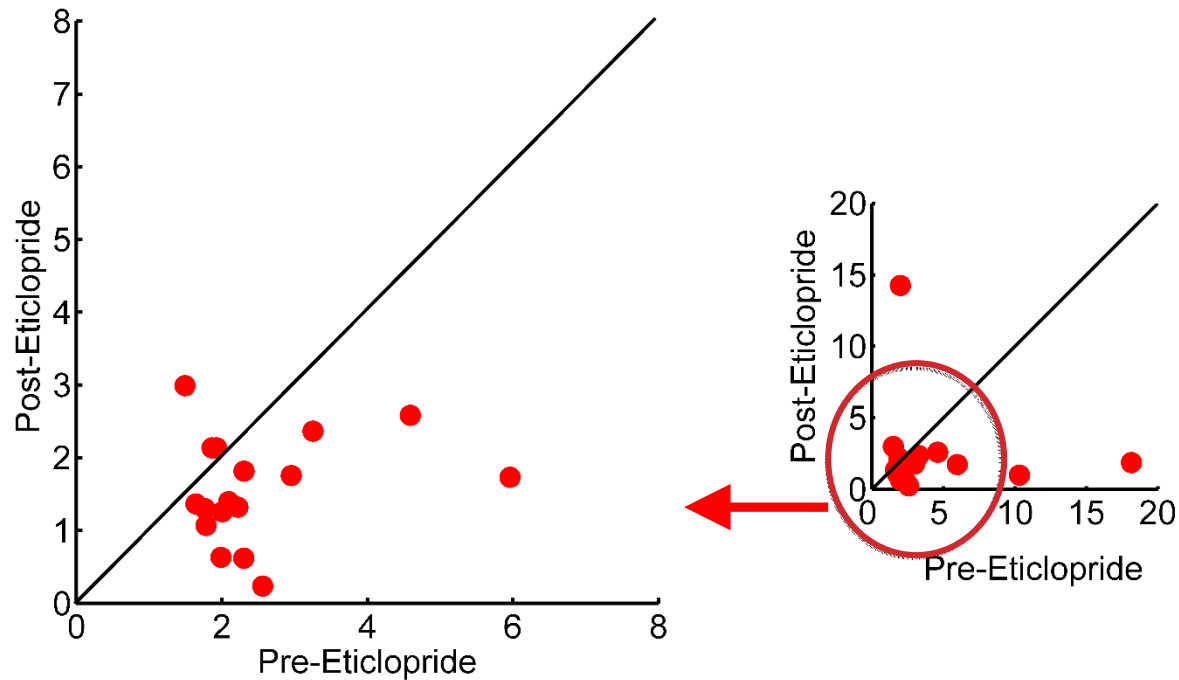


Fixed condition
(knowledge of sequence)
Pre-Fixed —
Post-Fixed —

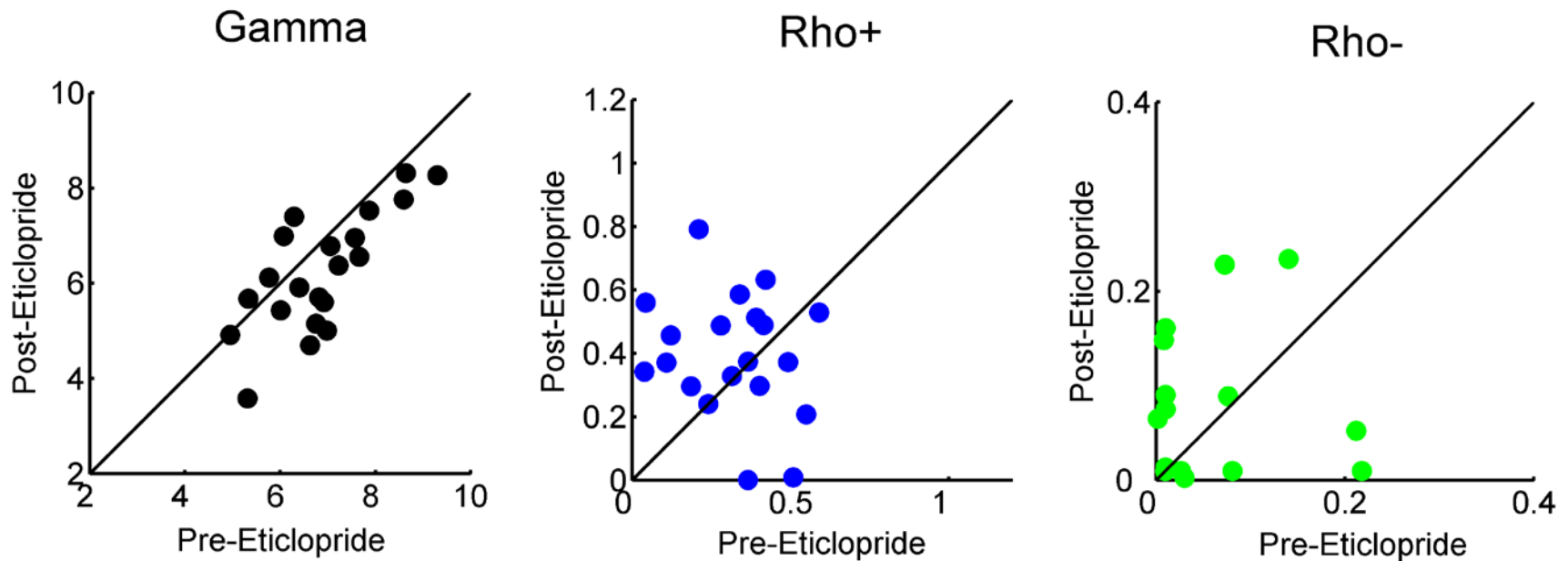
Random condition
(pay attention to stimulus)
Pre-Rand —
Post-Rand —

