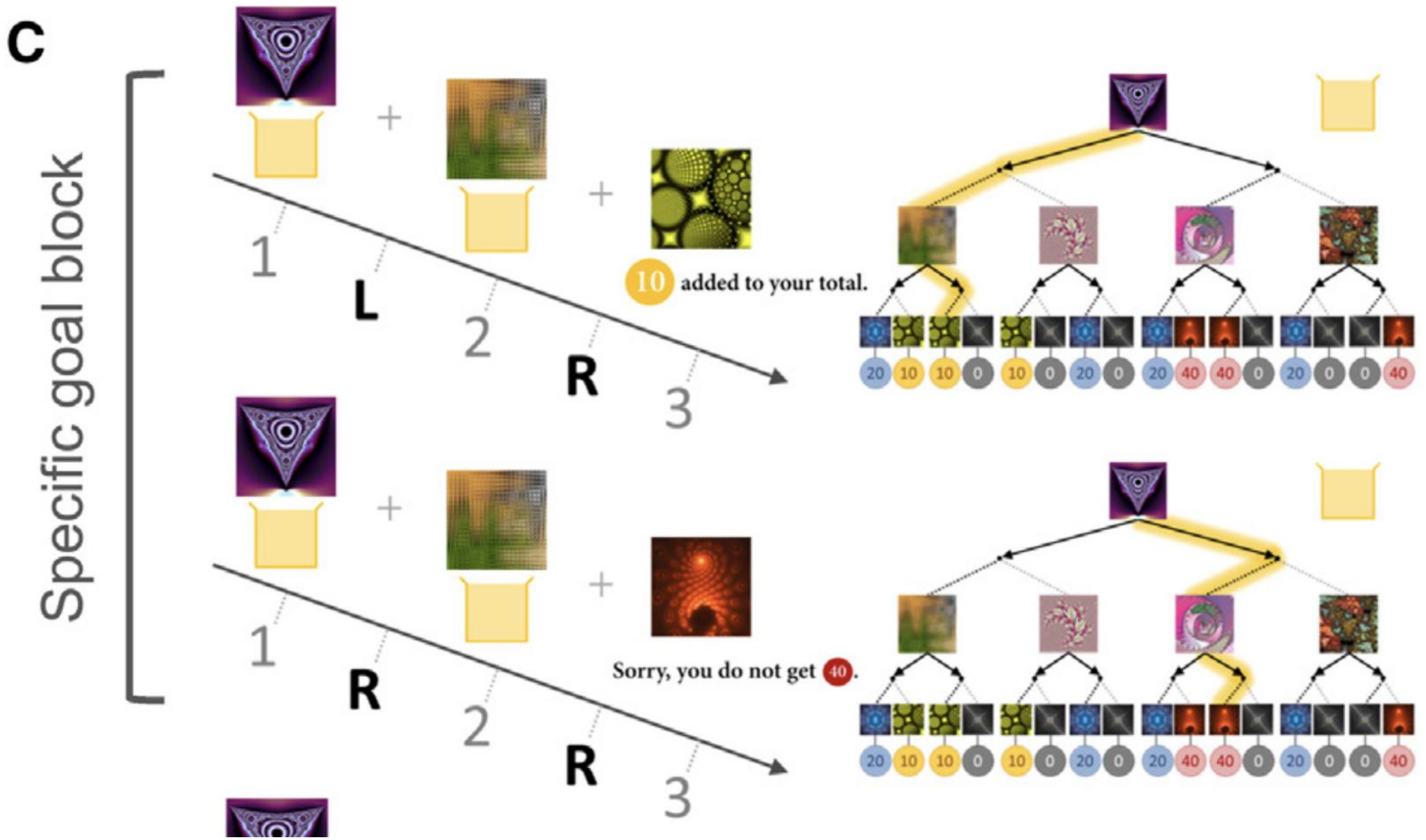
Competition between Model-based and Model-free Learning

