Task-space dimensions guide human exploration in complex environments

Jiahui An[1,2], Jiewen Hu[1,2], Yilin Elaine Wu[1,3], Siyu Ning[1,4], Fanyu Zhu[1,2], Yuhang Pan[1,5], Ruosi Wang[6], Ni Ji[1,2 †]

1.Chinese Institute for Brain Research, Beijing 2.Peking Union Medical College 3.Nanjing University

4.China Agricultural University 5.Zhejiang University 6.Shanghai International Studies University † niji@cibr.ac.cn

Motivation & Scientific Question

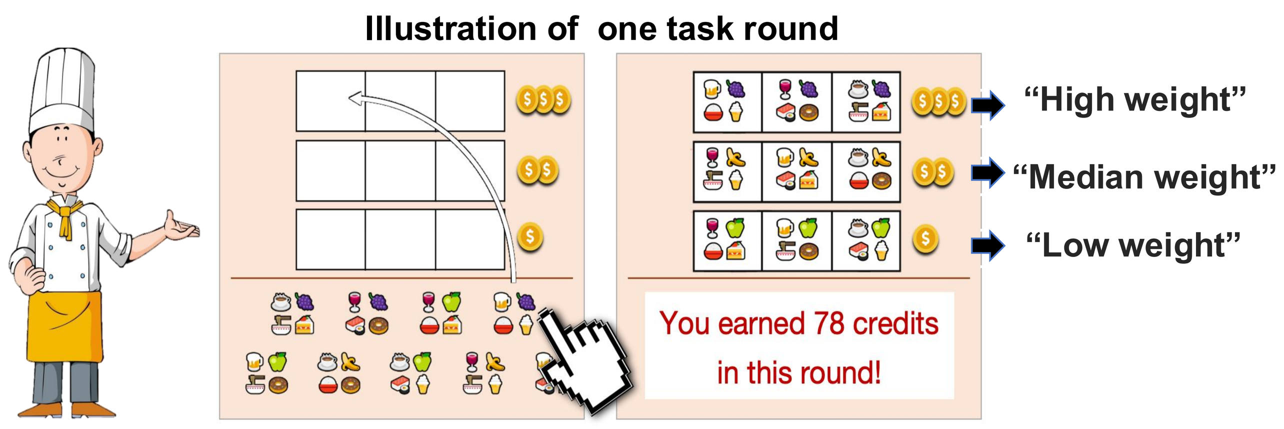
- Real-life decision-making are done in ***high-dimensional environments***
- Learning in high-D environments with ***sparse feedback*** is challenging, and requires ***efficient exploration***
- How humans explore in multi-dimensional task environments has not been systematically examined

Key question: How do humans explore efficiently in unknown multi-dimensional environments?

