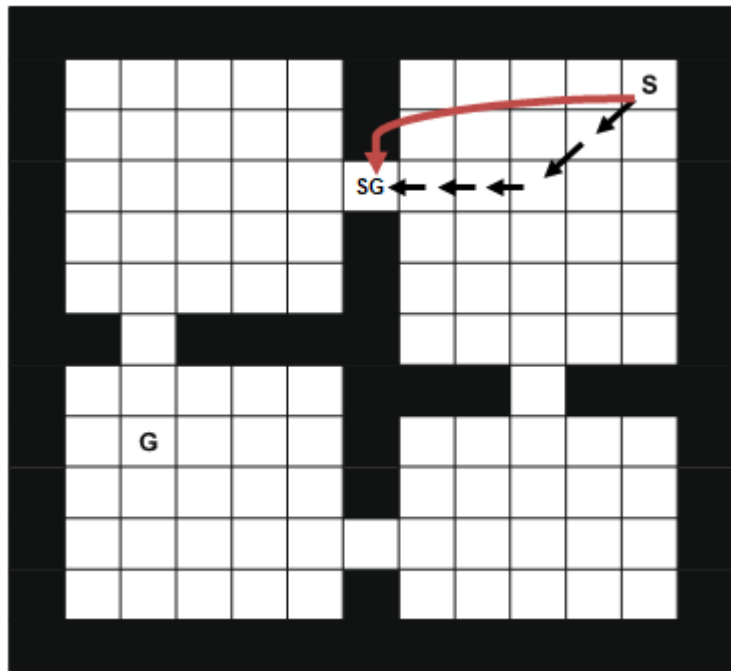# Option Transfer in Humans

Liyu (Jimmy) Xia, Anne Collins

University of California, Berkeley
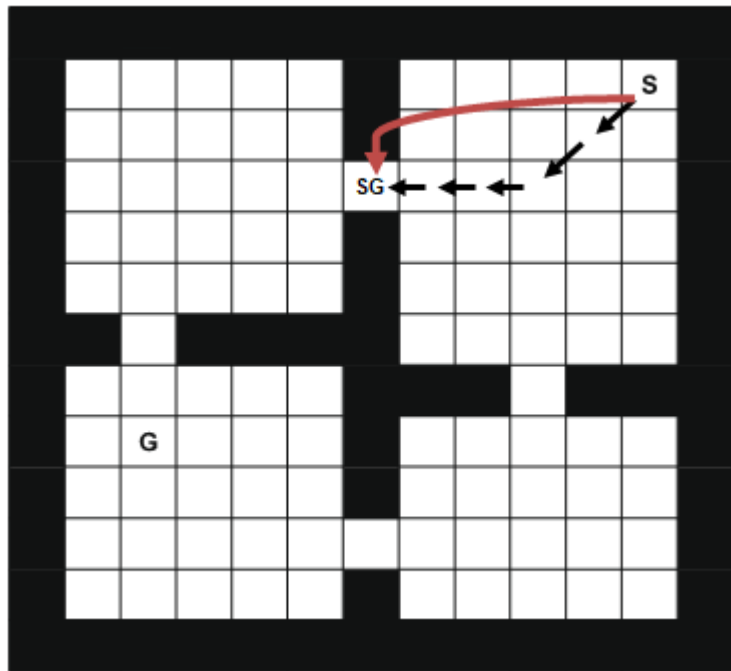
Hierarchical human behavior

The Options Framework

The Options Framework
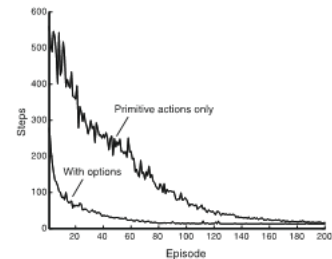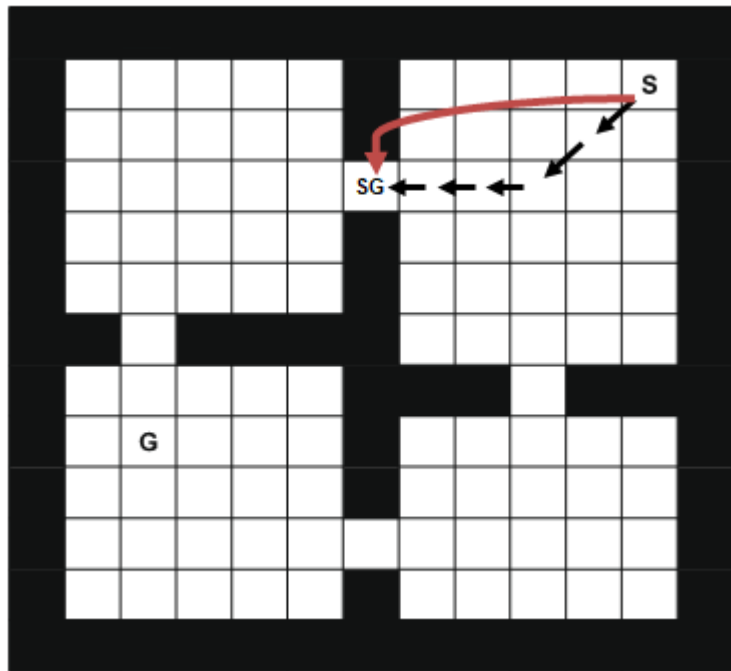

(Precup & Barto; Botvinick, Niv & Barto)

The Options Framework

1. Speed up learning

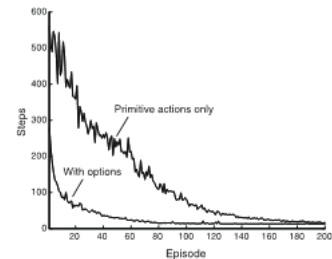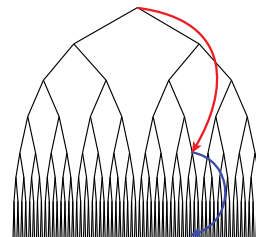# The Options Framework

1. Speed up learning

2. Long-term planning

(Precup & Barto; Botvinick, Niv & Barto)
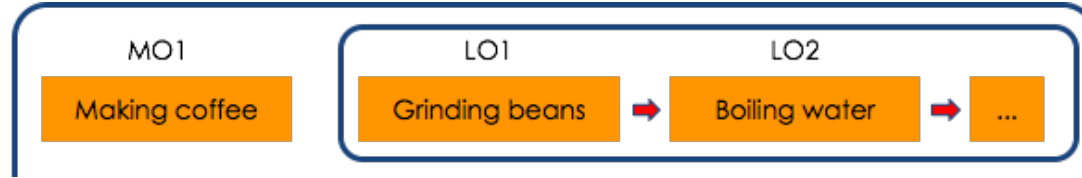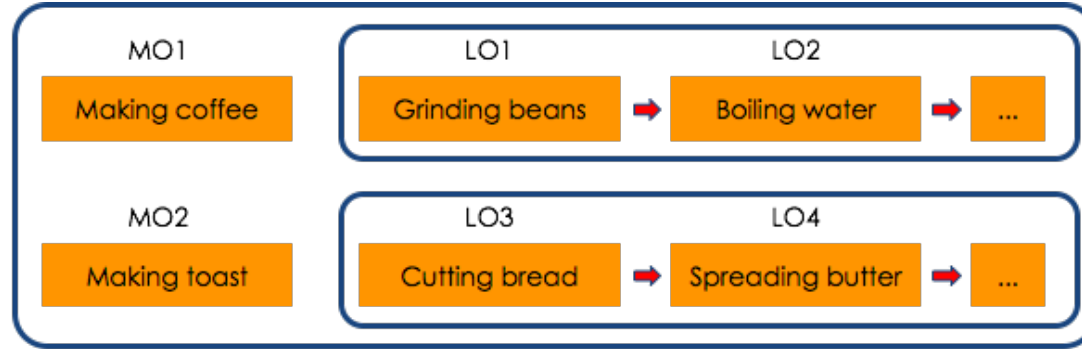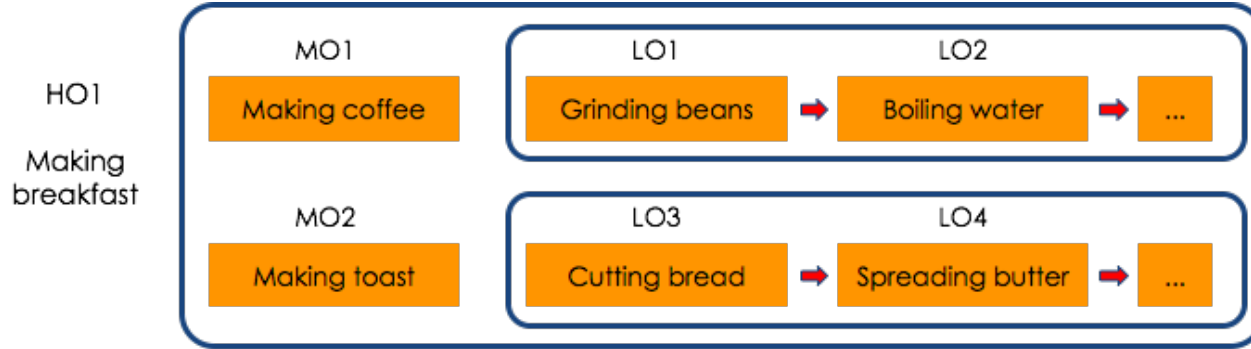
What about Human Learning?

What about Human Learning?

Hierarchically structured
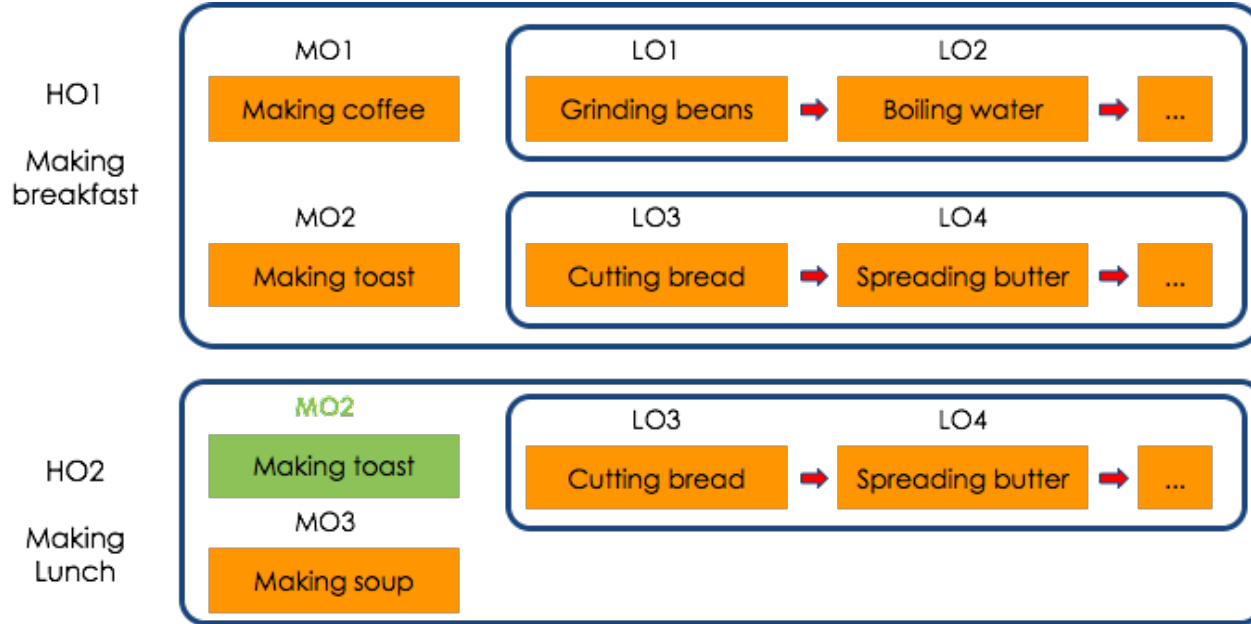
What about Human Learning?

Hierarchically structured

What about Human Learning?

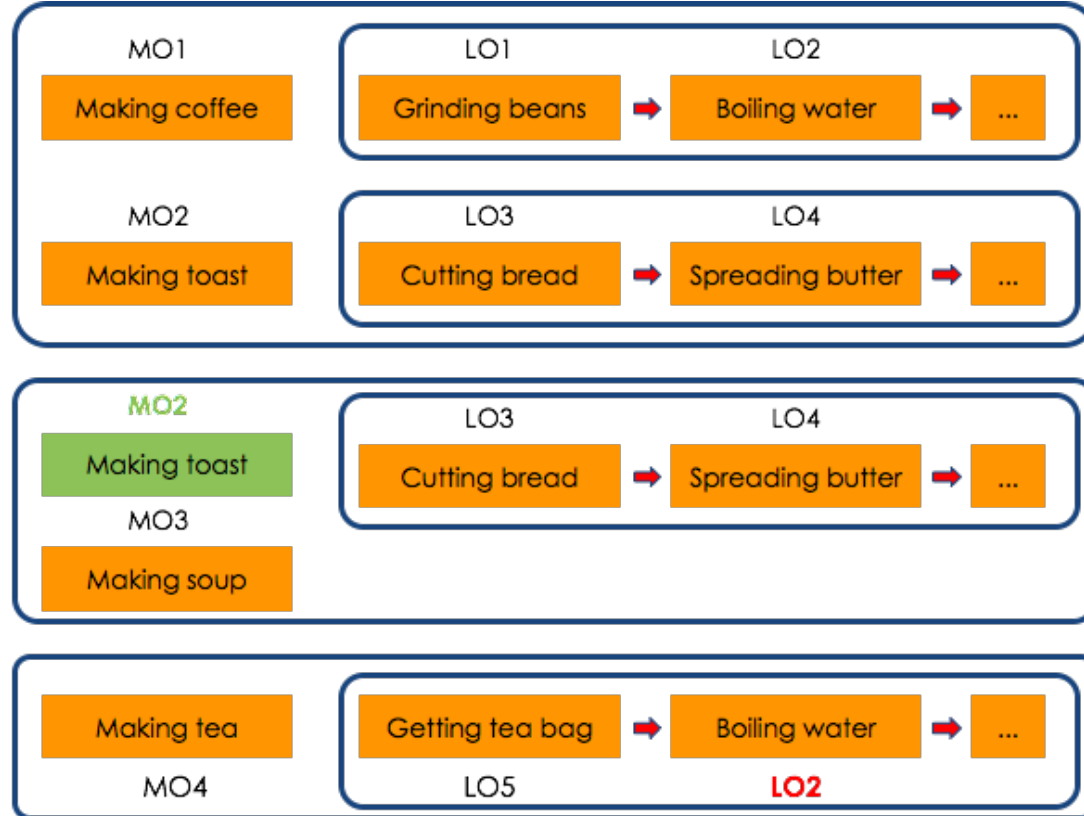Hierarchically structured

What about Human Learning?

Hierarchically structured

Flexible transfer

What about Human Learning?

Hierarchically structured

Flexible transfer

Options in Human Learning?

1. Do we learn options (through RL)? Options of options?

Options in Human Learning?

1. Do we learn options (through RL)? Options of options?
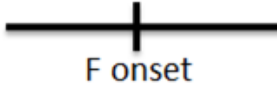

2. Can we transfer options? At any level?

Options in Human Learning?

1. Do we learn options (through RL)? Options of options?

2. Can we transfer options? At any level?

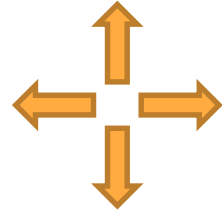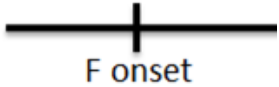3. Does option improve exploration and speed up human learning?