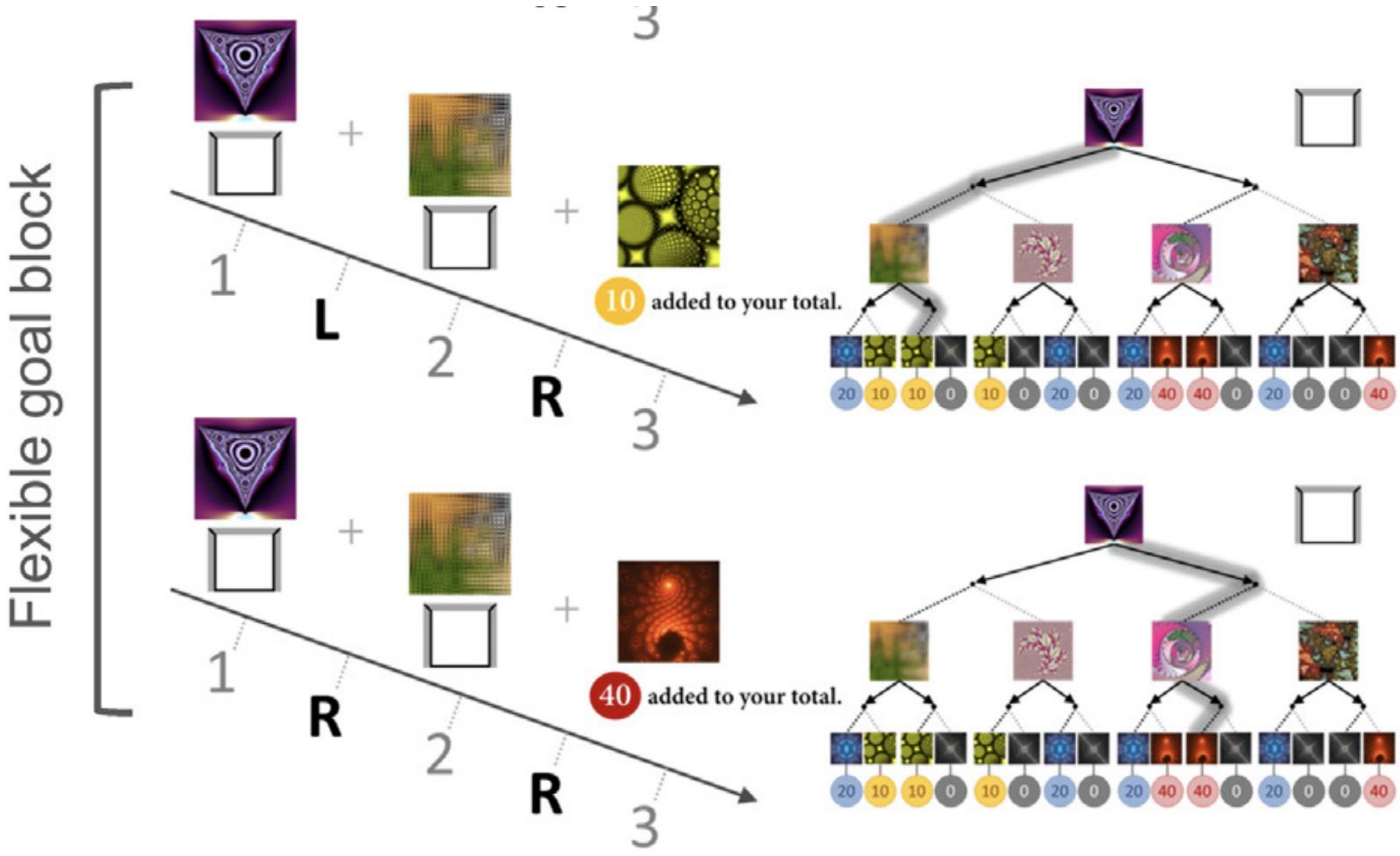
——Arbitration model

张博涛 进展汇报

Task setting: Two Stage Task

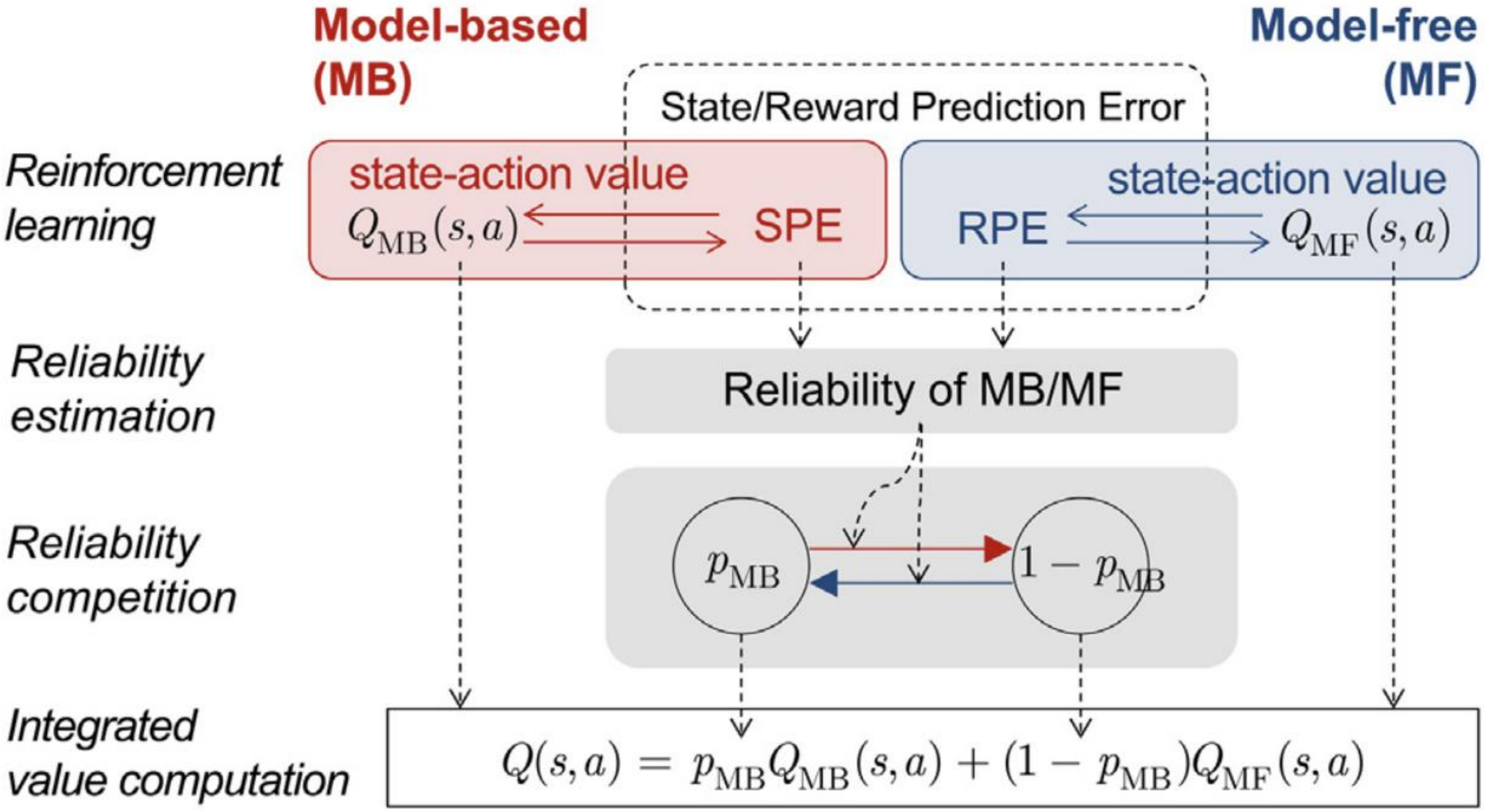


Task setting: Two Stage Task



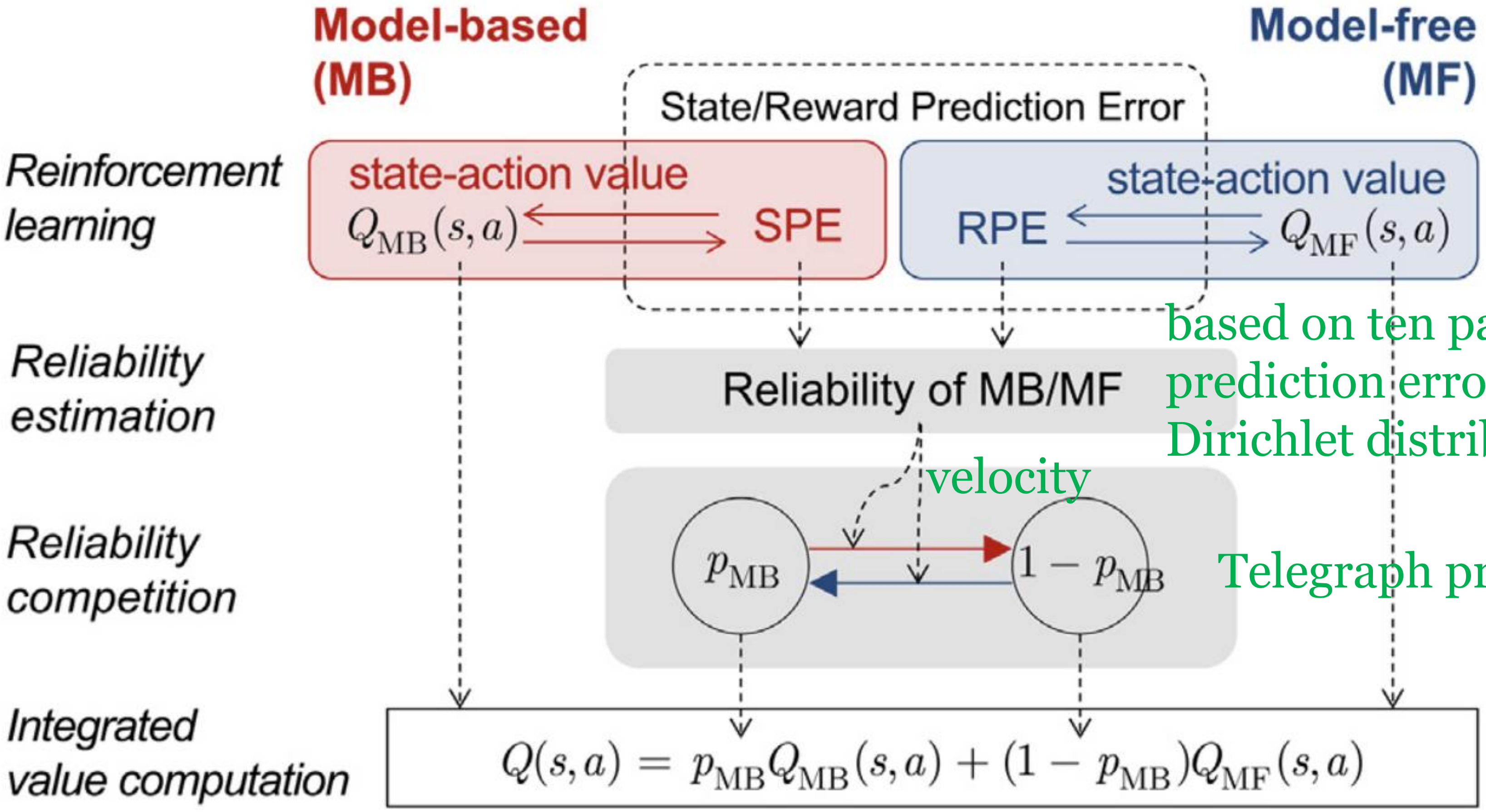
Neural Computations Underlying Arbitration between Model-Based and Model-free Learning