Dopamine type 2 receptors control inverse temperature (Choice Consistency).

Beta (Choice Consistency)



No changes in the learning rates and weights of mixture of two colors.



Conclusions

-

-

-

Conclusions

- D2R antagonist injection impacts performance accuracy and reaction times only when animal makes a decision based on past reinforcement (Fixed condition).

▪

▪

Conclusions

- D2R antagonist injection impacts performance accuracy and reaction times only when animal makes a decision based on past reinforcement (Fixed condition).
- There were no behavioral effects with D1R antagonist or saline administration, nor in the random condition (forced to use information in fixation stimulus) with D2R antagonists.

▪

Conclusions

- D2R antagonist injection impacts performance accuracy and reaction times only when animal makes a decision based on past reinforcement (Fixed condition).
- There were no behavioral effects with D1R antagonist or saline administration, nor in the random condition (forced to use information in fixation stimulus) with D2R antagonists.
- In the reinforcement learning model, effects of the D2R antagonist were on inverse temperature, i.e. response variability increased (choice consistency decreased).

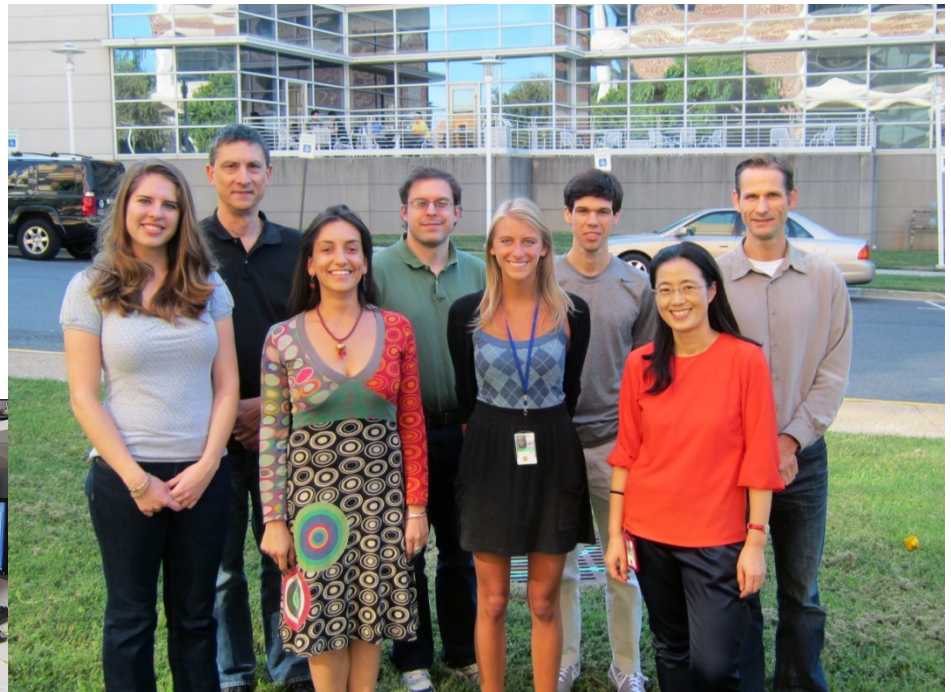
Acknowledgement

Bruno Averbeck

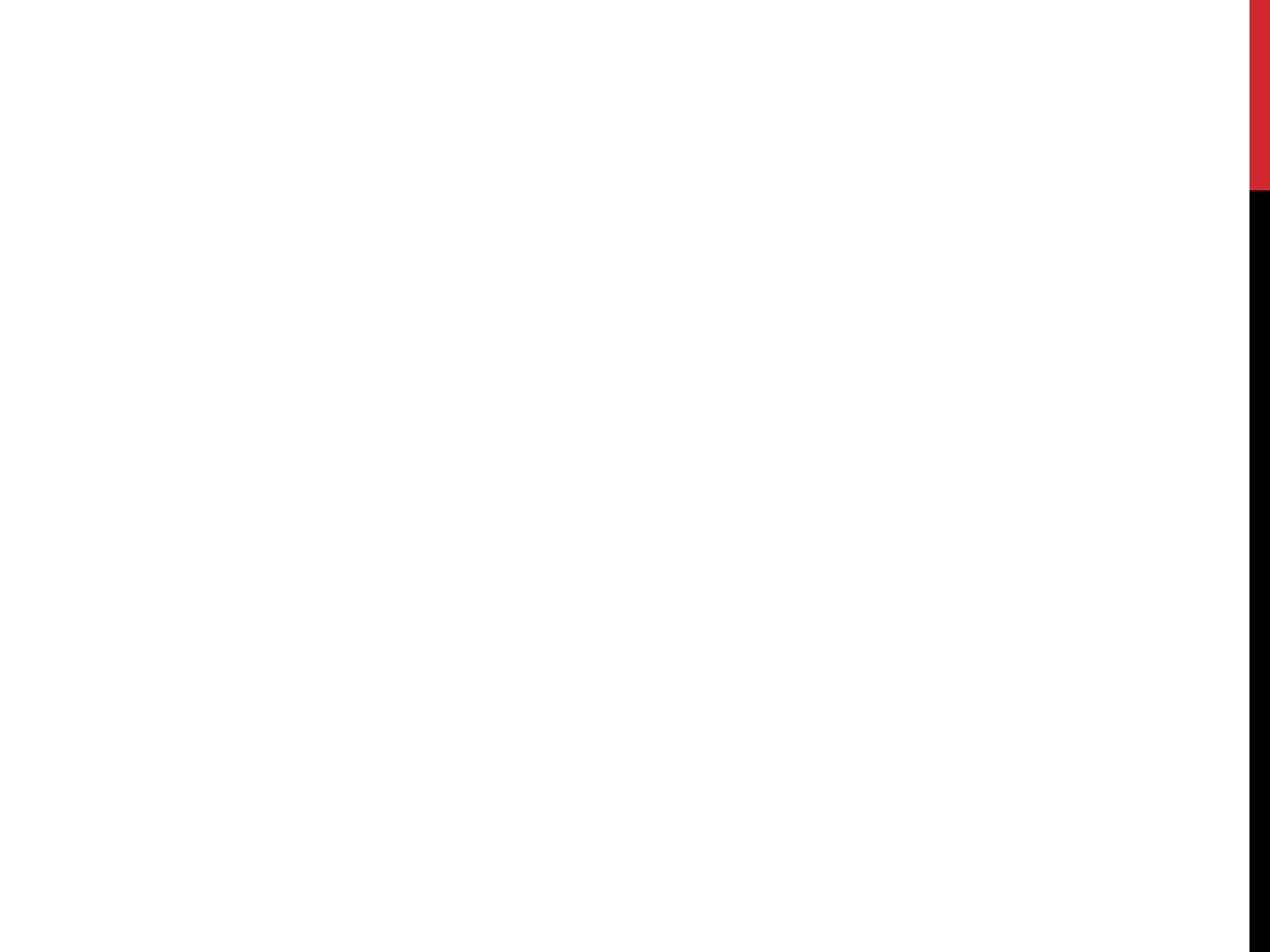
Moonsang Seo

Olga Dal Monte

Andy Mitz



Unit on Learning and Decision Making@LN/NIMH/NIH



❑ **Background**

- Behavior task : Oculomotor Sequential decision making Task
- Brain areas of interest : Prefrontal cortex and Striatum

❑ **Neurophysiology Results**

- Striatal neurons represent task relevant variables

❑ **Neuropharmacology Results**

- How is dopamine involved in the decision making process?

❑ **Background**

- Behavior task : Oculomotor Sequential decision making Task
- Brain areas of interest : Prefrontal cortex and Striatum

❑ **Neurophysiology Results**

- Striatal neurons represent task relevant variables

❑ **Neuropharmacology Results**

- How is dopamine involved in the decision making process?