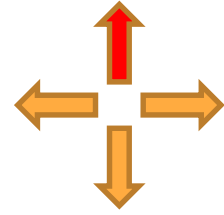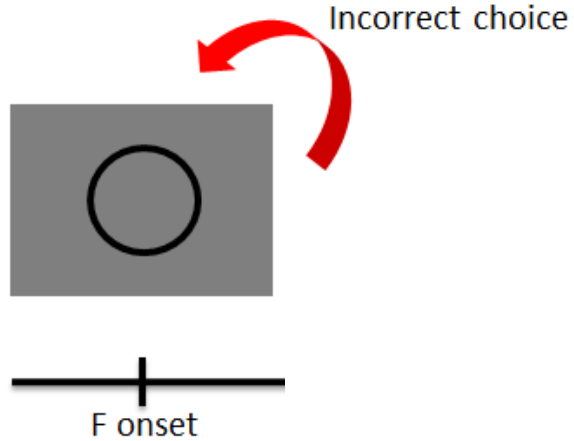# Experimental Design: Trial Structure

Experimental Design: Trial Structure

F onset

Experimental Design: Trial Structure

Experimental Design: Trial Structure

Experimental Design: Trial Structure
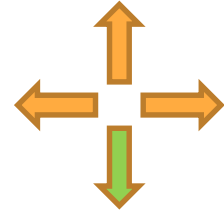
Experimental Design: Trial Structure

Experimental Design: Trial Structure

Experimental Design: Trial Structure

Experimental Design: Trial Structure
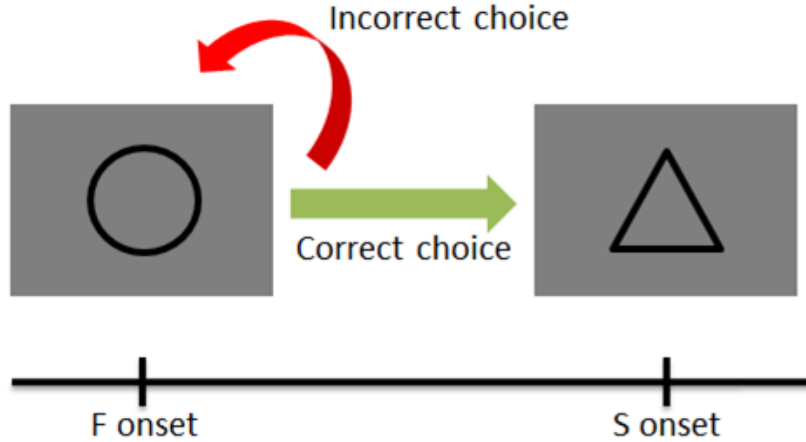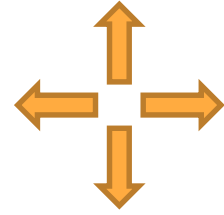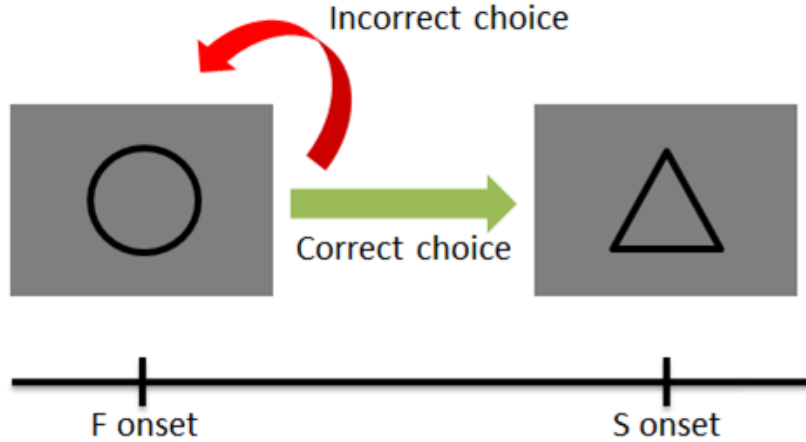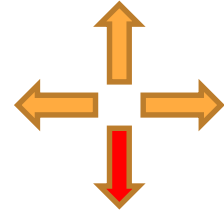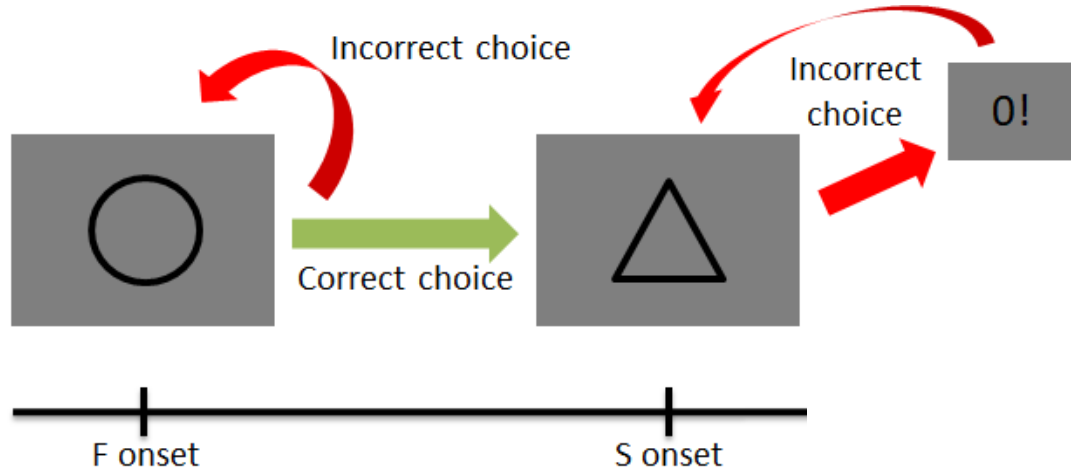

2nd stage shape is random:
Rule out pure sequence learning

Experimental Design: Trial Structure

2nd stage shape is random:
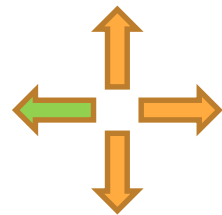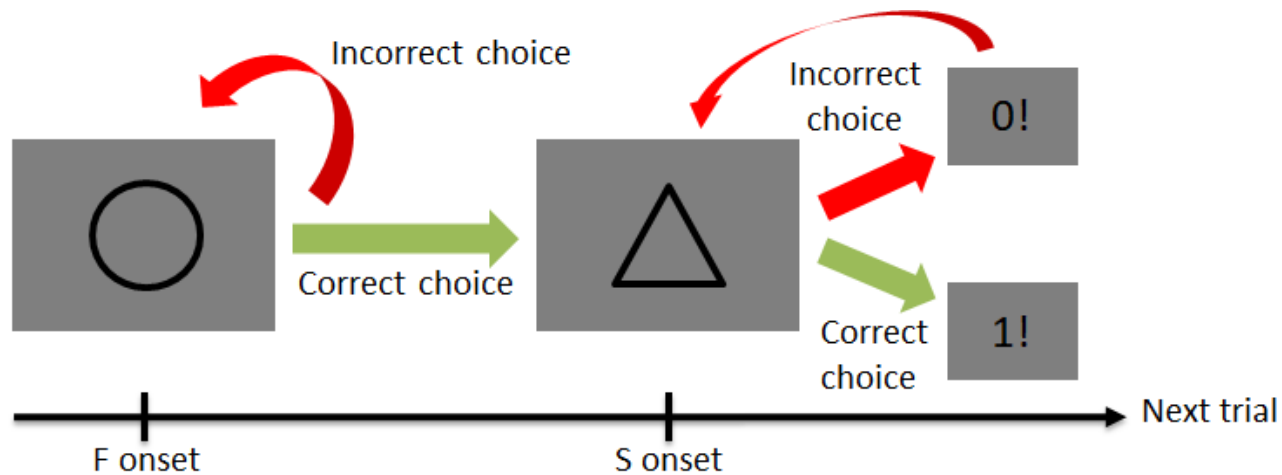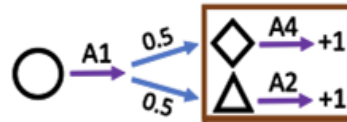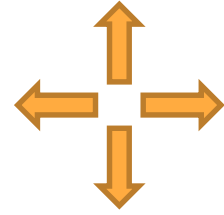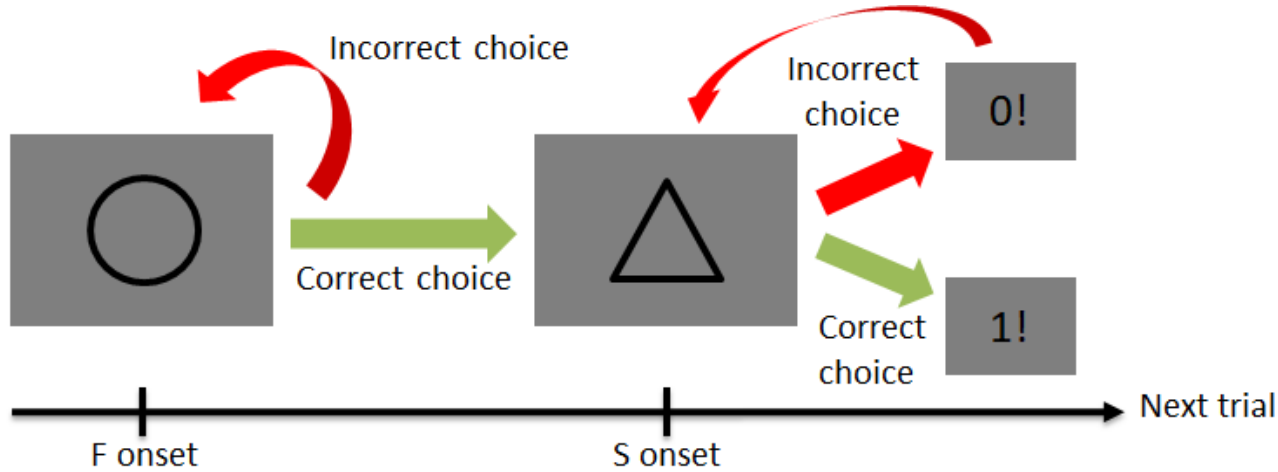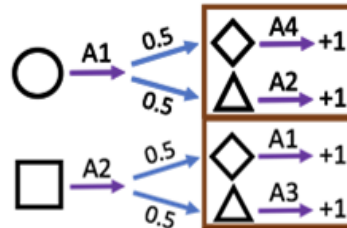Rule out pure sequence learning

2nd stage depends on 1st stage:
Option-specific policies

Experimental Design: Trial Structure

Incorrect choice

Correct choice

Incorrect choice

Correct choice

0!

1!

F onset

S onset

Next trial

MO1

Making coffee

MO2

Making toast

2nd stage shape is random:
Rule out pure sequence learning

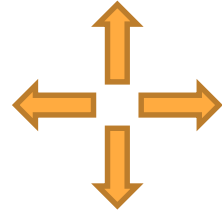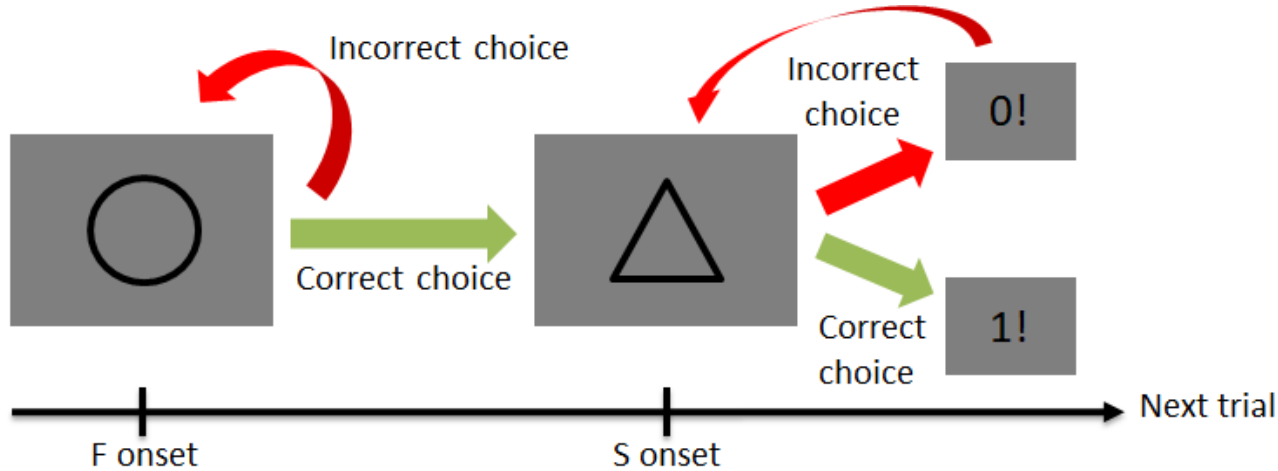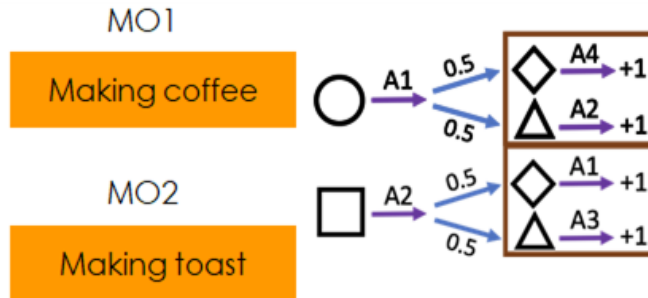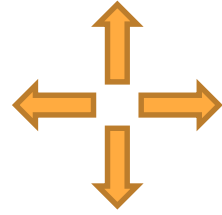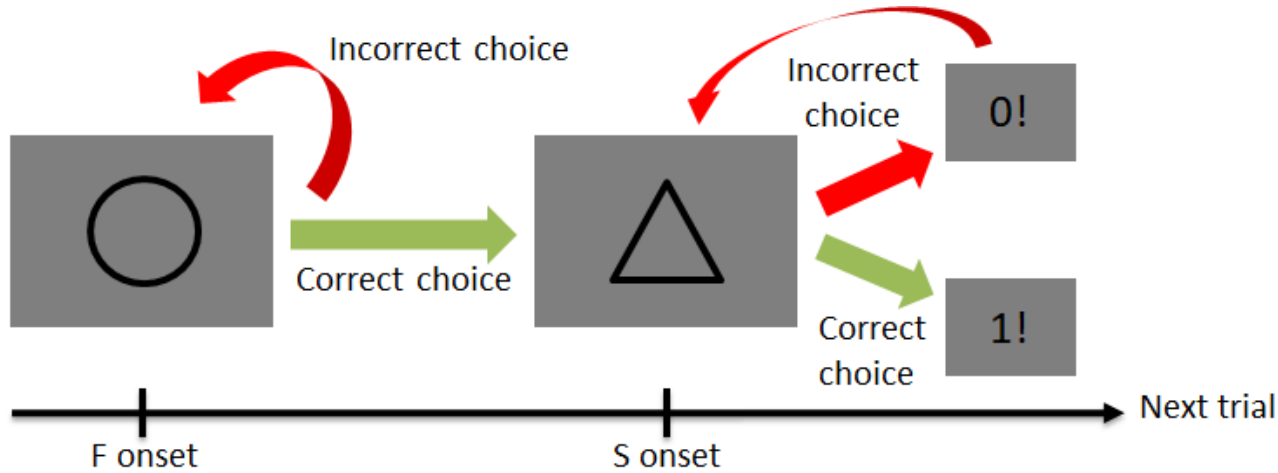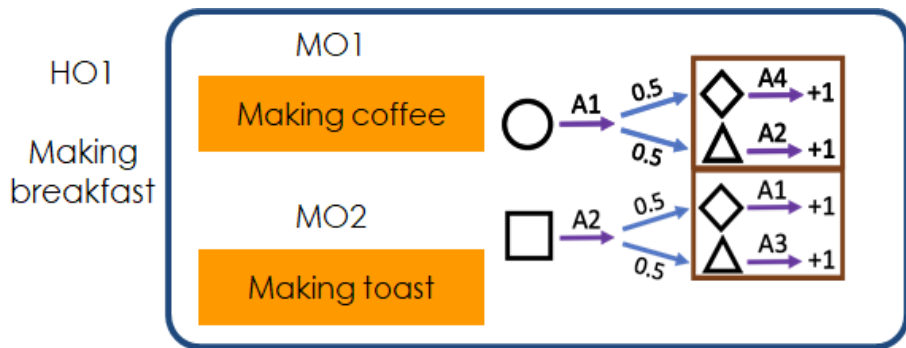2nd stage depends on 1st stage:
Option-specific policies
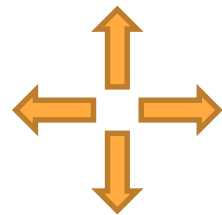
Experimental Design: Trial Structure