Task Design

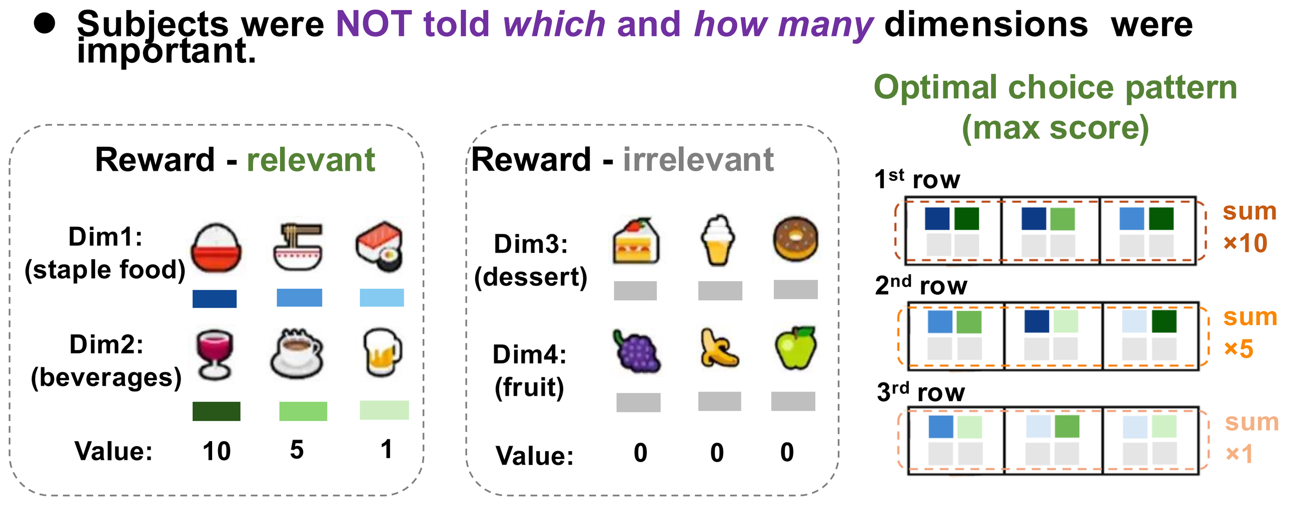
Background

"You are a chef in a restaurant. Your task is to recommend meals to the customer and infer his/ her tastes based on the rating."

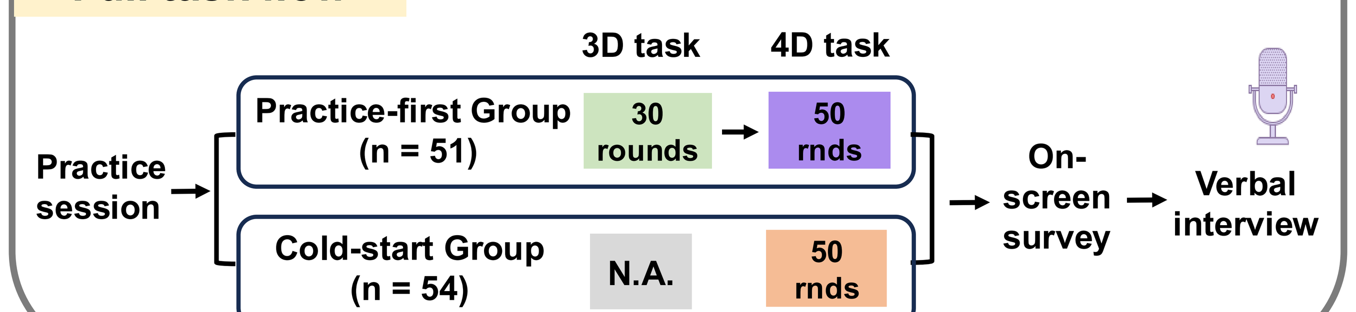


Scoring rules

- Only 2 dimensions are relevant for the score
- Subjects were NOT told *which* and *how many* dimensions were important.