Sang Wan Lee,[1,2,3,*] Shinsuke Shimojo,[1,2,4] and John P. O'Doherty[1,2,3]



Neural Computations Underlying Arbitration between Model-Based and Model-free Learning

Sang Wan Lee,[1,2,3,*] Shinsuke Shimojo,[1,2,4] and John P. O'Doherty[1,2,3]



based on ten past trials' prediction errors. model with Dirichlet distribution

velocity

Telegraph process

Loss4arbitration Model

Integrated value computation

$$\boxed{Q(s,a) = p_{\mathrm{MB}} Q_{\mathrm{MB}}(s,a) + (1 - p_{\mathrm{MB}}) Q_{\mathrm{MF}}(s,a)}$$

$$w = argmin_w Q_w(s,a) - Q_{true}(s,a) \text{ where}$$

$$Q_w(s,a) = wQ_{MB}(s,a) + (1-w)Q_{MF}(s,a)$$

Loss4arbitration Model

Integrated value computation

$$\boxed{Q(s,a) = p_{\mathrm{MB}} Q_{\mathrm{MB}}(s,a) + (1 - p_{\mathrm{MB}}) Q_{\mathrm{MF}}(s,a)}$$

$$w = argmin_w Q_w(s,a) - Q_{true}(s,a) \text{ where}$$

$$Q_w(s,a) = w Q_{MB}(s,a) + (1-w) Q_{MF}(s,a)$$

$$\text{gradient descent on } \mathcal{L} = (Q_w - Q_{true})^2$$

Model Fitting

minimize negative loglikelihood

subjects' data: true actions, a_t

model's prediction: $\mathcal{P} \sim \mathrm{softmax}(Q_w)$

negative loglikelihood: $NLL = \sum \mathcal{P}(a_t)$

Model Fitting

150 trials.

trial 1 ~ 37: fixed goal(20), low uncertainty(0.9,0.1)

trial 38 ~ 75: flexible goal, low uncertainty

trial 76 ~ 112: fixed goal(10), high uncertainty(0.5)

trial 113 ~ 150: flexible goal, high uncertainty(0.5)

back planning(MB)

Model Comparison

MB: pure model-based learning

MF: pure model-free learning

Hybrid: $Q = wQ_{MB} + (1-w)Q_{MF}$, but fixed w

Loss4arbitration:

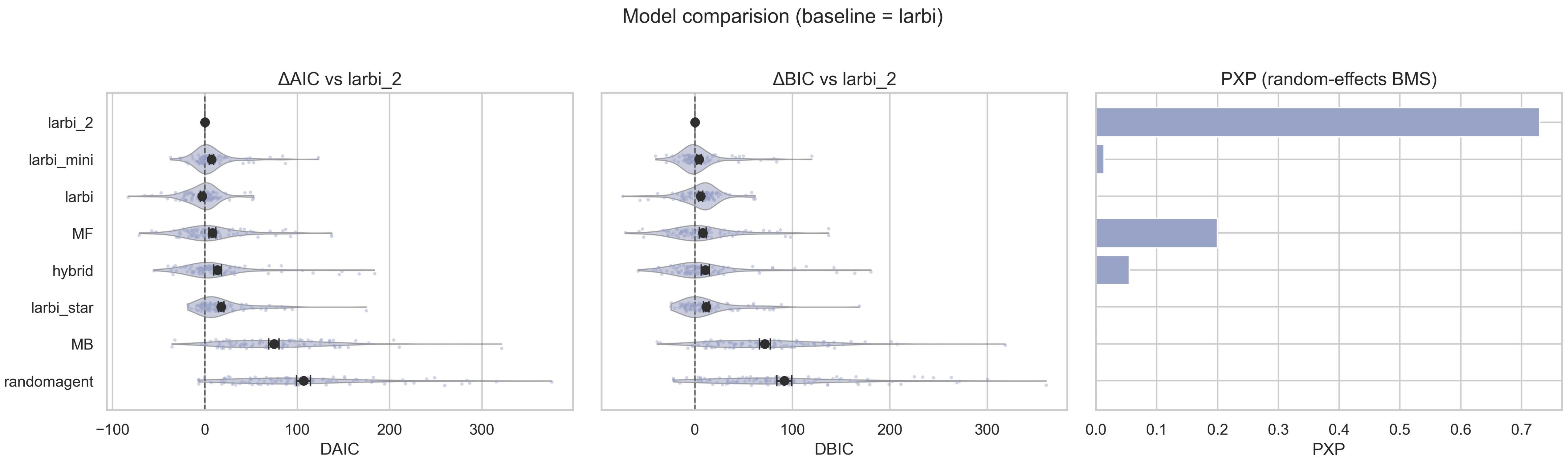
1. larbi: MF learning rates * 2, MB … * 2, discount, w's learning rate, exploration temperature * 2

2. larbi_mini: MF, MB, w learning rates, exploration temperature * 1

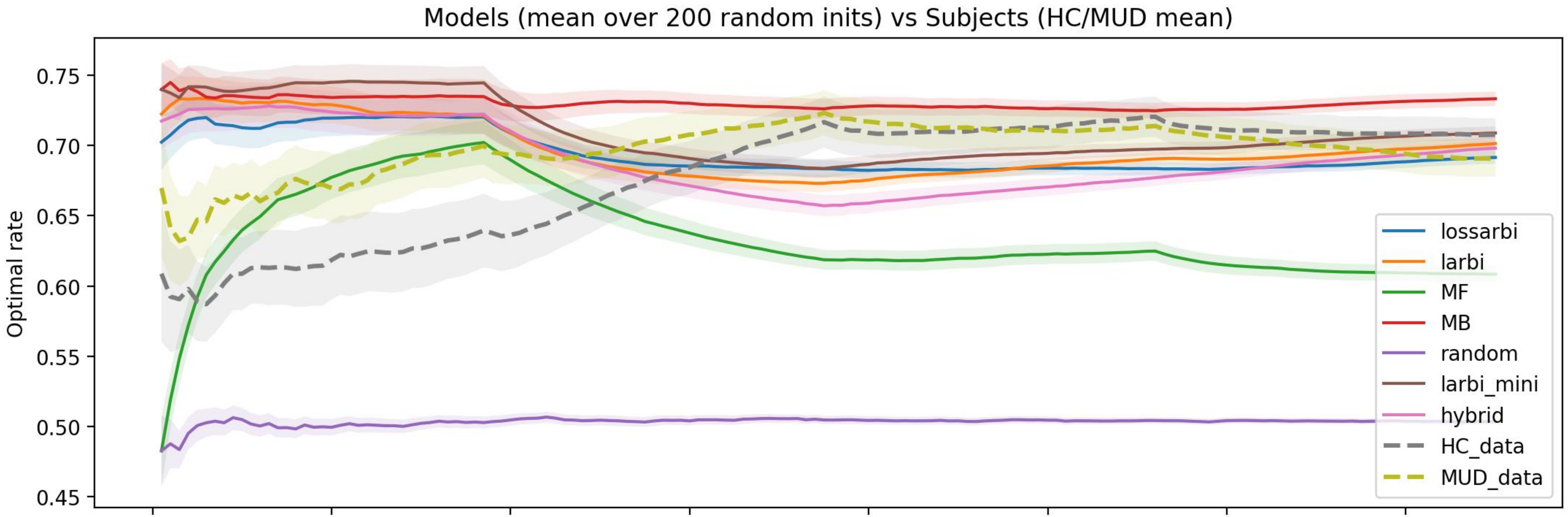
3. larbi_star: based on mini, but MB learning rate = MF's