2nd stage shape is random:
Rule out pure sequence learning

2nd stage depends on 1st stage:
Option-specific policies

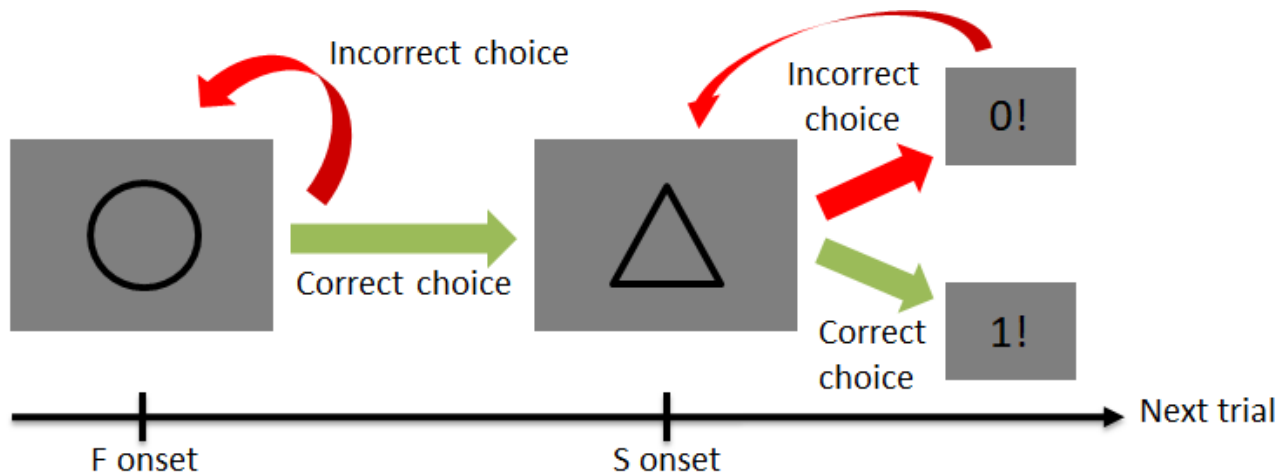# Experimental Design: Block Structure
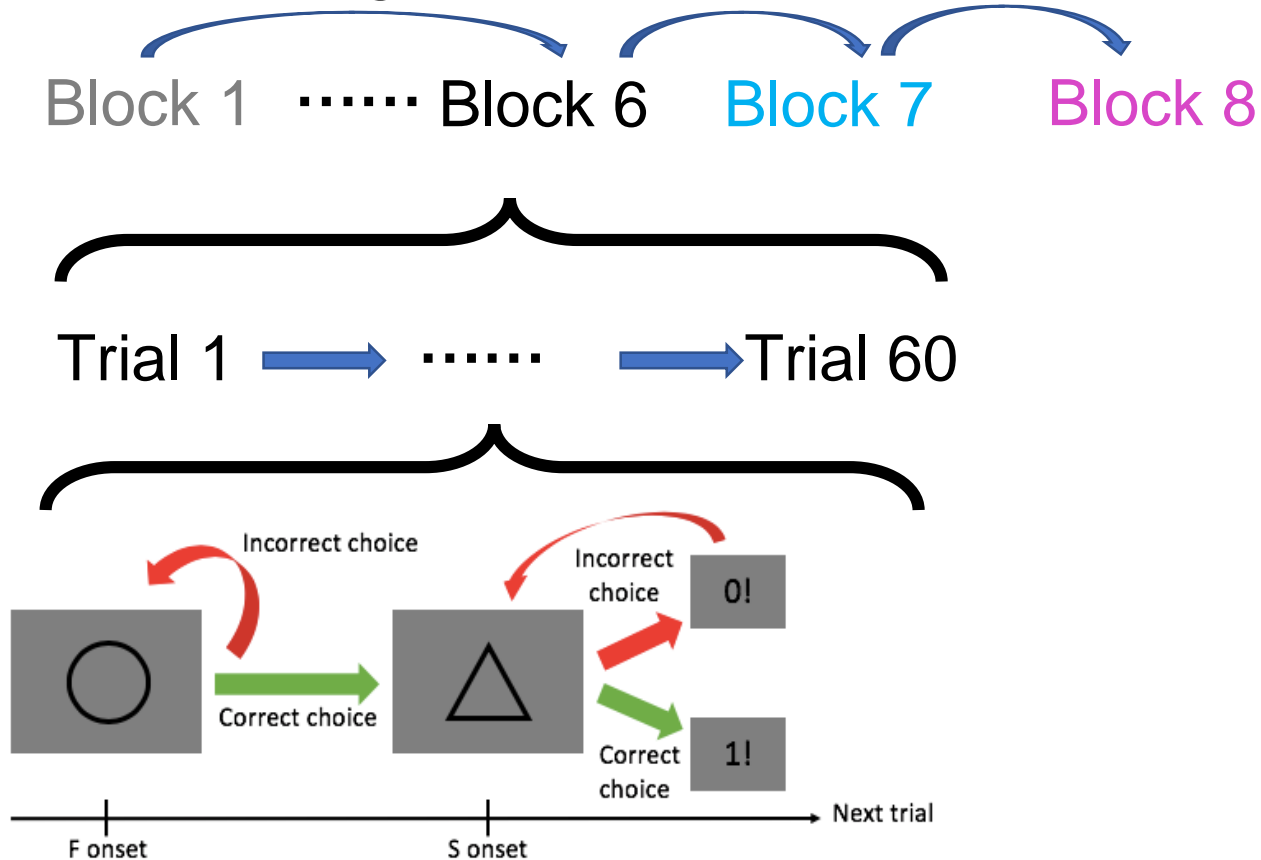
Block 1 ⋯⋯ Block 6   Block 7   Block 8

Trial 1 ⟶ ⋯⋯ ⟶ Trial 60

Incorrect choice

Correct choice

Incorrect choice

0!

Correct choice

1!

Next trial

F onset   S onset

Experimental Design: Action Assignments

Experimental Design: Action Assignments

Blocks 1, 3, **5:**

MO1: ○ —A1→ 0.5 / 0.5 → ◇ —A4→ +1 / △ —A2→ +1   **LO1**
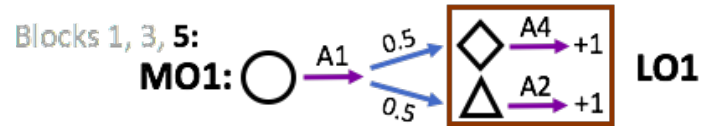
Experimental Design: Action Assignments

Experimental Design: Action Assignments

Experimental Design: Action Assignments

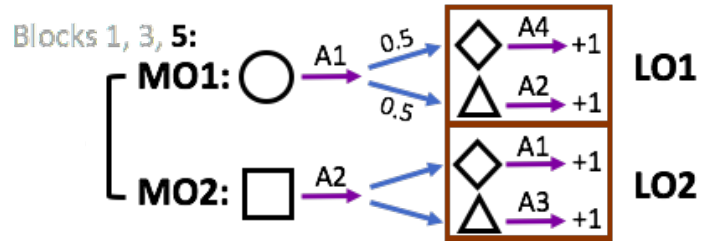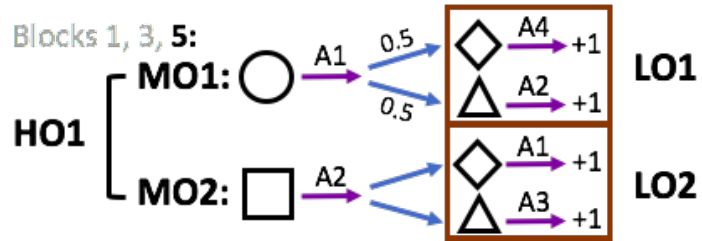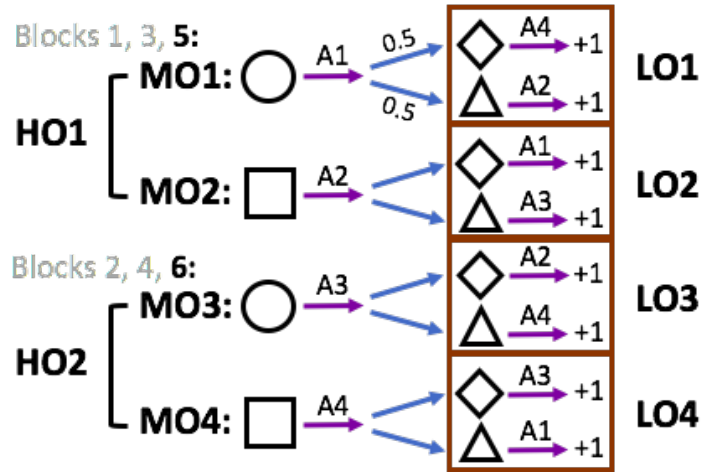Unsignaled context: test
- **creation** of two sets of high level options
- **reactivation** of high-level options in Blocks 3-6

Experimental Design: Action Assignments

Unsignaled context: test
- **creation** of two sets of high level options
- **reactivation** of high-level options in Blocks 3-6

Test
- **positive transfer** of low-level options
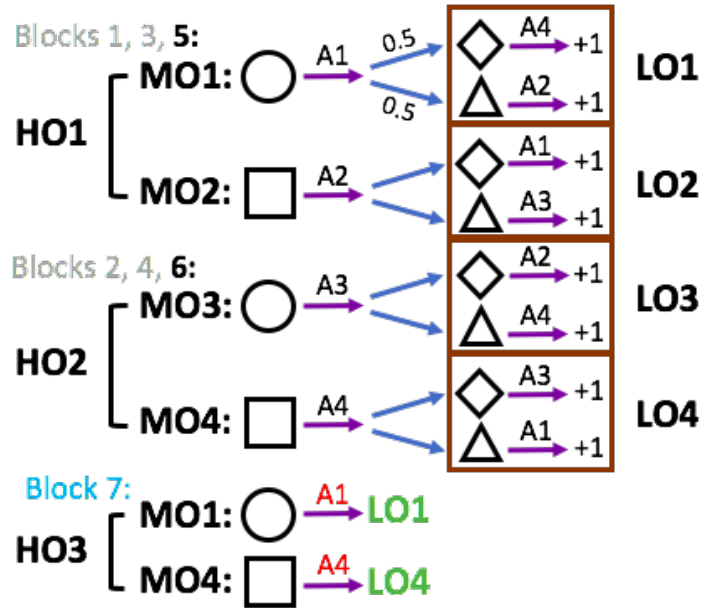- **negative transfer** of high-level options
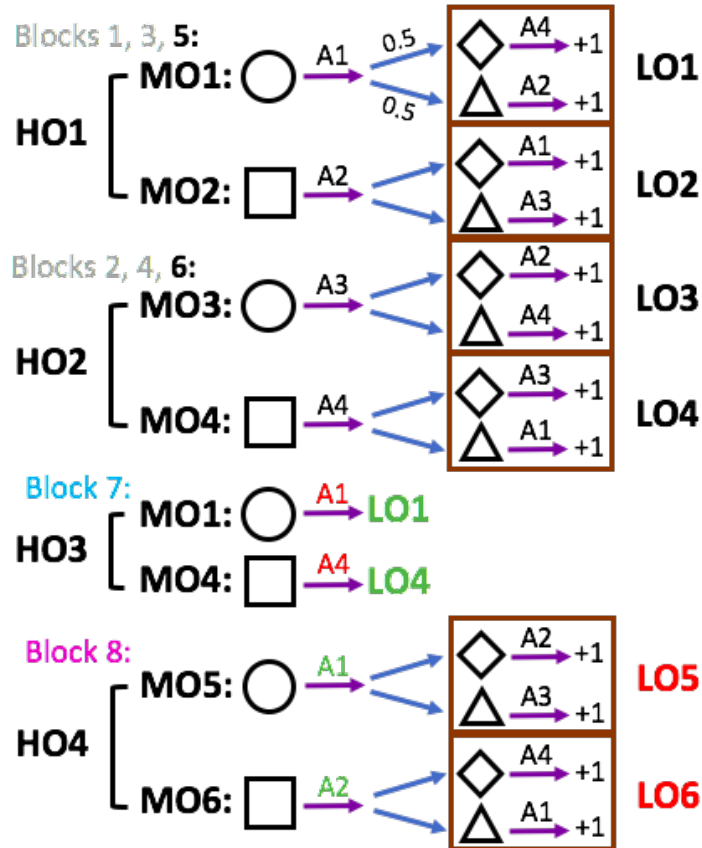
Experimental Design: Action Assignments

Unsignaled context: test
- **creation** of two sets of high level options
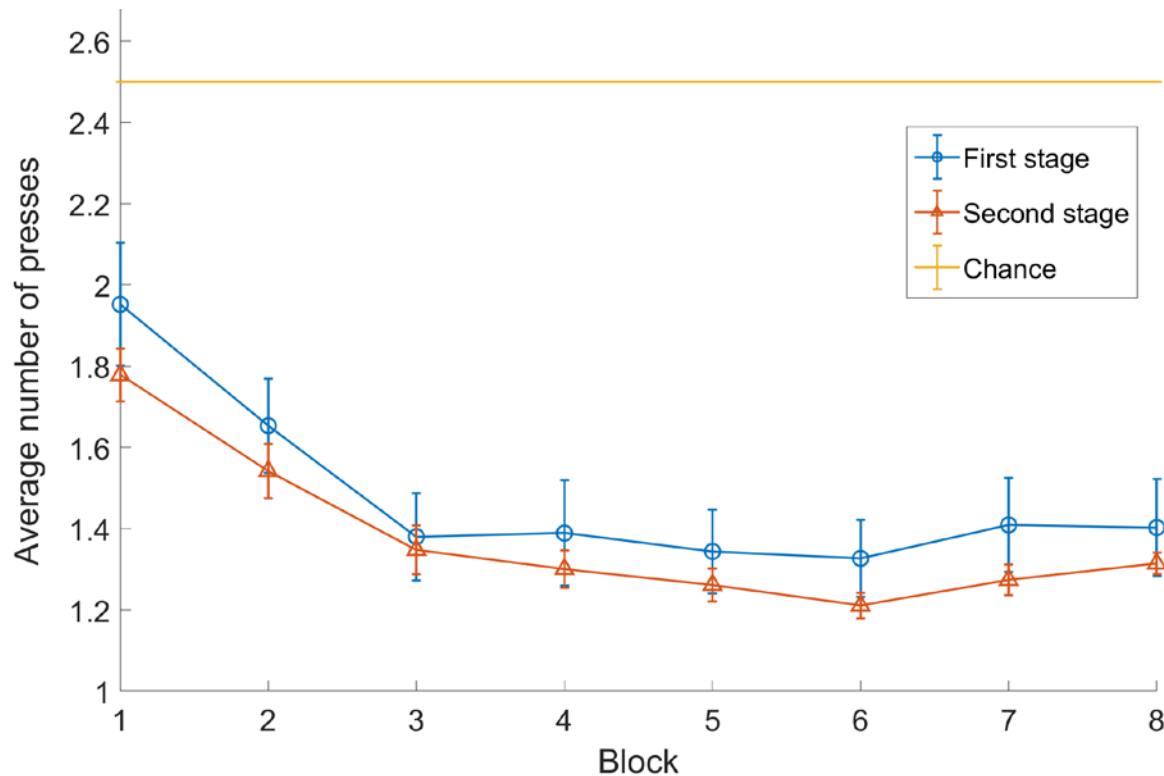- **reactivation** of high-level options in Blocks 3-6