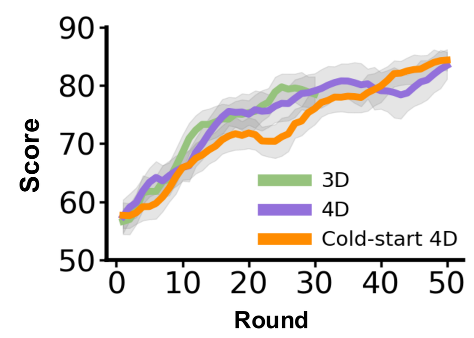


Full task flow

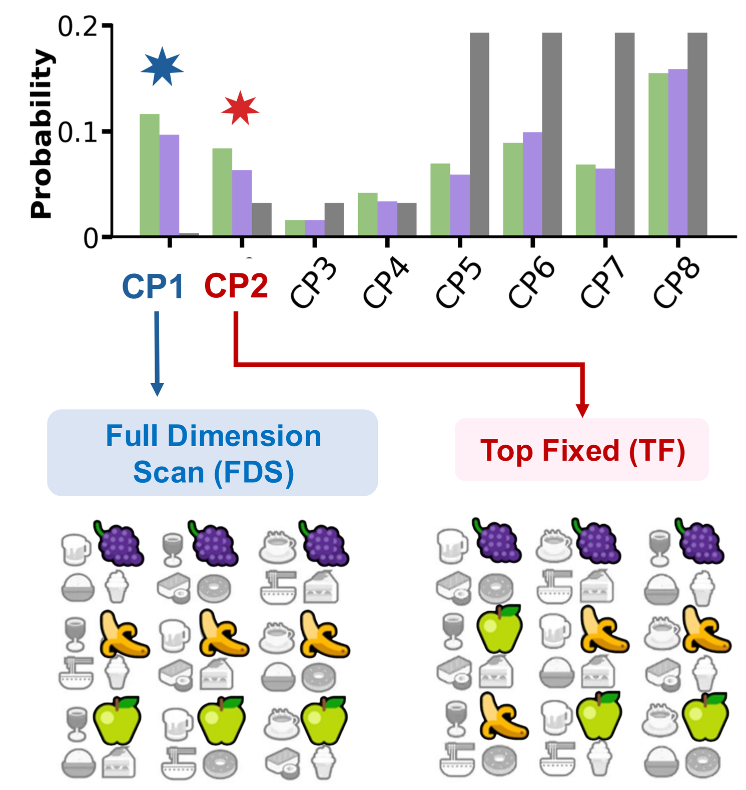


1. Humans learned efficiently and showed two *stereotyped* and *persistent* choice patterns

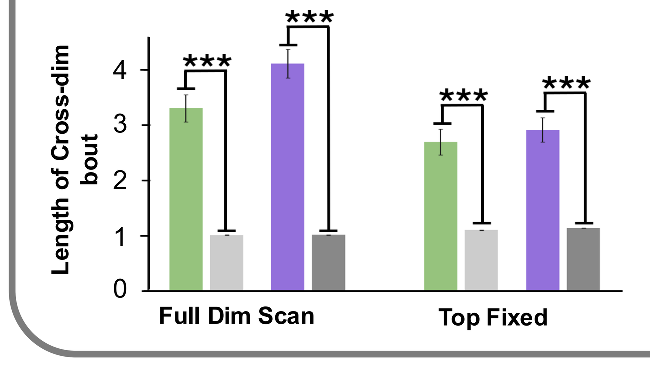
1. Subjects successfully learned to play the game