4. larbi_2: based on mine, but a separate arbitration w at each stage

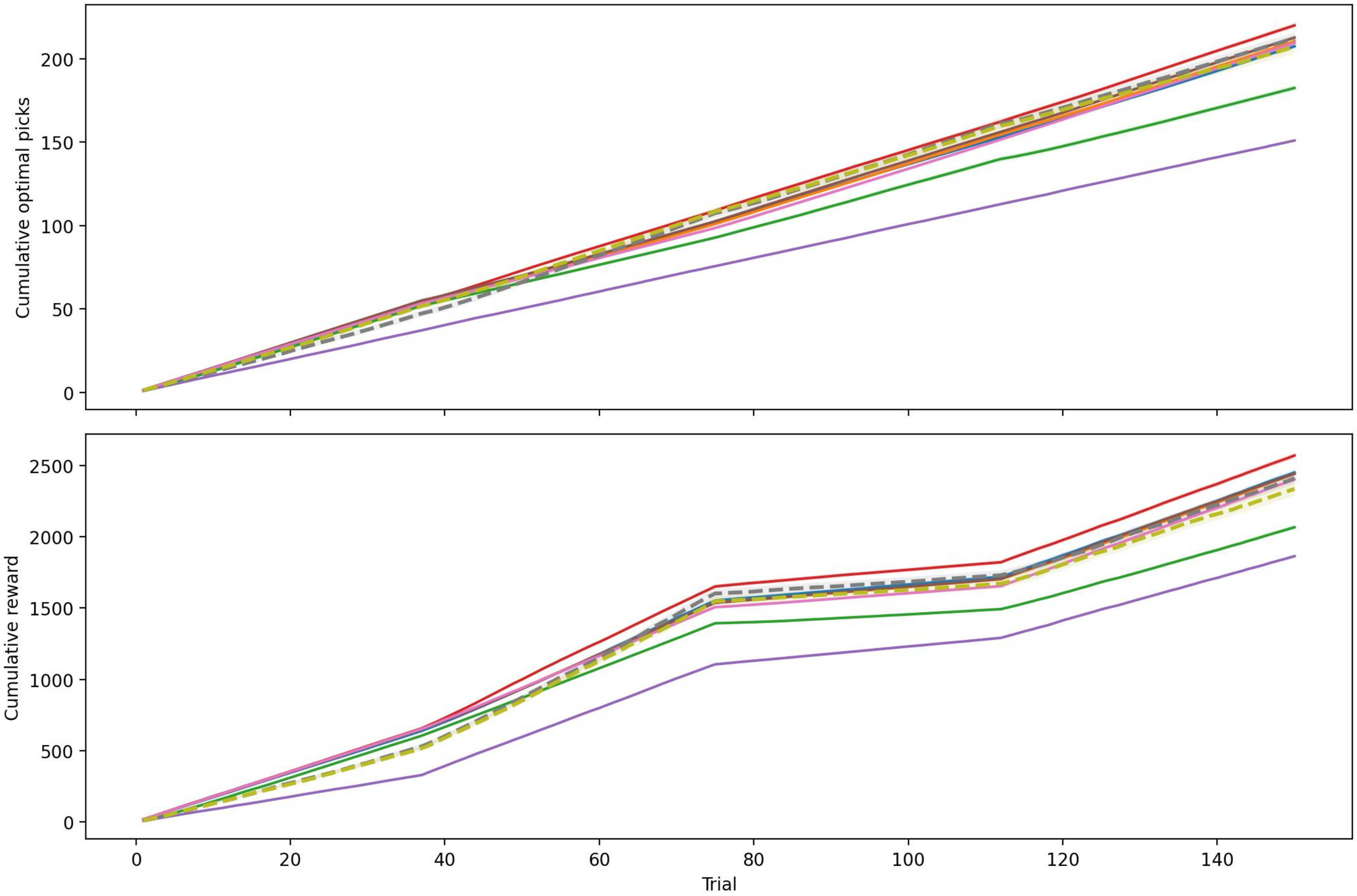
Model Comparison



Model Comparison

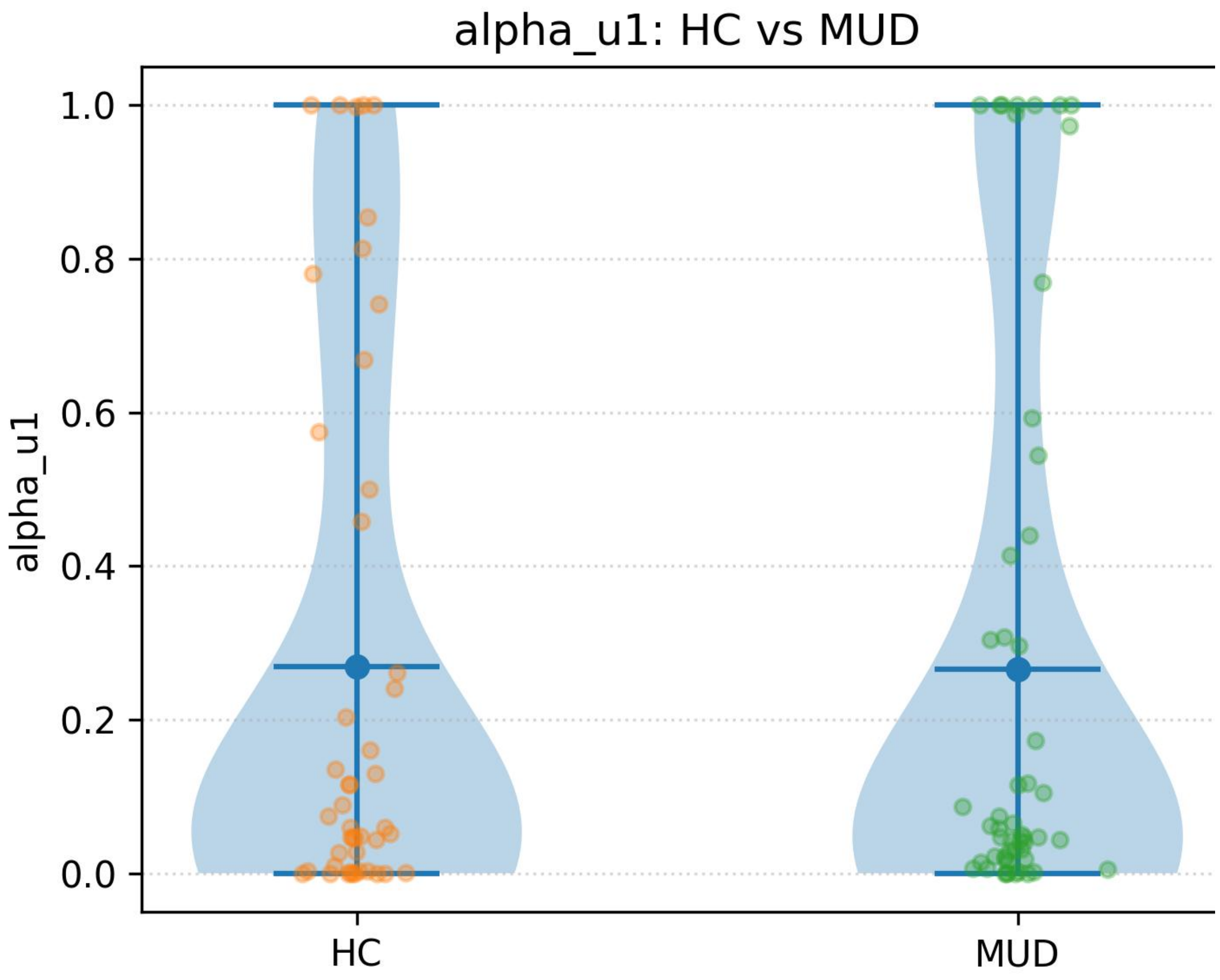


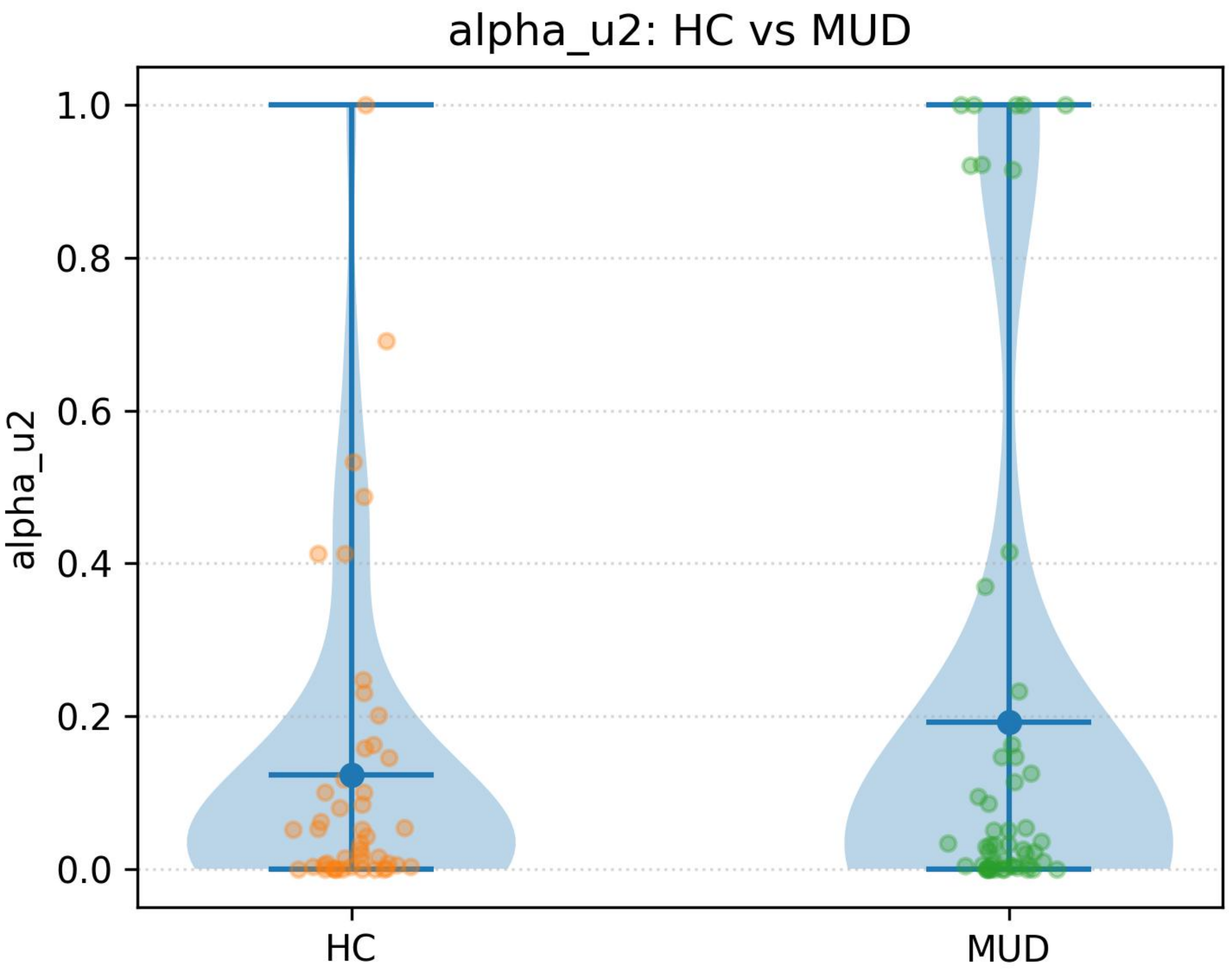
Model Comparison



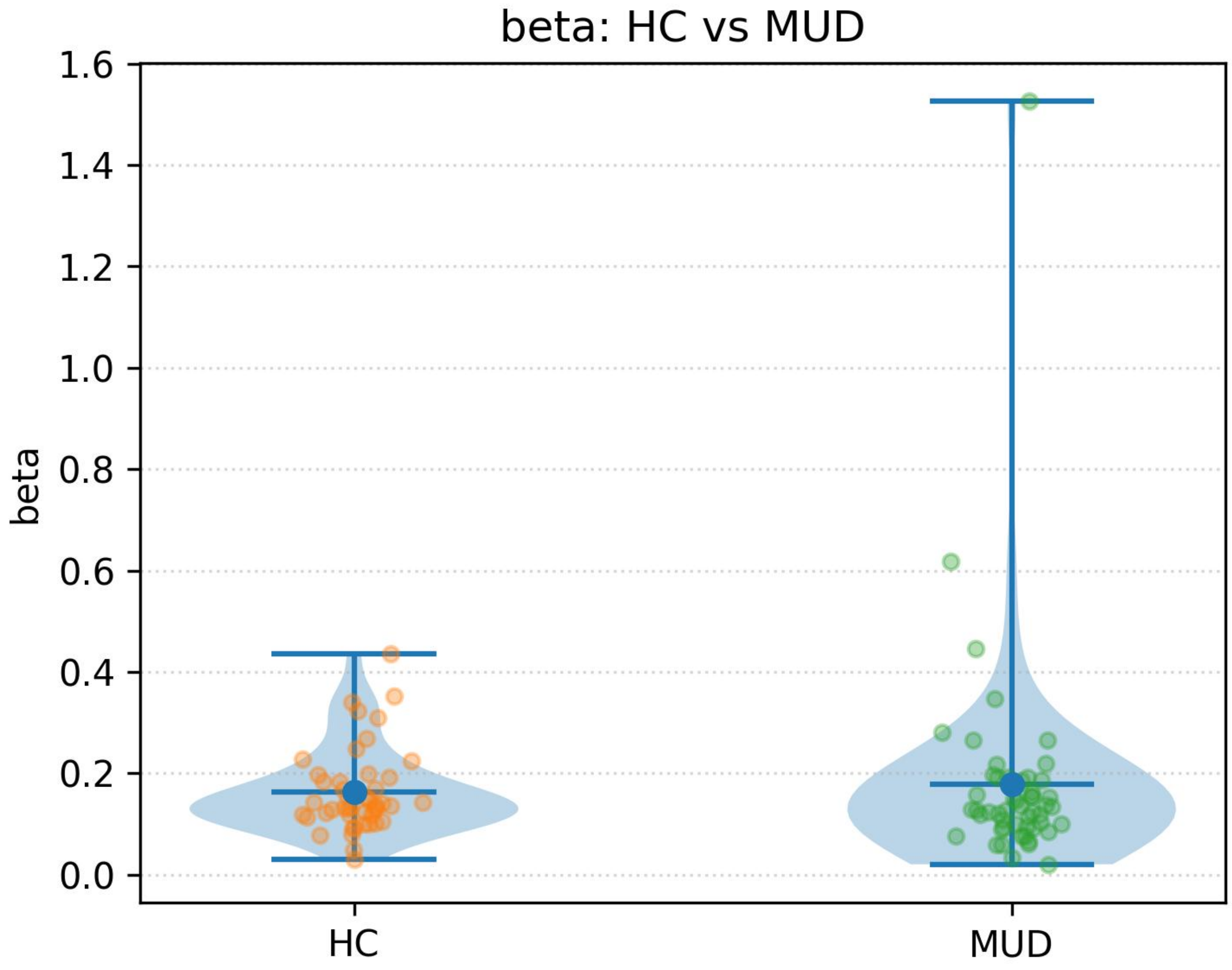
Model Comparison

Results Analysis

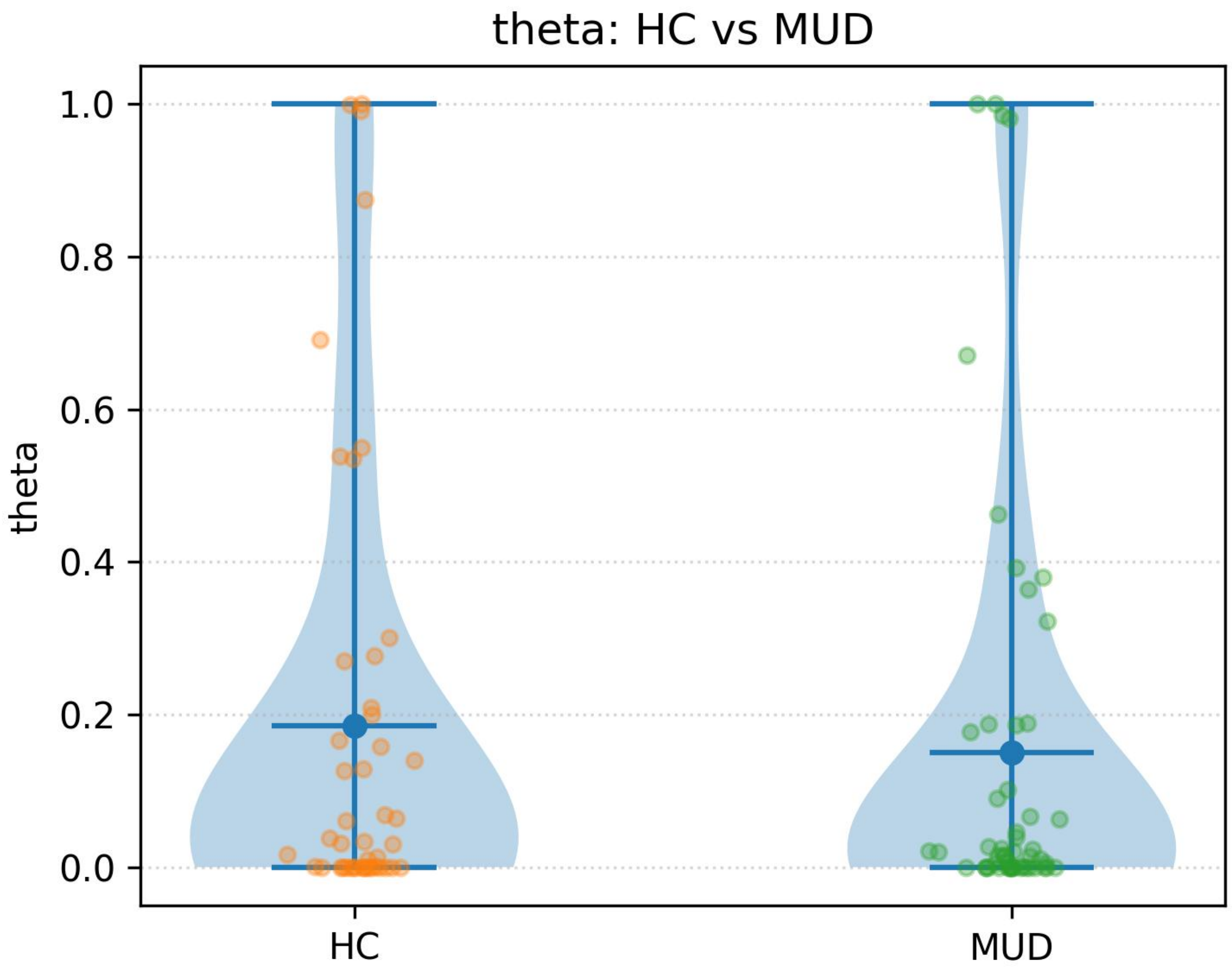




Results Analysis

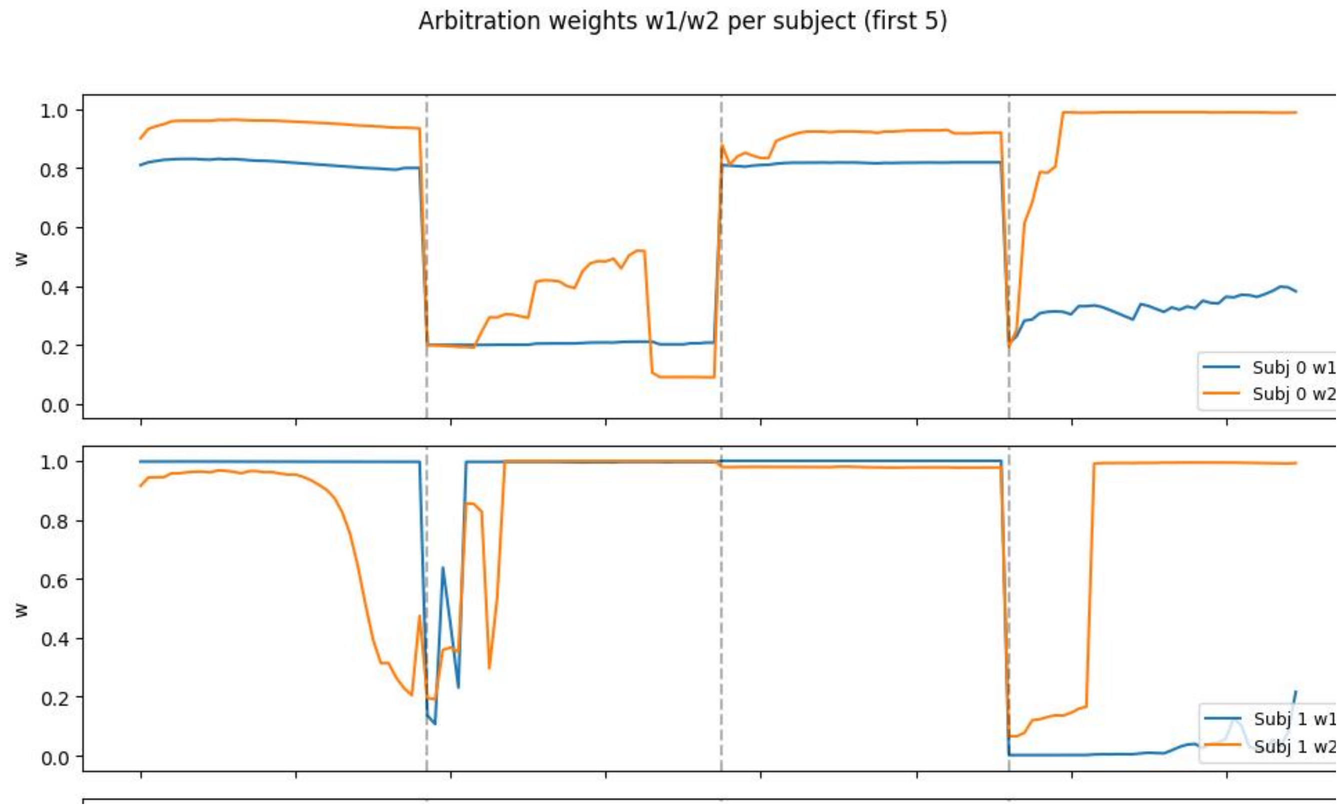


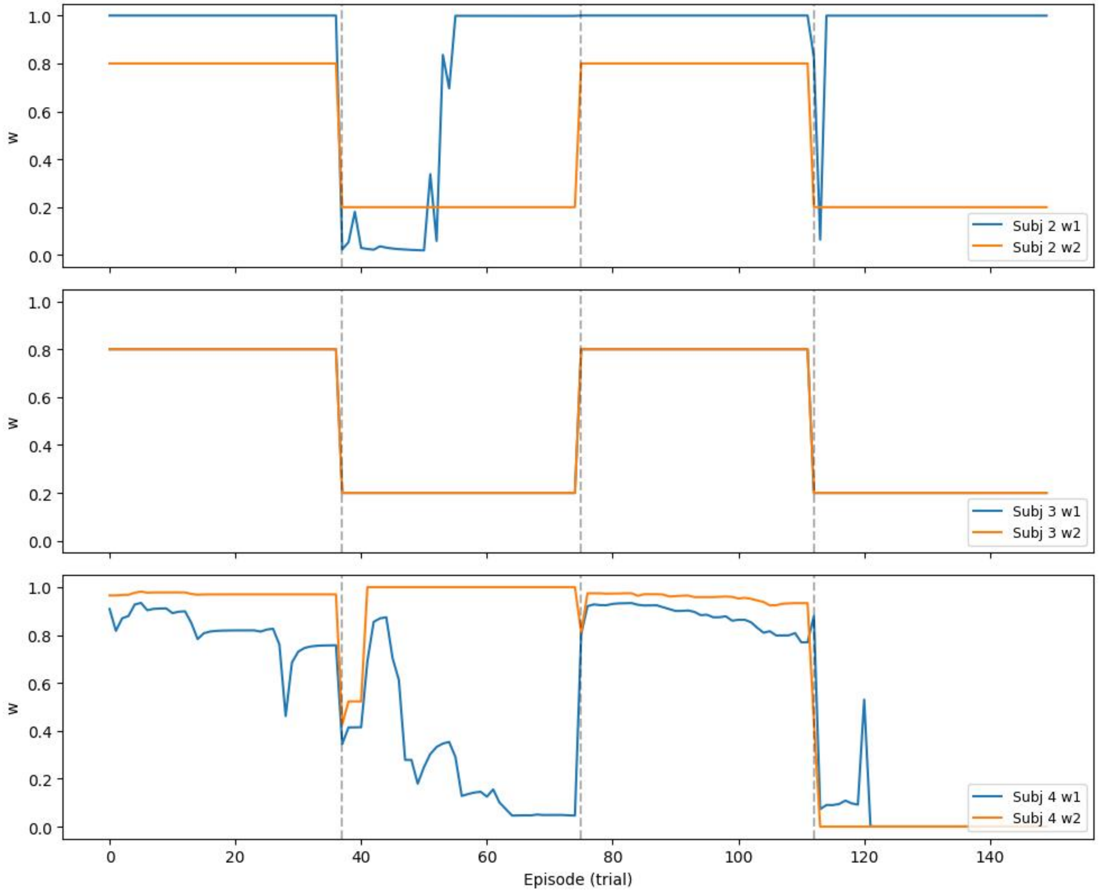
Results Analysis



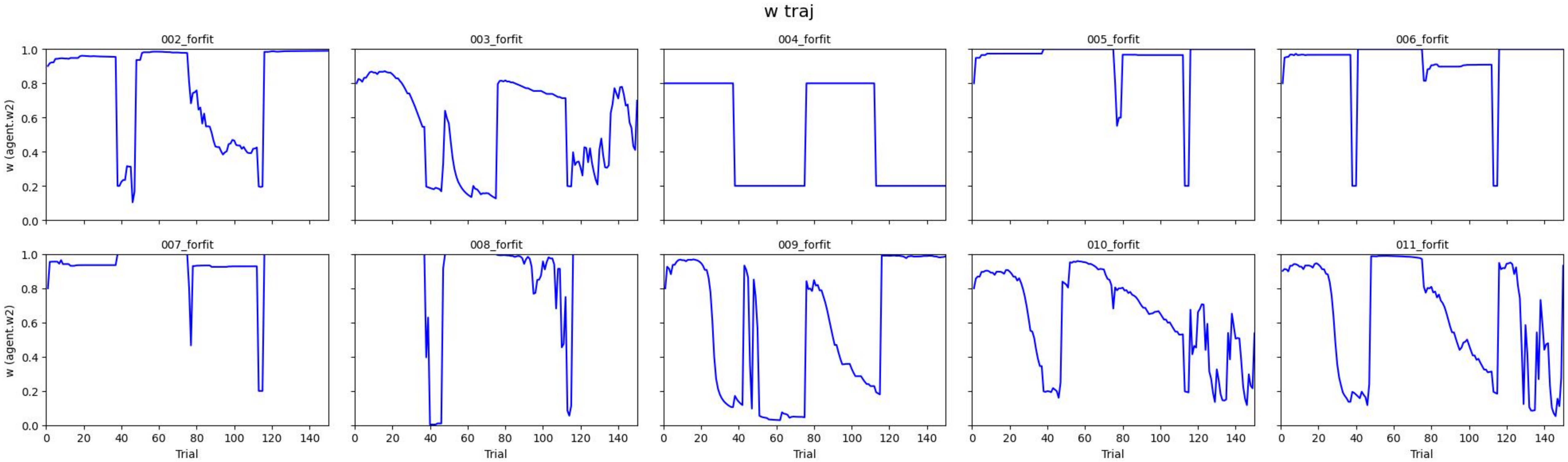
around 1/3 is 0

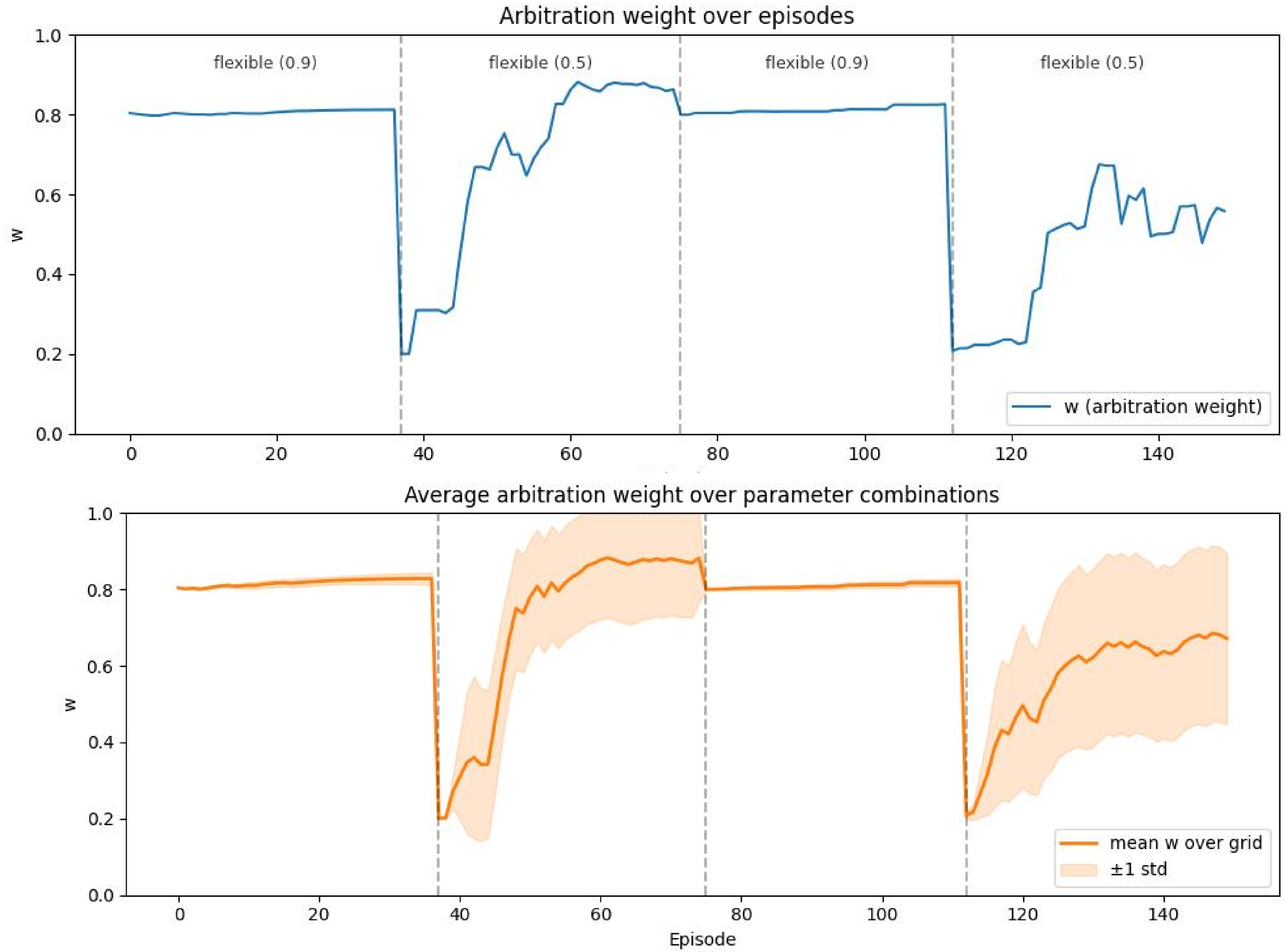
Results Analysis



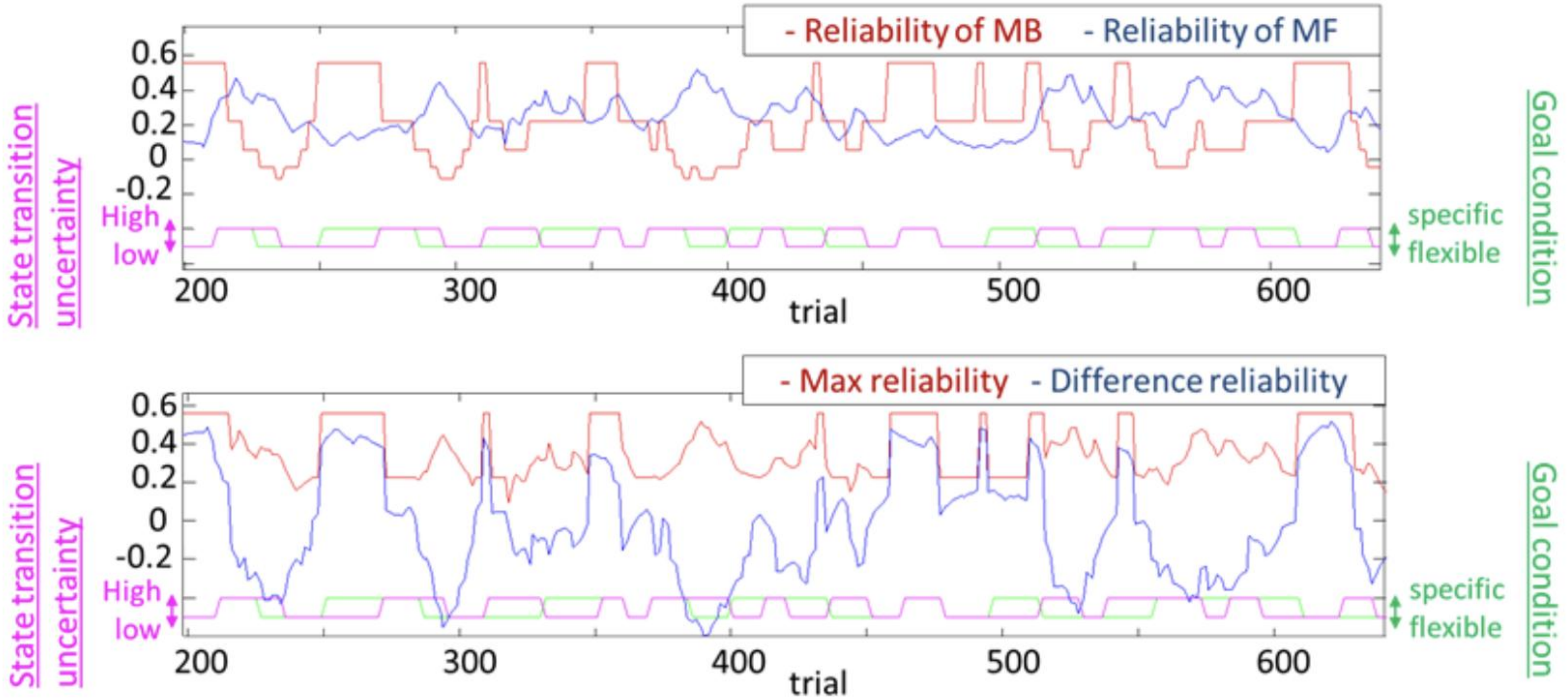