Test
- **positive transfer** of low-level options
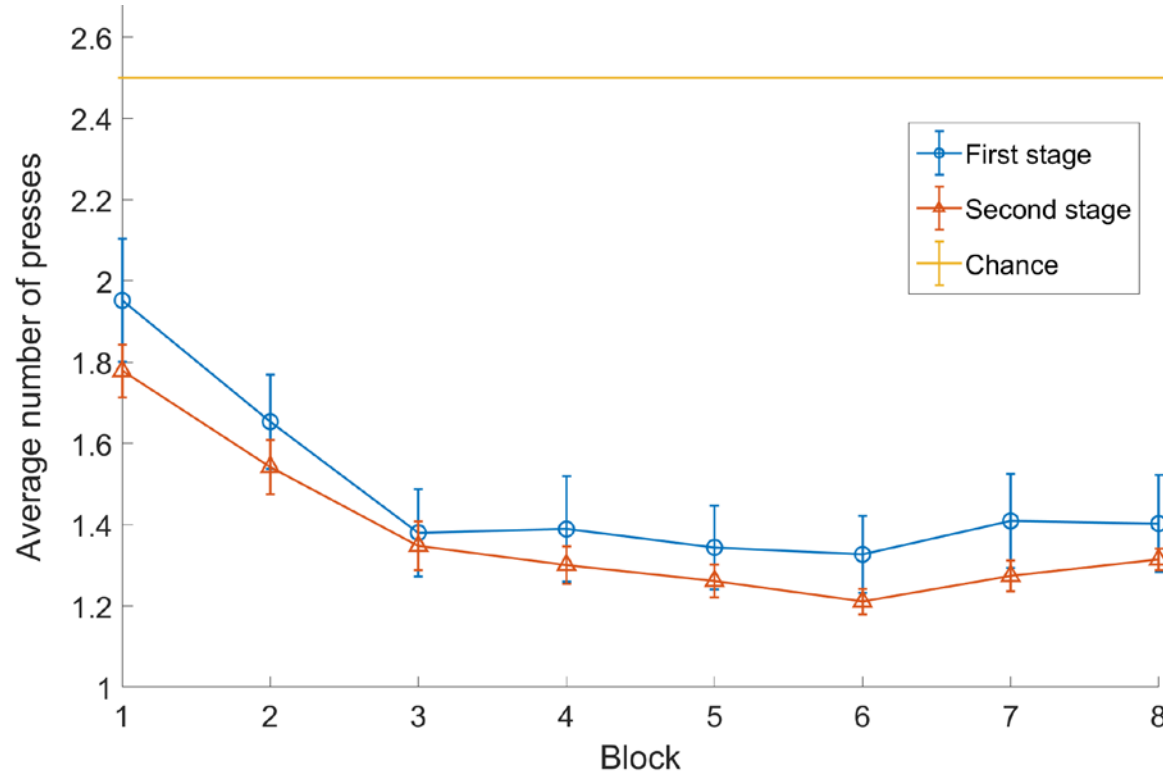- **negative transfer** of high-level options

Test
- **positive transfer** of high-level options
- **negative transfer** of mid-level options

General Behavior: Average Number of Presses
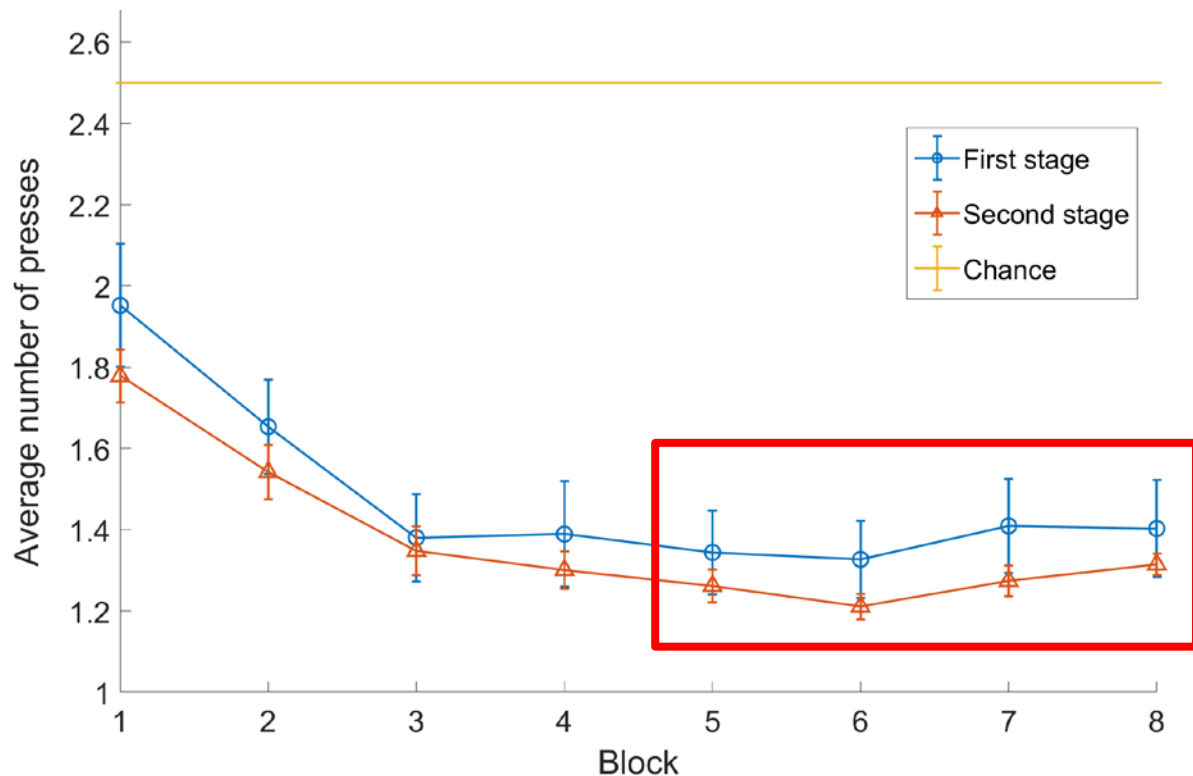
General Behavior: Average Number of Presses

General Behavior: Average Number of Presses

Participants (N = 23) are able to learn the task fairly well.
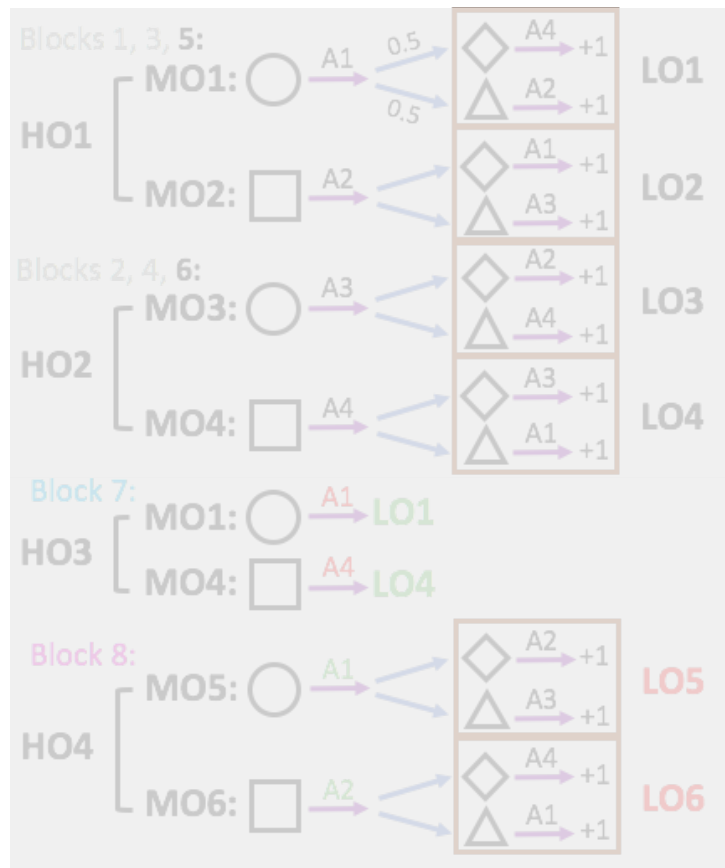
General Behavior: Zoom in on Blocks 5-8

Participants (N = 23) are able to learn the task fairly well.

Blocks 5-6: Learning saturation (benchmark)

Block 7: First testing block

Block 8: Second testing block

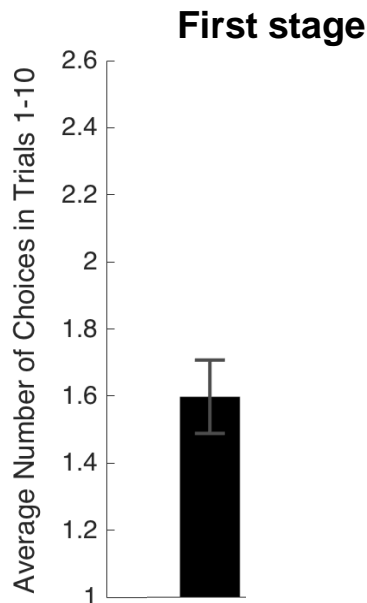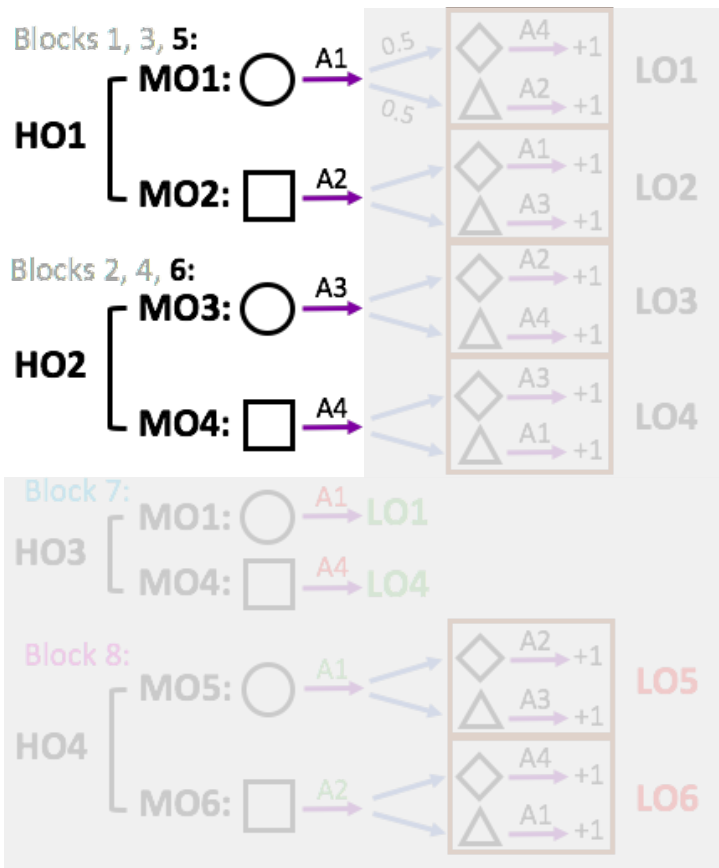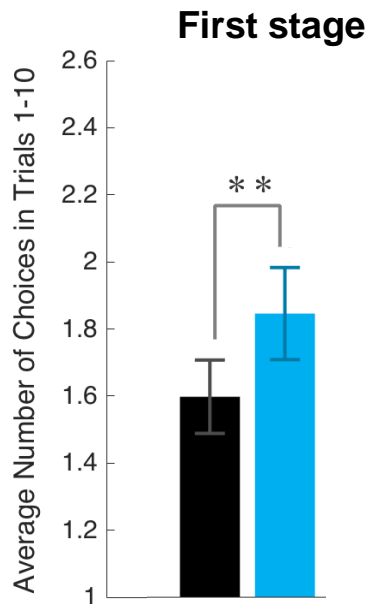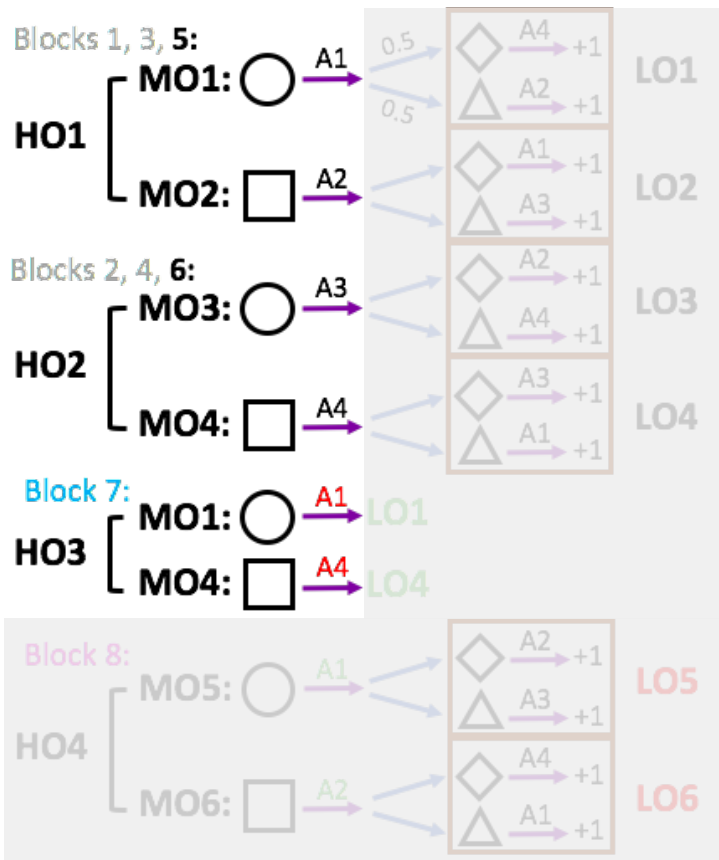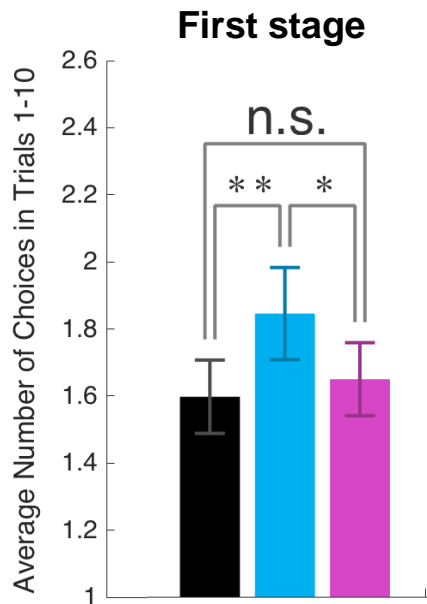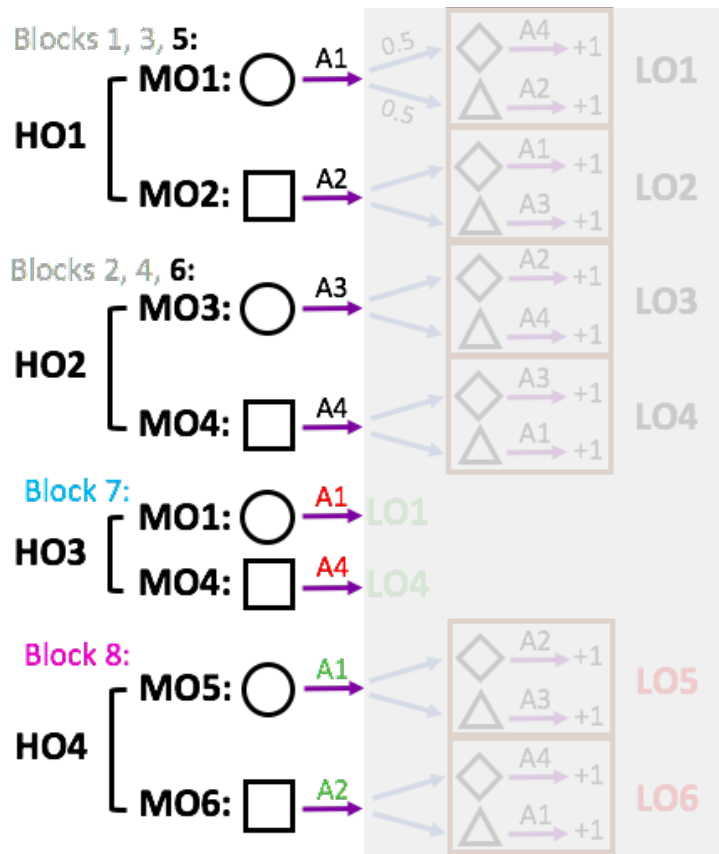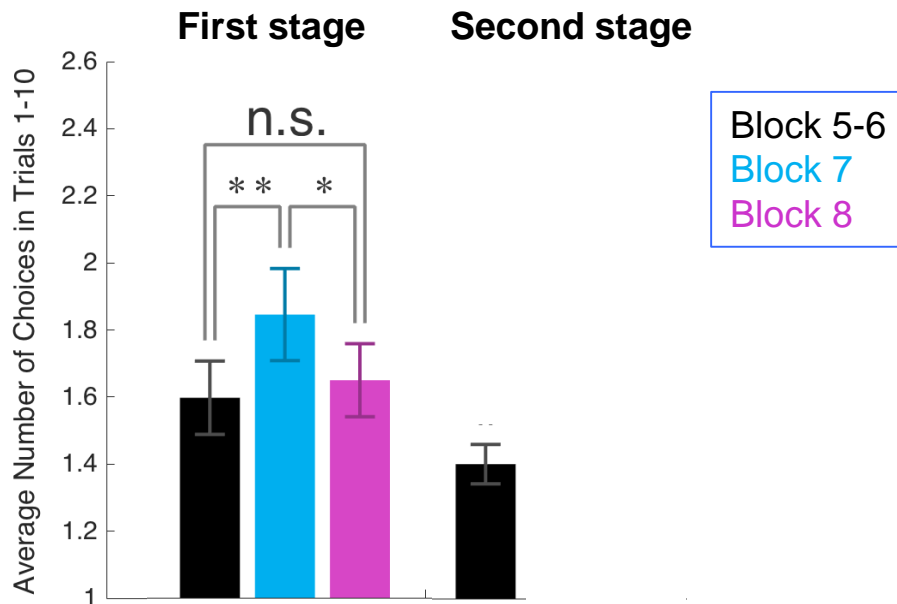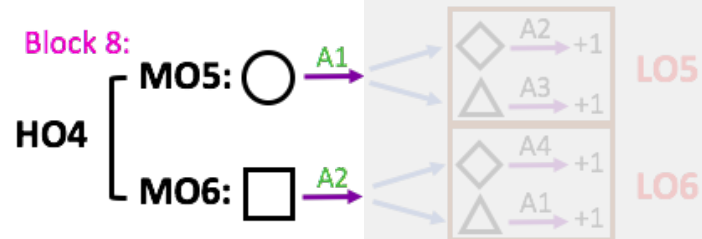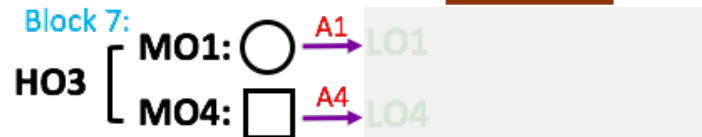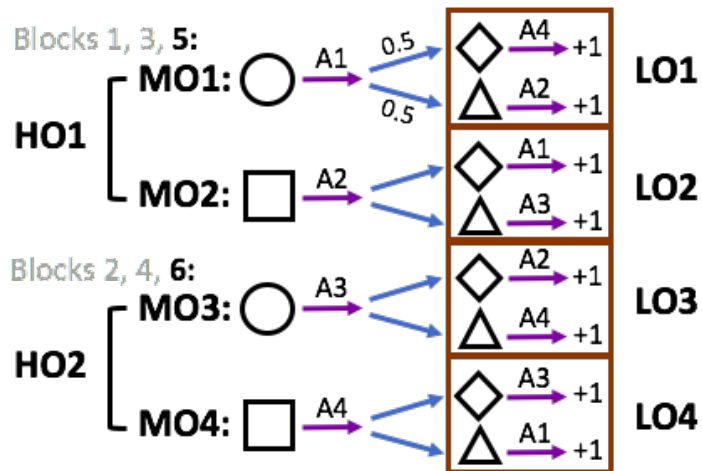# Zoom in Blocks 5-8: First 10 Trials