Interim Summary

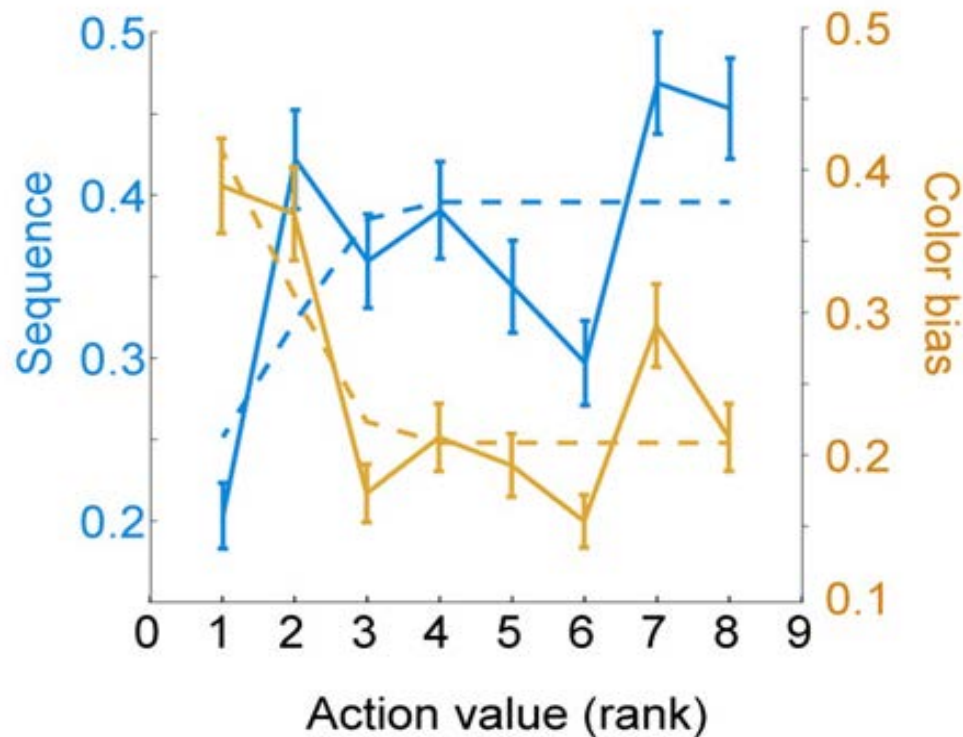
- The dorsal striatum represents the value of actions driven by immediately available perceptual information (mixture of two colors) or the reinforcement of previous choices (Knowledge of sequences).

-

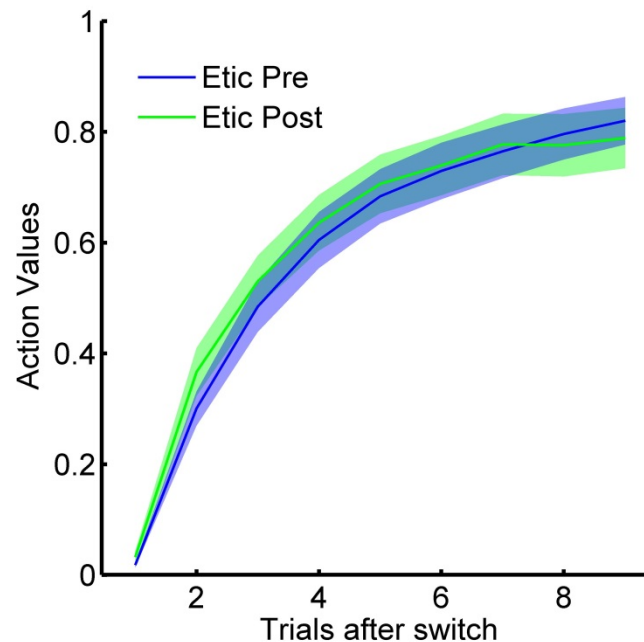
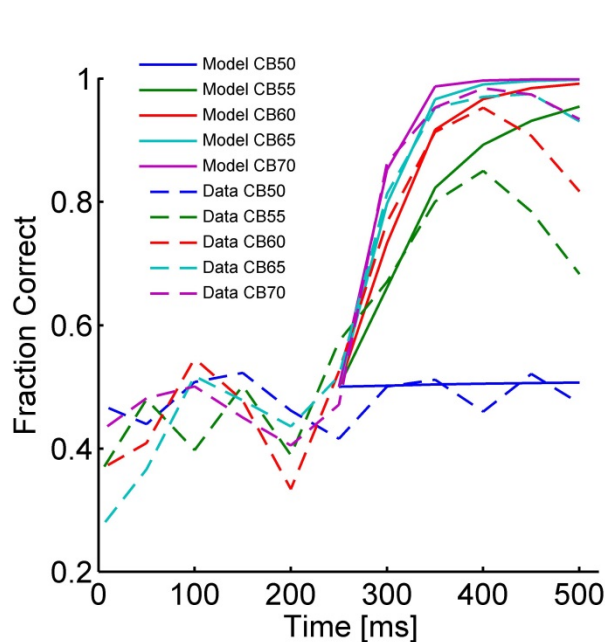
Interim Summary