Results Analysis




Arbitration weight over episodes
flexible (0.9)
flexible (0.5)
flexible (0.9)
flexible (0.5)
w
w (arbitration weight)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140
Average arbitration weight over parameter combinations
w
mean w over grid
±1 std
Episode

Possible Explanations



Possible Explanations

However, to incentivize subjects to continue learning, throughout the task, the chances of payoff associated with the four second-stage options were changed slowly and independently, according to Gaussian random walks.

In the Lee 2014 task setting, perhaps Model-Based ***is*** oftentimes better?
(also, from the total reward attained at each block)

The current state of research on MB-MF competition/cooperation

Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control

Nathaniel D Daw[1], Yael Niv[1,2] & Peter Dayan[1]

Speed/Accuracy Trade-Off between the Habitual and the Goal-Directed Processes

Mehdi Keramati[1]*, Amir Dezfouli[2], Payam Piray[2]

1 School of Management and Economics, Sharif University of Technology, Tehran, Iran, **2** Control and Intelligent Processing Center Of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

When Does Model-Based Control Pay Off?

Wouter Kool[1]*, Fiery A. Cushman[1], Samuel J. Gershman[1,2]

1 Department of Psychology, Harvard University, Cambridge, Massachusetts, United States of America,
2 Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America

Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems

Wouter Kool[1], Samuel J. Gershman[1,2], and Fiery A. Cushman[1]
[1]Department of Psychology, Harvard University, and [2]Center for Brain Science, Harvard University

Beyond dichotomies in reinforcement learning

Anne G. E. Collins and Jeffrey Cockburn

MB-MF competition/cooperation

Two Stage Task

Humans are primarily model-based and not model-free learners in the two-stage task

Carolina Feher da Silva[1] and Todd A. Hare[1]

[1]Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Zurich, Switzerland

(In) Relation to Memory

How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis

Anne G. E. Collins and **Michael J. Frank**

"systematically varying the size of the learning problem and delay between stimulus repetitions"

Working-memory capacity protects model-based learning from stress

A. Ross Otto[a,1], Candace M. Raio[b], Alice Chiang[b], Elizabeth A. Phelps[a,b,c], and Nathaniel D. Daw[a,b]

(In) Relation to Memory

Hippocampal Contributions to Control: The Third Way

Máté Lengyel and Peter Dayan
{lmate,dayan}@gatsby.ucl.ac.uk