Block 5-6
Block 7
Block 8

Zoom in Blocks 5-8: First 10 Trials

Blocks 1, 3, **5:**

HO1
- **MO1:** ○ —A1→ 0.5 / 0.5
- **MO2:** □ —A2→

LO1
- ◇ —A4→ +1
- △ —A2→ +1

LO2
- ◇ —A1→ +1
- △ —A3→ +1

Blocks 2, 4, **6:**

HO2
- **MO3:** ○ —A3→
- **MO4:** □ —A4→

LO3
- ◇ —A2→ +1
- △ —A4→ +1

LO4
- ◇ —A3→ +1
- △ —A1→ +1

Block 7:

HO3
- MO1: ○ —A1→ LO1
- MO4: □ —A4→ LO4

Block 8:

HO4
- MO5: ○ —A1→
- MO6: □ —A2→

LO5
- ◇ —A2→ +1
- △ —A3→ +1
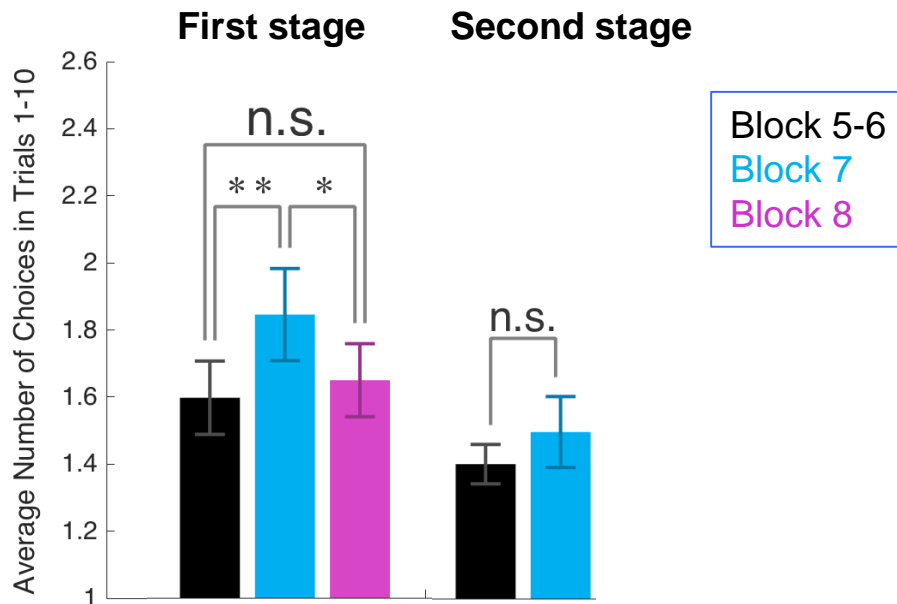
LO6
- ◇ —A4→ +1
- △ —A1→ +1

First stage

Average Number of Choices in Trials 1-10

Block 5-6
Block 7
Block 8

Xia & Collins, in prep

Zoom in Blocks 5-8: First 10 Trials

Block 5-6
Block 7
Block 8

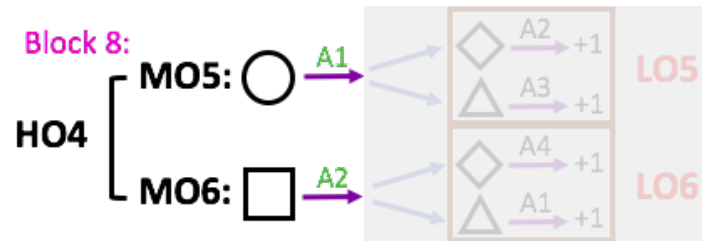Xia & Collins, in prep

Zoom in Blocks 5-8: First 10 Trials

First stage
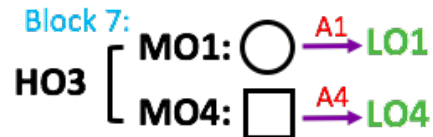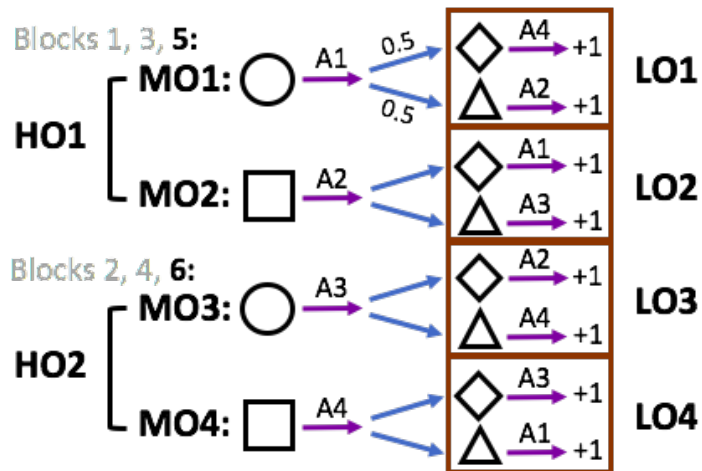
Block 5-6
Block 7
Block 8

43

Zoom in Blocks 5-8: First 10 Trials

Xia & Collins, in prep

Zoom in Blocks 5-8: First 10 Trials

Xia & Collins, in prep

Zoom in Blocks 5-8: First 10 Trials

Xia & Collins, in prep

Zoom in Blocks 5-8: First 10 Trials

Participants learn and
(positive/negative) transfer **options**.

Xia & Collins, in prep
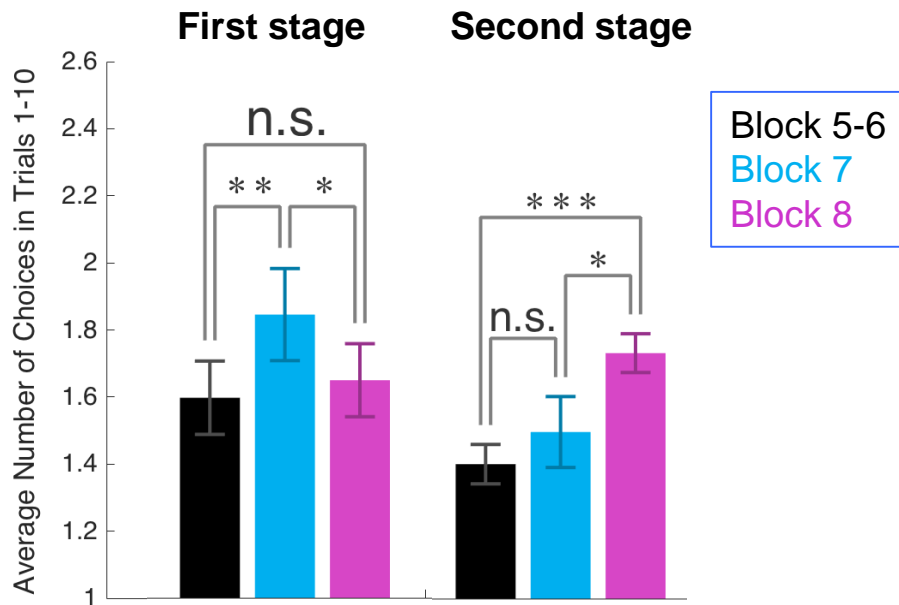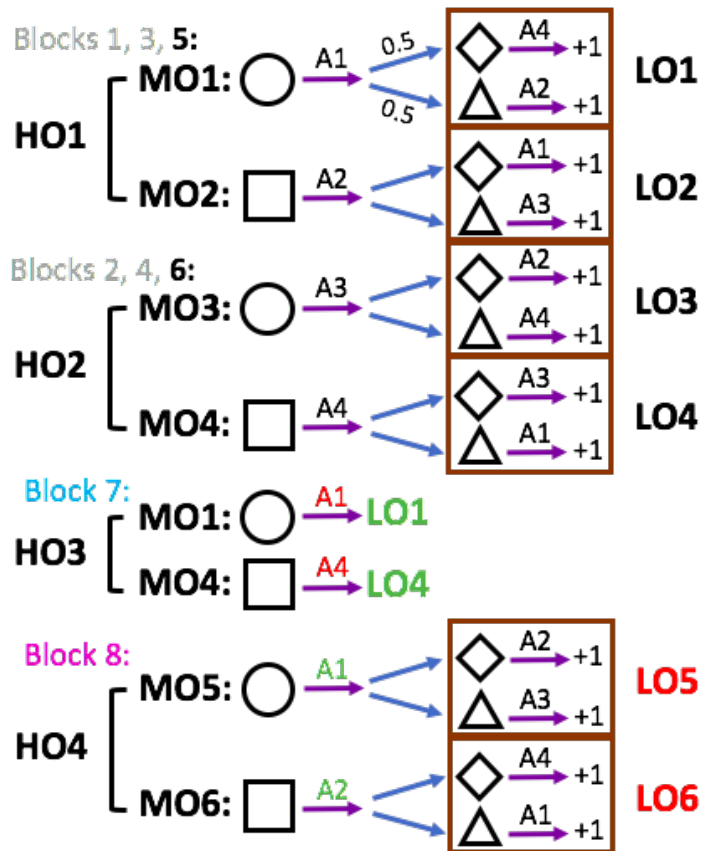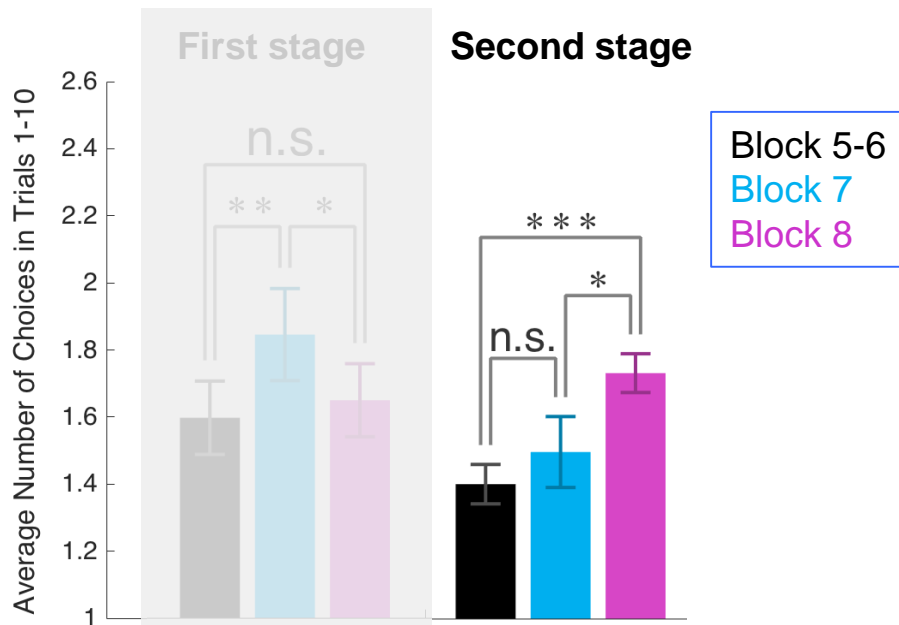
Zoom in Blocks 5-8: First 10 Trials
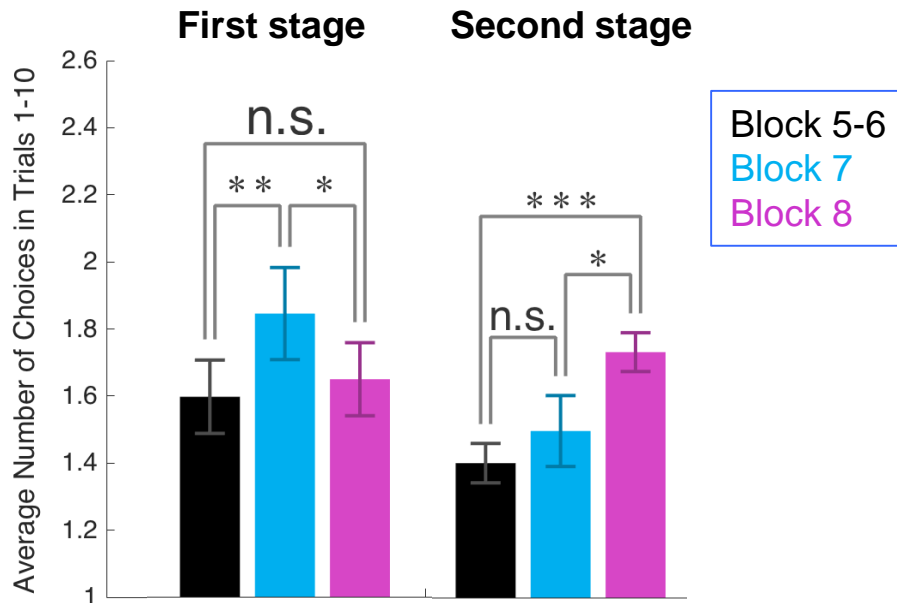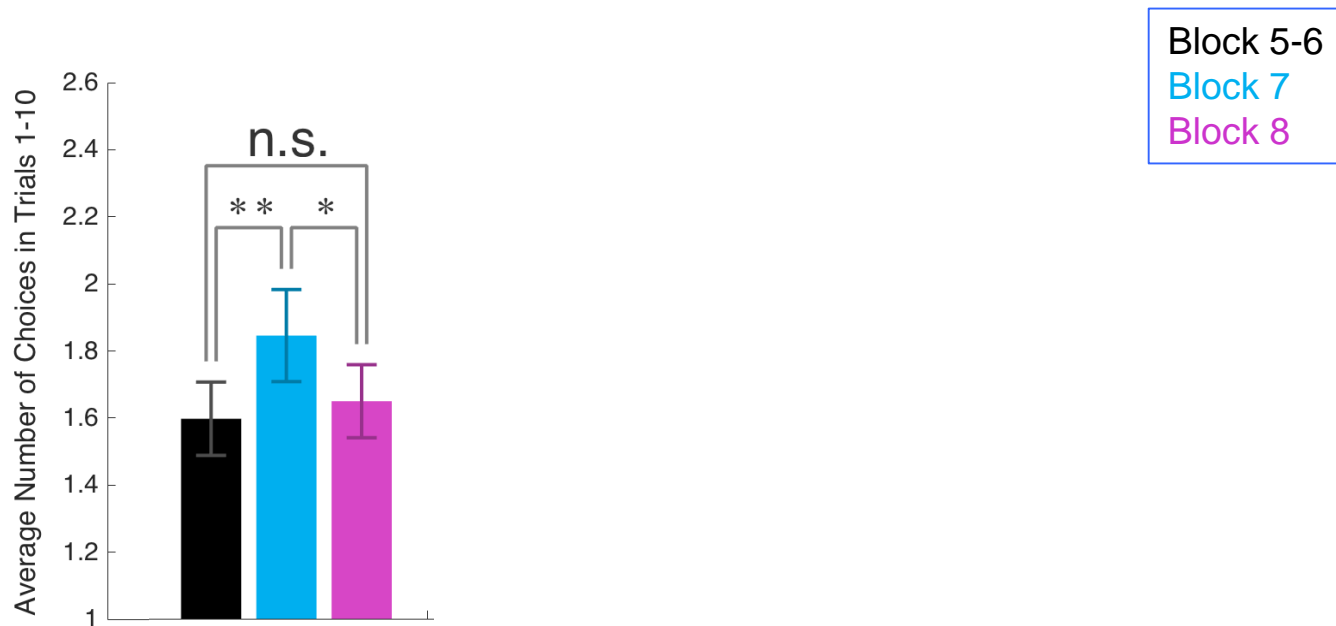
Participants learn and
(positive/negative) transfer **options**.