- The dorsal striatum represents the value of actions driven by immediately available perceptual information (mixture of two colors) or the reinforcement of previous choices (Knowledge of sequences).
- The dorsal striatum integrates knowledge of sequences and color bias information.

Integration of sequence and color bias information in the dorsal Striatum

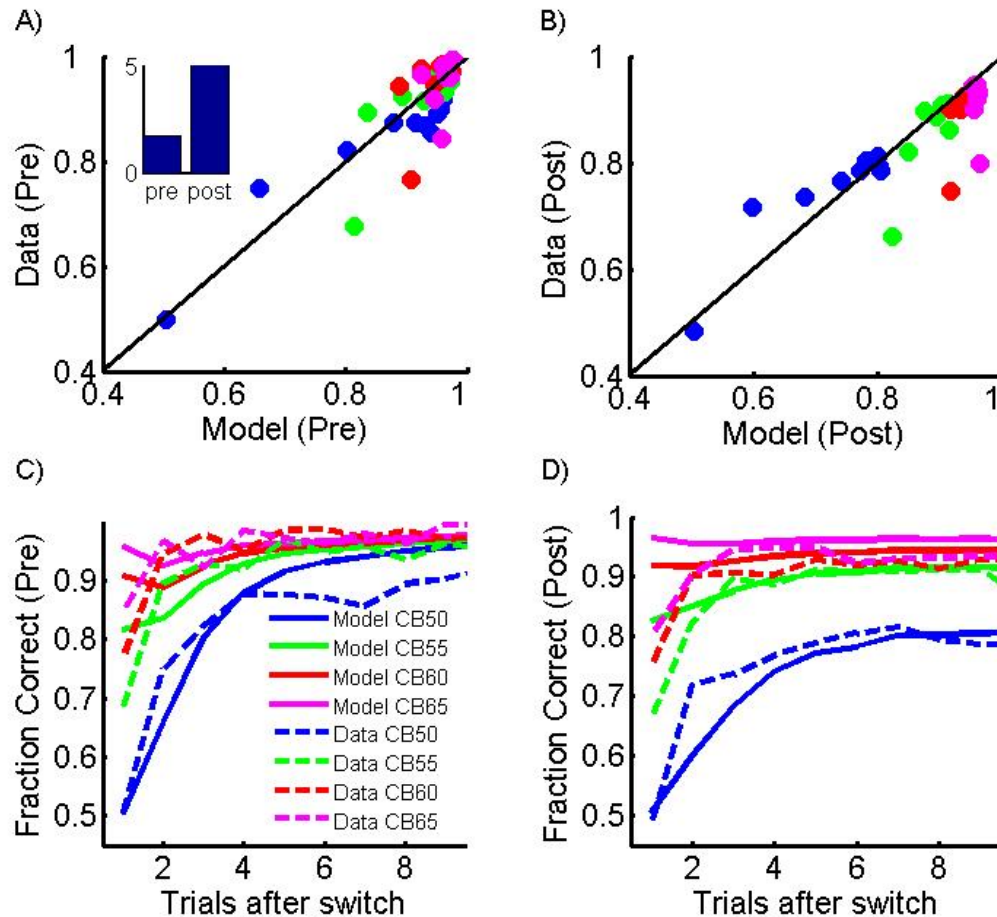


Model fits for perceptual inference and reinforcement learning.



$$P(\text{Choice} | V(t), N_{\text{blue}}(t)) = \frac{P(V(t) | \text{Choice}) P(N_{\text{blue}}(t) | \text{Choice}) P(\text{Choice})}{P(V(t), N_{\text{blue}}(t))}$$

D2A into the dStr increased noise in the value representation that was integrated by choice mechanisms.



❑ Background

- Behavior task : Ocularmotor Sequential decision making Task
- Brain area of interest : Prefrontal cortex and Striatum

❑ Neurophysiology Results

- How striatum neurons represent the decision making process.

❑ Neuropharmacology Results

- How dopamine involved in the decision making process.