2. Two choice patterns were highly employed

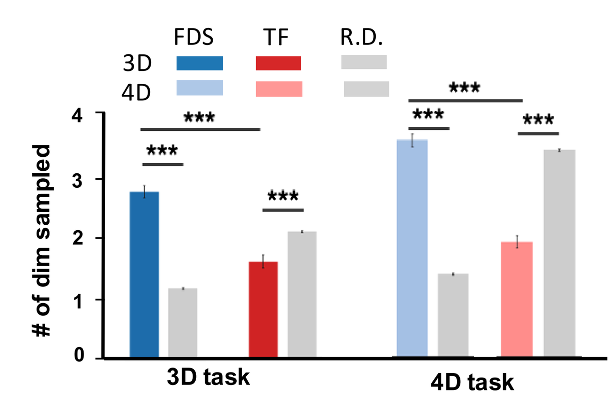


3. Both patterns were persistently employed

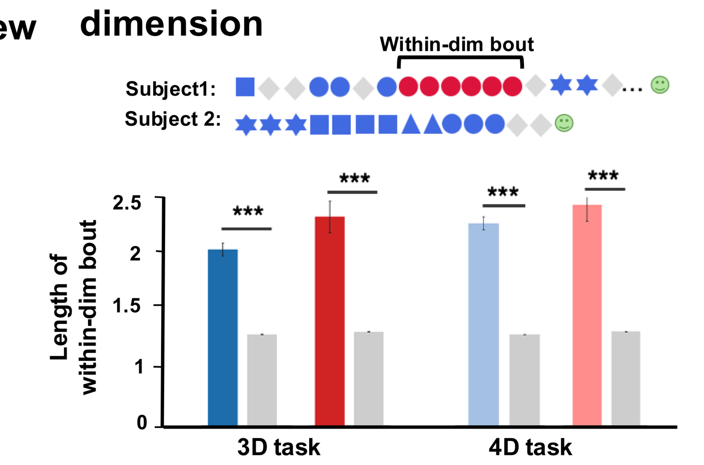


2. FDS was organized around task dimensions

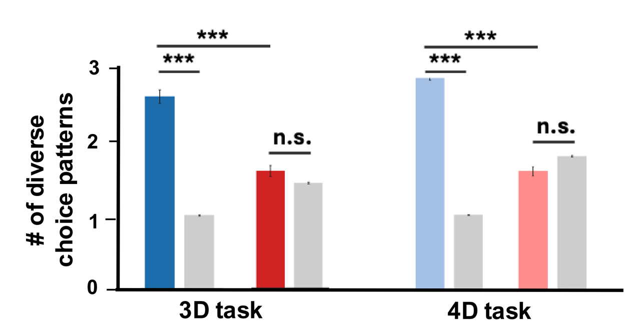
1. FDS was applied across multiple dimensions while TF limited to a few



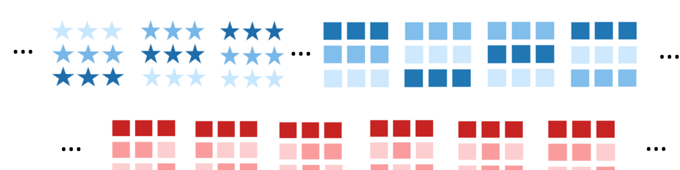
2. FDS & TF were persistent within dimension



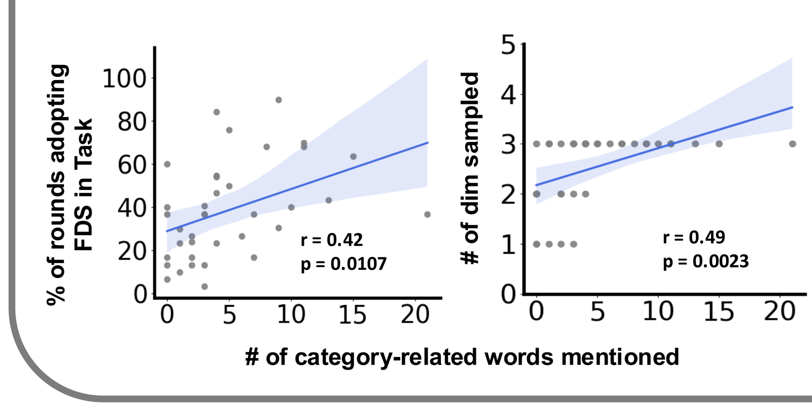
3. FDS covered more diverse choice patterns within the same dimension, whereas TF tends to repeat the same pattern



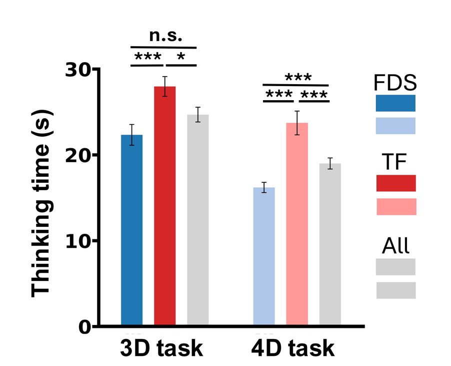
Example choice sequences employing FDS/ TF