Participants learn and
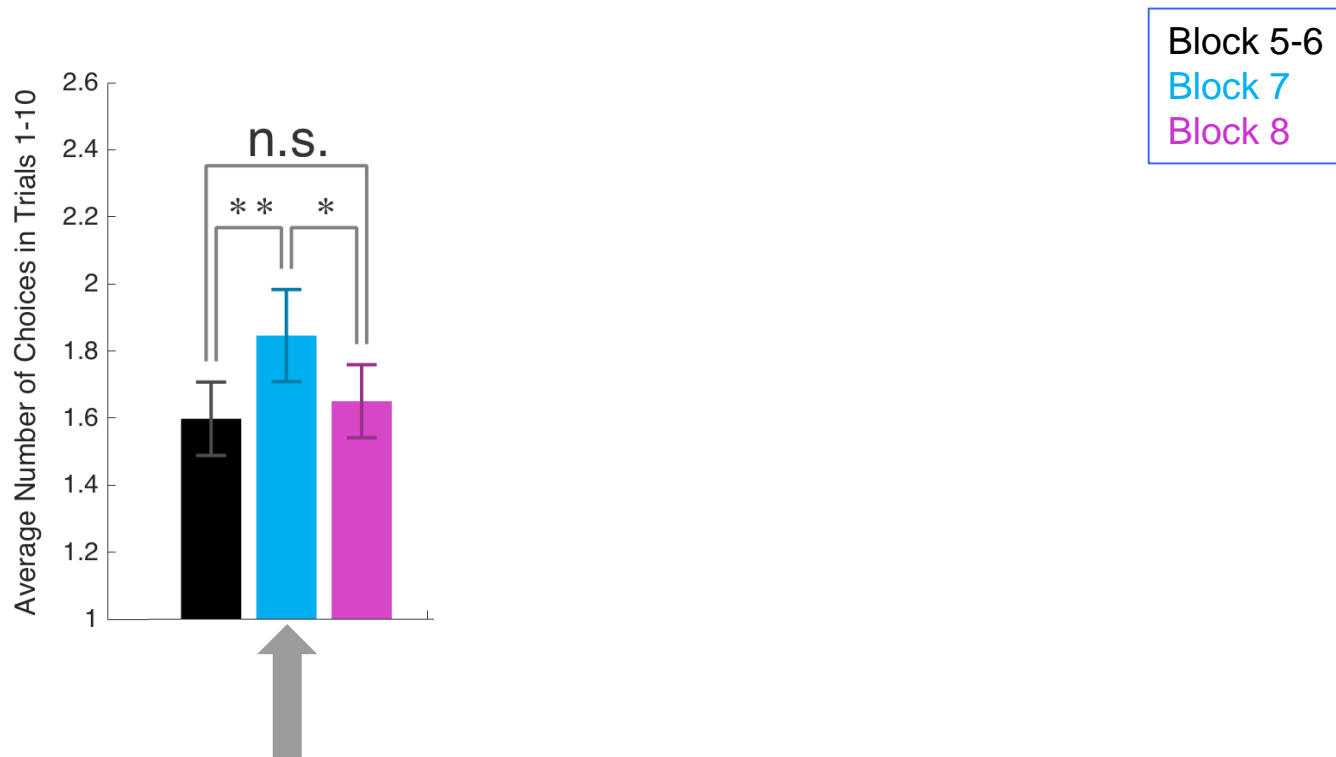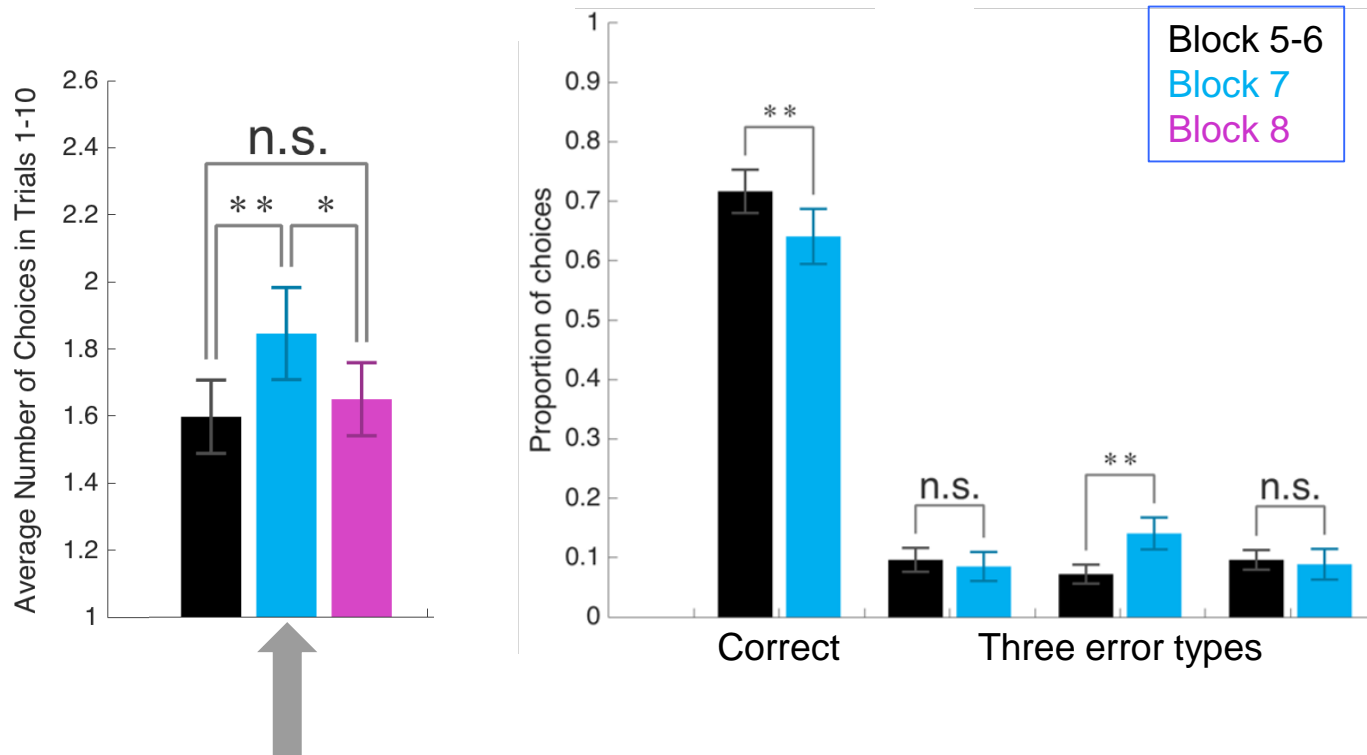(positive/negative) transfer **options of options**.

Xia & Collins, in prep

Zoom in the First Stage: Choice Type

Block 5-6
Block 7
Block 8

Zoom in the First Stage: Choice Type

Block 5-6
Block 7
Block 8

Xia & Collins, in prep

Zoom in the First Stage: Choice Type

Block 5-6
Block 7
Block 8

Xia & Collins, in prep
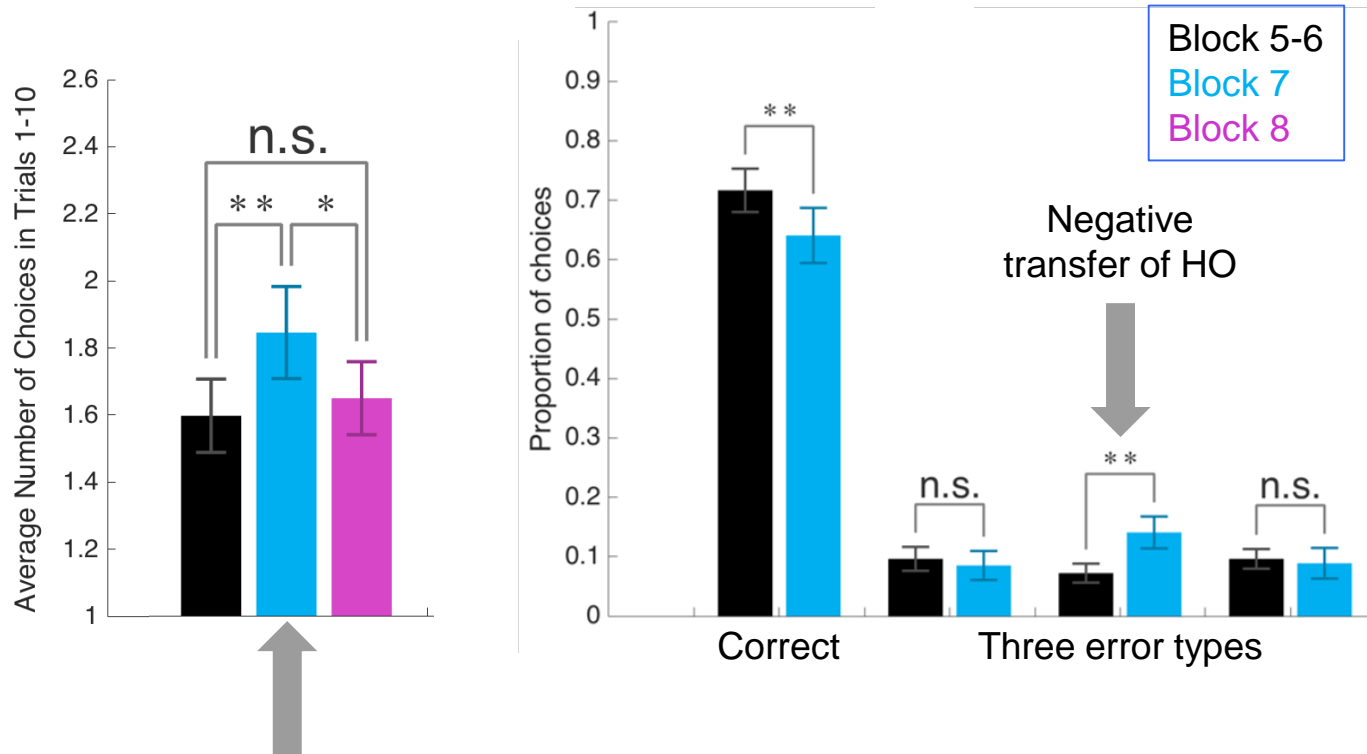
Zoom in the First Stage: Choice Type

Xia & Collins, in prep
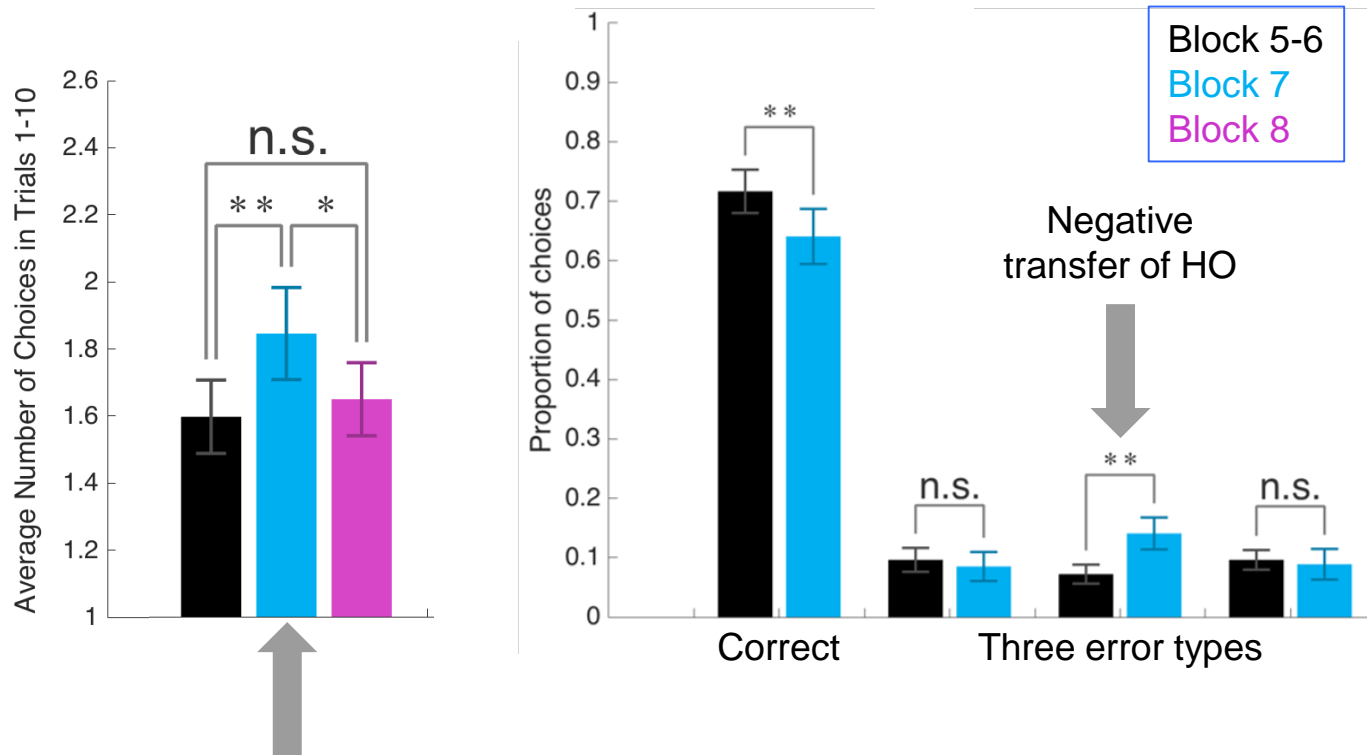
Zoom in the First Stage: Choice Type

Similar choice type analysis for all other transfer effects

Xia & Collins, in prep

Modeling

Modeling

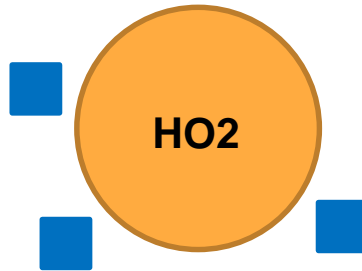- The Options Framework: Learn option-specific policies through Q-learning
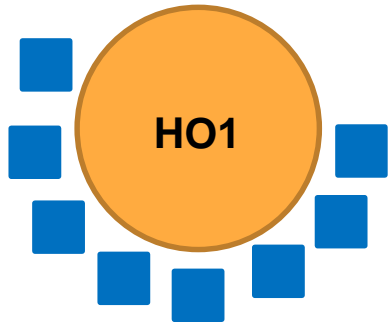
Modeling

- **The Options Framework:** Learn option-specific policies through Q-learning

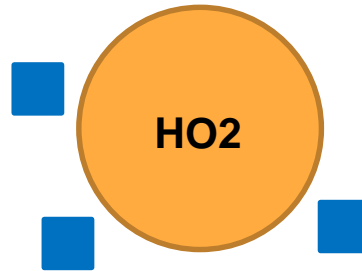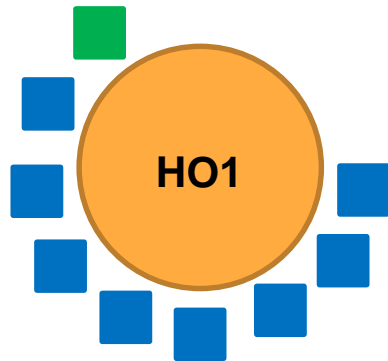- **Chinese Restaurant Process (CRP):** Cluster hidden contexts to achieve transfer
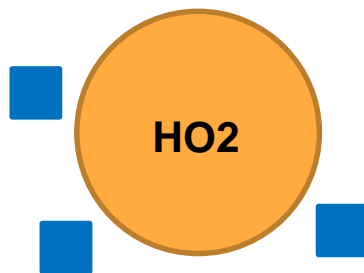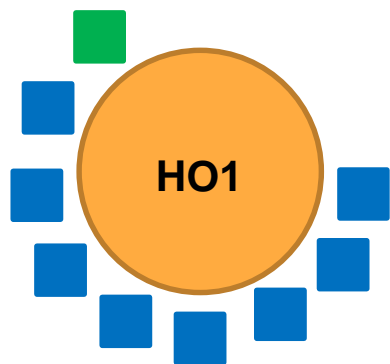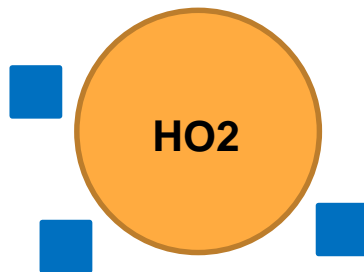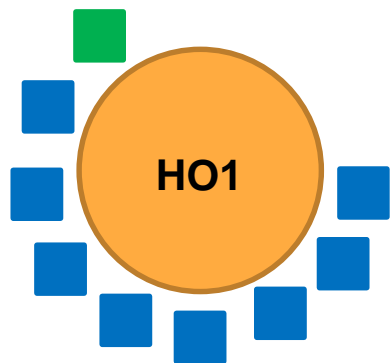
Modeling: First Stage CRP

Modeling: First Stage CRP

# Modeling: First Stage CRP

Context: temporal/block context

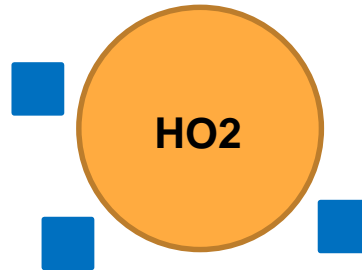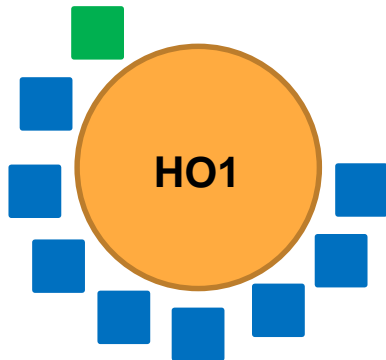# Modeling: First Stage HO1-specific Policy (Breakfast)

HO1	A1	A2	A3	A4
○				
□	Q-values learned through Q-learning			

Modeling: First Stage HO1-specific Policy (Breakfast)

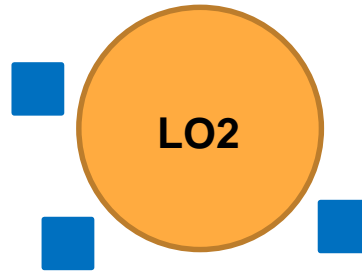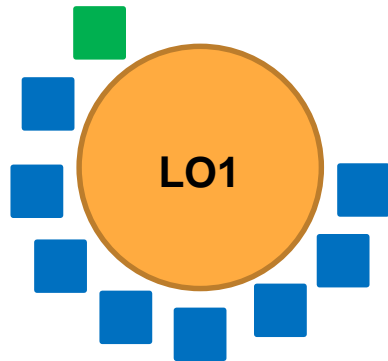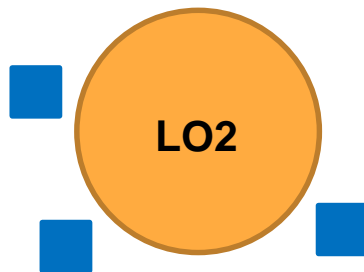| HO1	A1	A2	A3	A4
◯	Choosing MO1 (Coffee) as a whole			
▢				

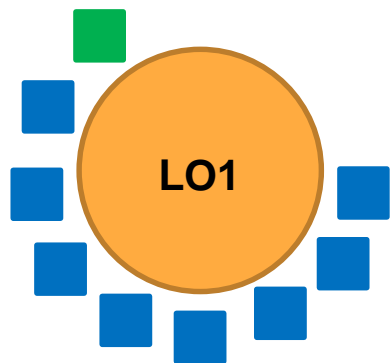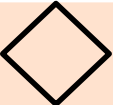# Modeling: Second Stage CRP

Which table?

LO1

LO2

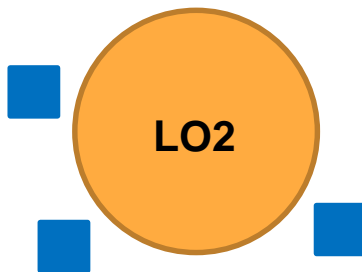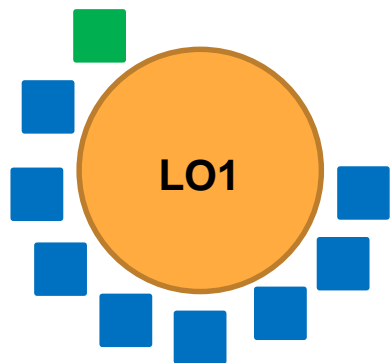New LO
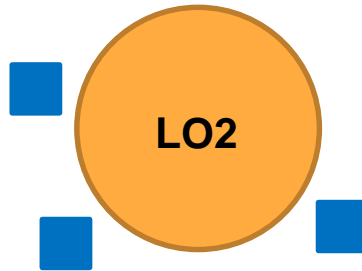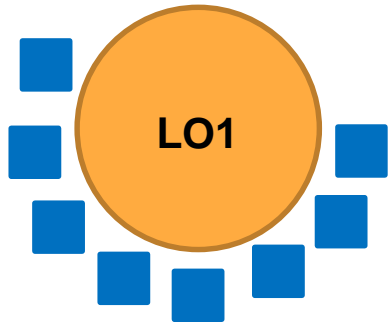
Modeling: Second Stage CRP

Modeling: Second Stage CRP

Context: MO
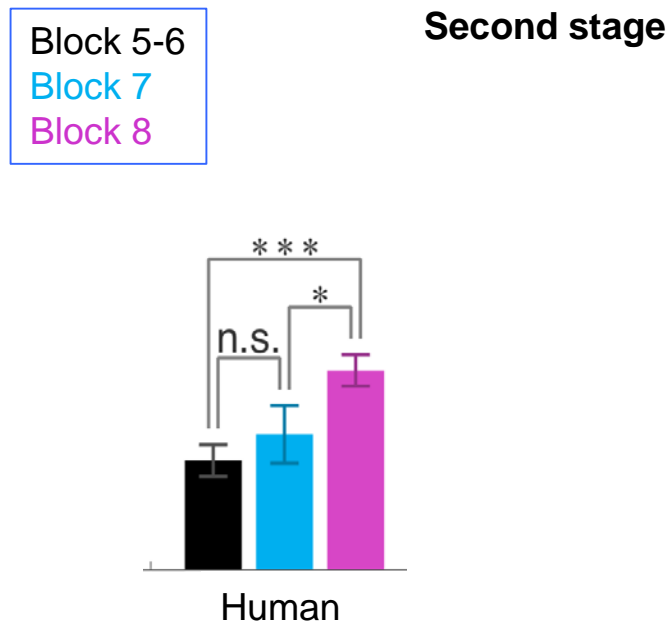
Modeling: Second Stage LO1-specific Policy (Boil)

LO1	A1	A2	A3	A4
◇	Q-values learned through Q-learning			
△				

Simulation Results
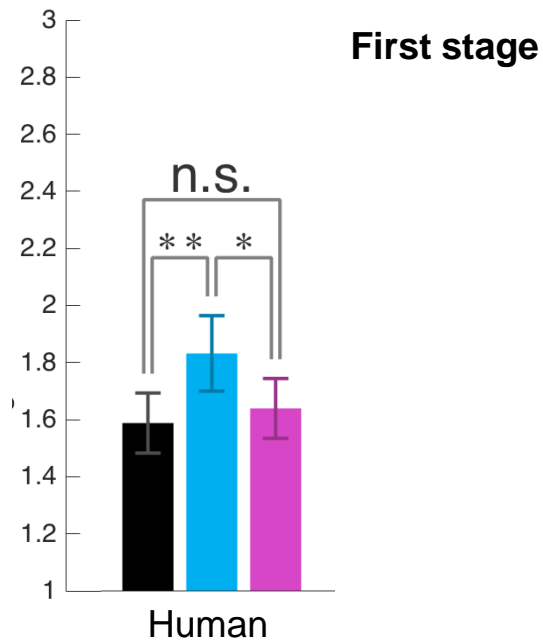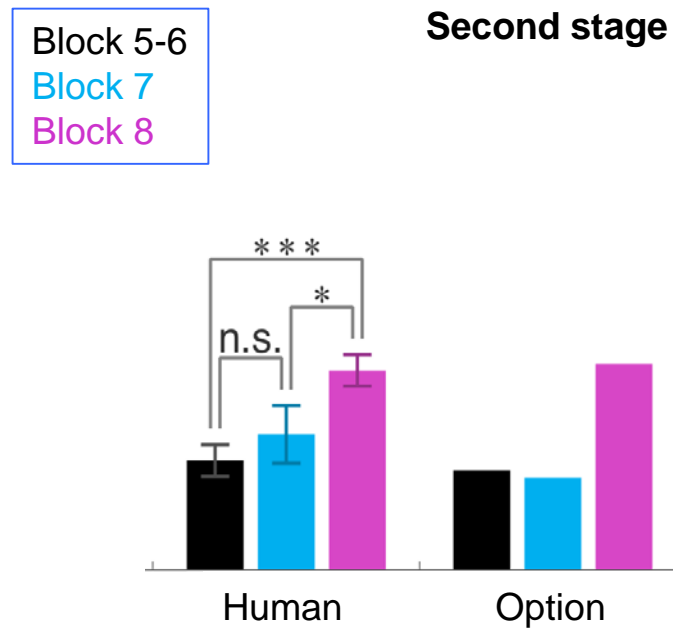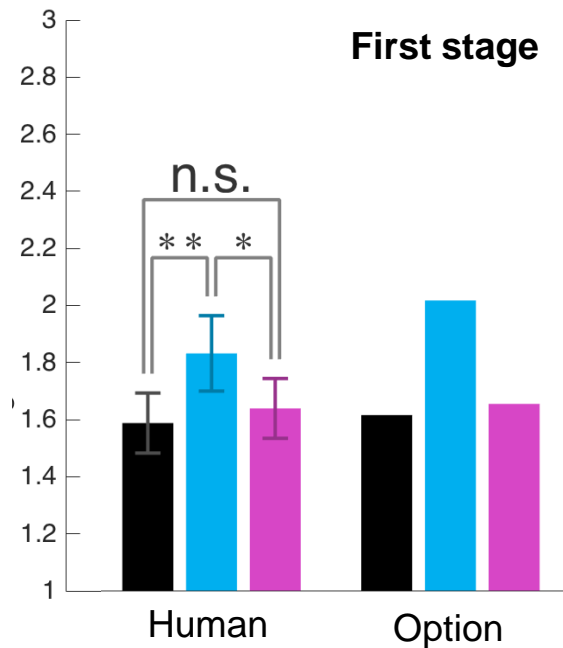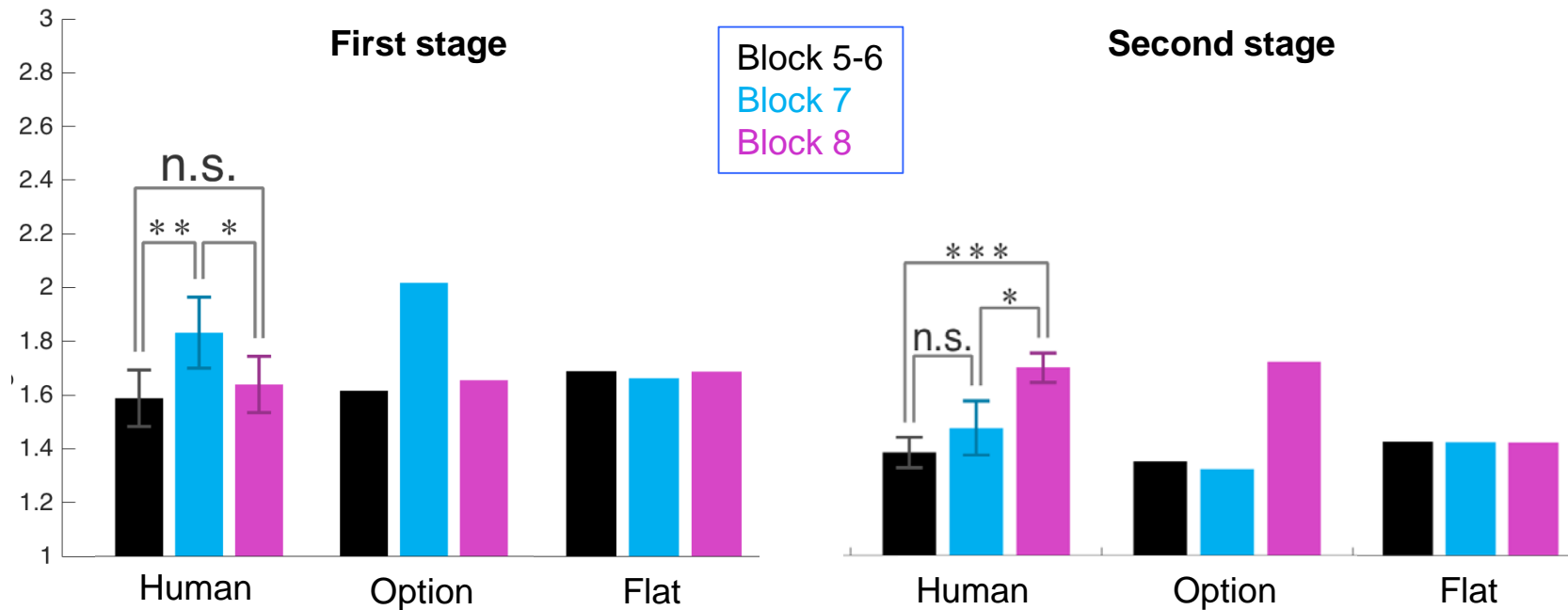
First stage

Block 5-6
Block 7
Block 8

Second stage

Xia & Collins, in prep

Simulation Results

Xia & Collins, in prep

Simulation Results

First stage

Block 5-6
Block 7
Block 8

Second stage

Xia & Collins, in prep

Summary

1. **Psy:** Behavioral signatures of option learning and transfer (and options of options) through a novel behavioral paradigm.

Block 5-6
Block7
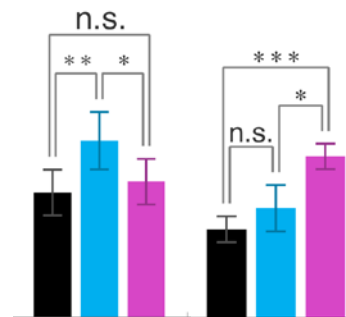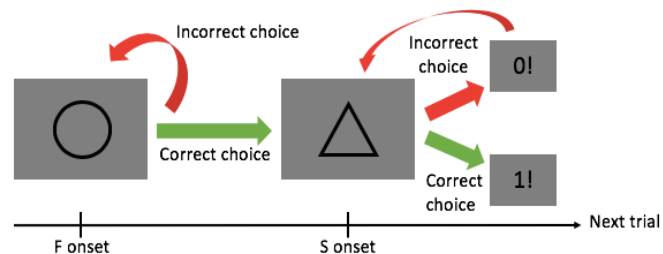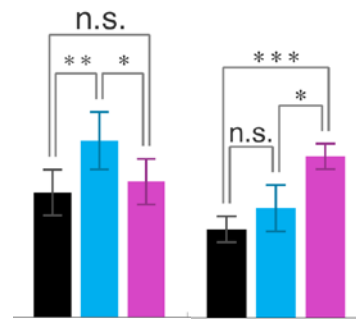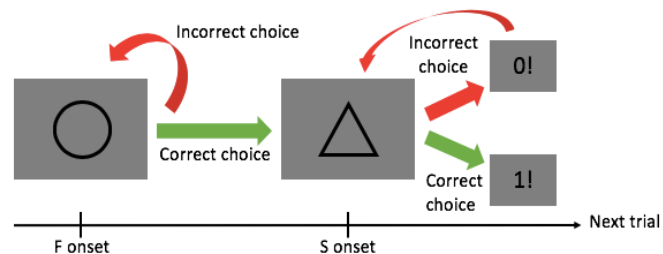Block8

Summary

1. **Psy:** Behavioral signatures of option learning and transfer (and options of options) through a novel behavioral paradigm.

2. **AI:** The Option Model + CRP is human-like (when flat RL fails) and enables flexible option transfer at multiple levels.
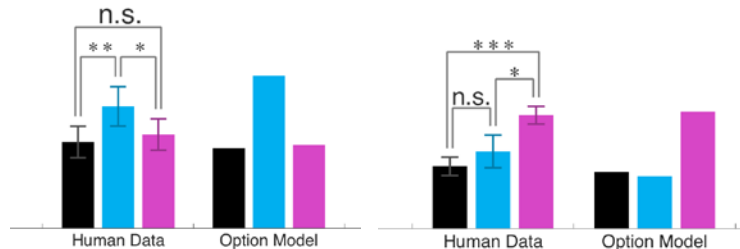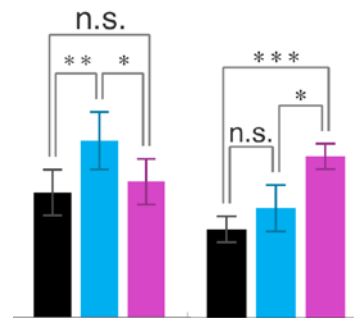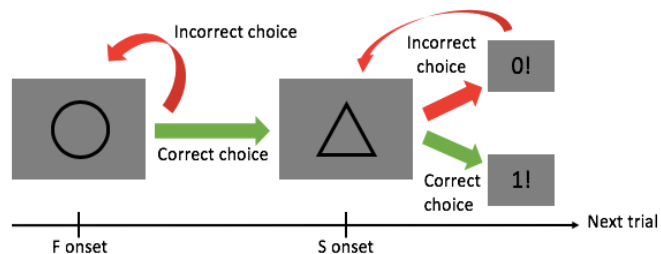

Block 5-6
Block7
Block8

Summary

1. **Psy:** Behavioral signatures of option learning and transfer (and options of options) through a novel behavioral paradigm.

2. **AI:** The Option Model + CRP is human-like (when flat RL fails) and enables flexible option transfer at multiple levels.

3. Three more experiments testing other aspects of options: **naturalistic**, **compositional**, etc, and replicating.

Block 5-6
Block7
Block8

71
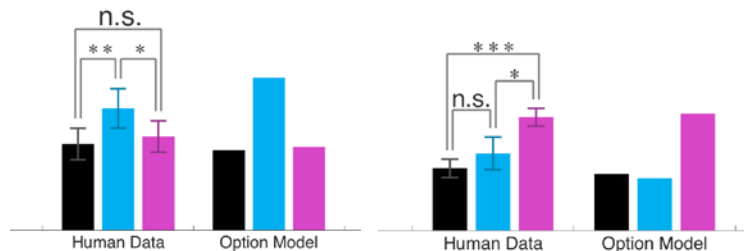
Acknowledgements

Contact me: jimmyxia@math.berkeley.edu

CCN Lab PI:
Anne Collins

Lab manager:
Sarah Master

Undergrad RAs:
Katya Brooun
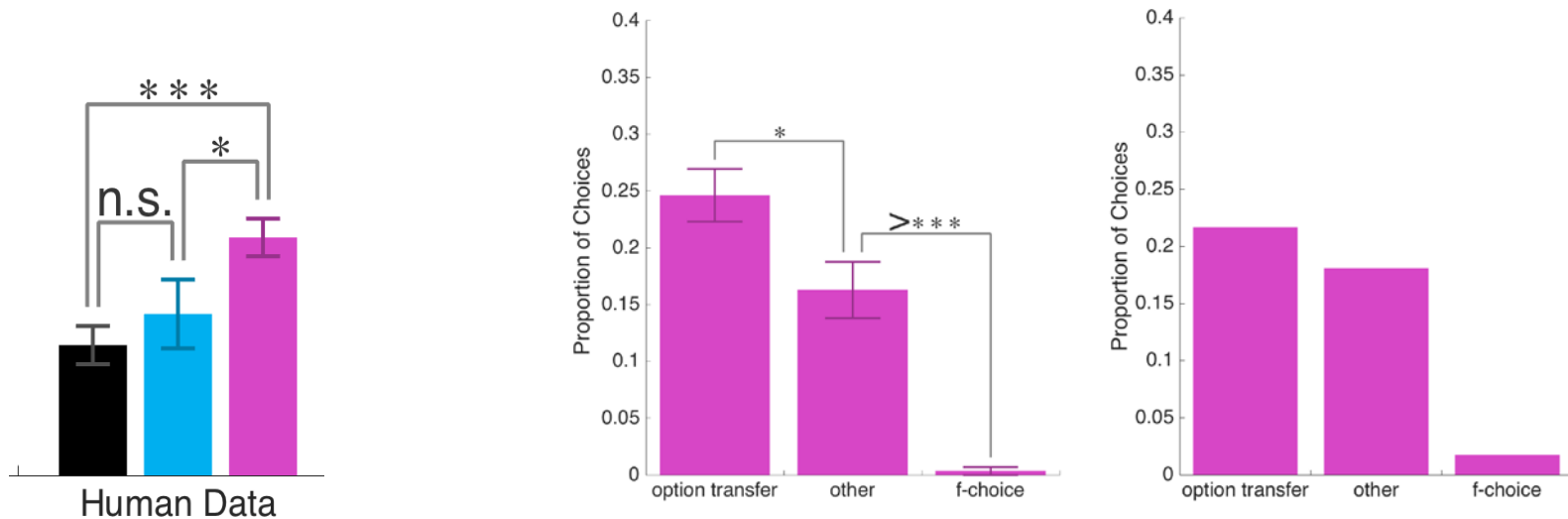Ham Huang
Helen Lu
Wendy Shi
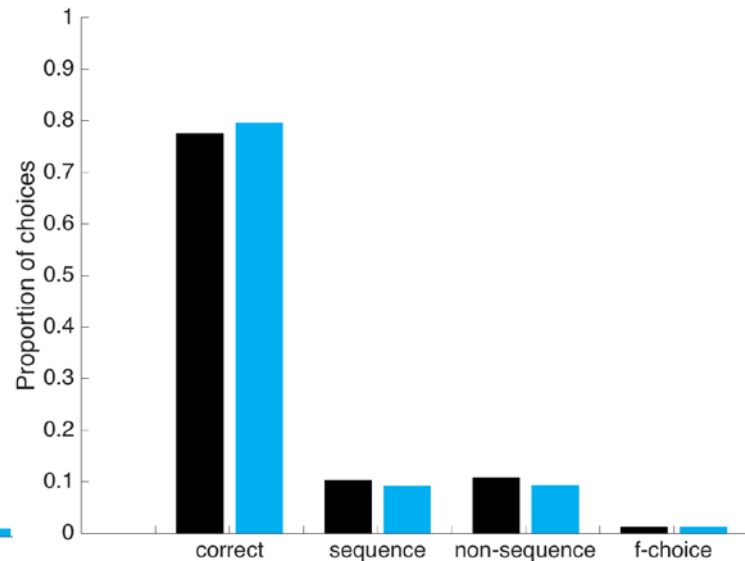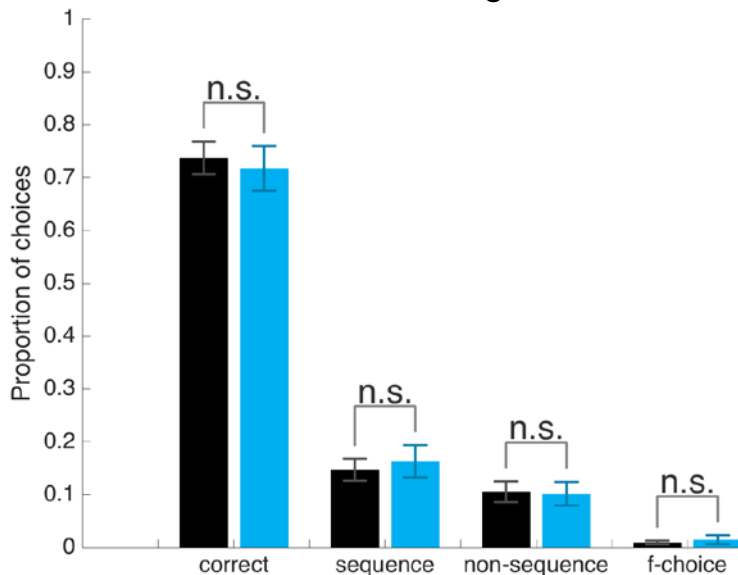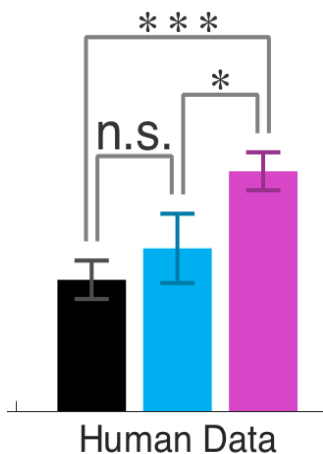
CCN Lab
NIH R01MH119383-01

RLDM

Block 8 Second Stage Error Type

Negative transfer dominated by the error type predicted by option transfer. Modeling replicates.
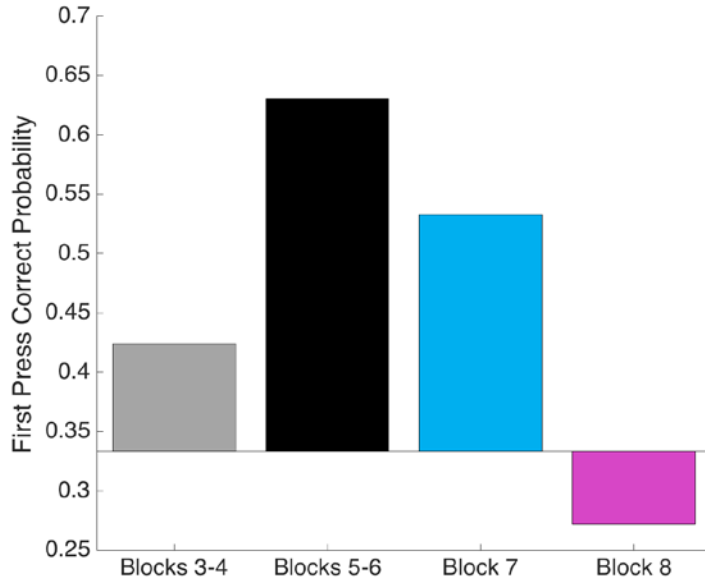

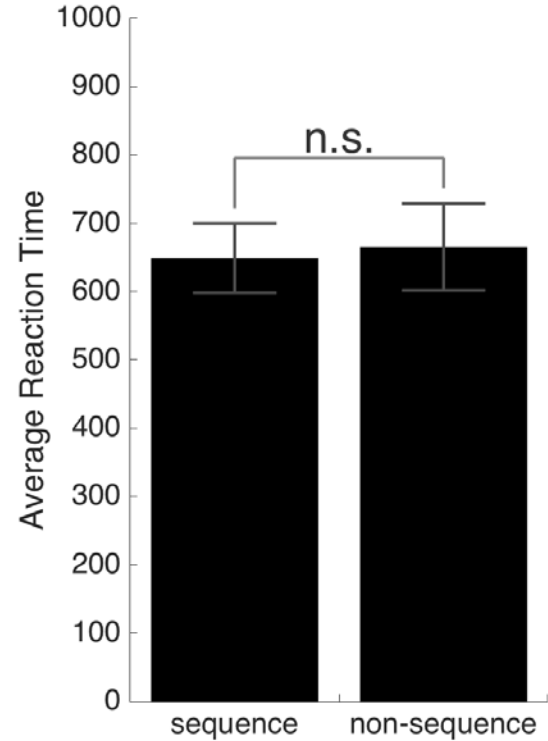Human Data

Block 7 Second Stage Error Type

Positive transfer: no significant difference across all types. Modeling replicates.

More Second Stage Analysis

First press in the second stage: significant (positive/negative) in the first press of the block.

Reaction time: rule out pure sequence learning effects