4. Subjects who expressed more category-related words are more likely to perform FDS across dimensions.

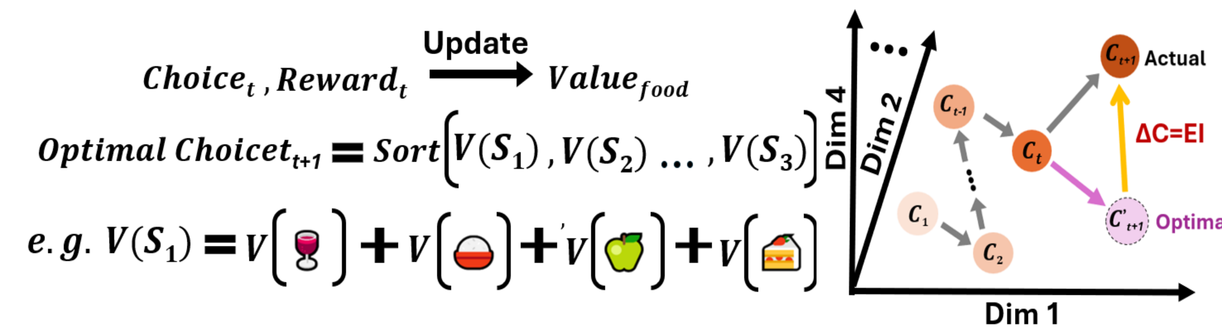


5. Subjects responded faster when using FDS strategy

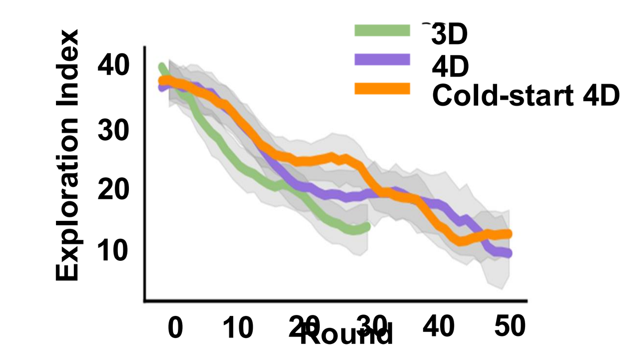


3. FDS served as an exploration strategy

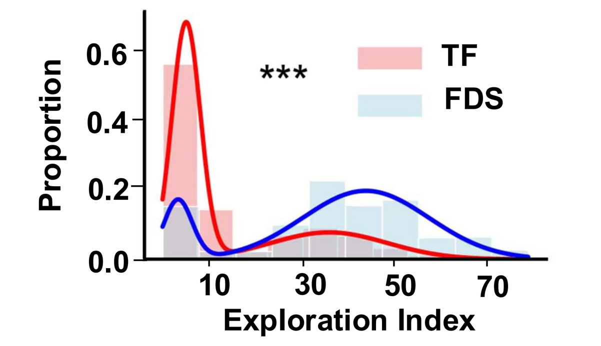
➢ Use **exploration index (EI)** to quantify explorative behavior:



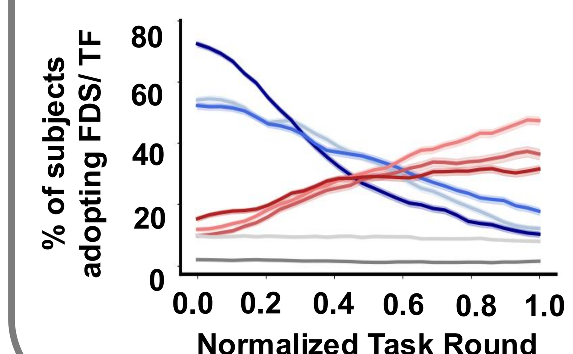
1. EI decreased as game proceeded



2. FDS is more associated with exploration than TF.



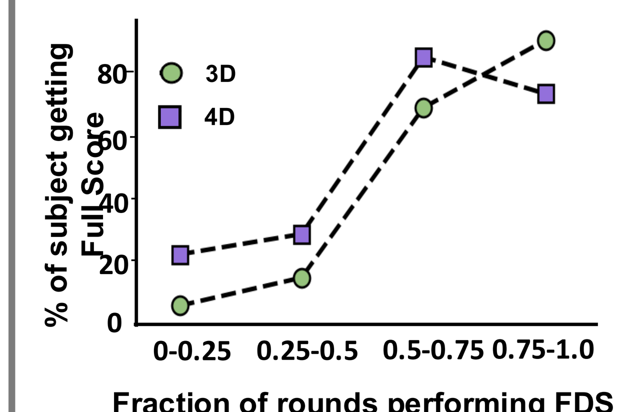
3. FDS were employed earlier in the task than TF



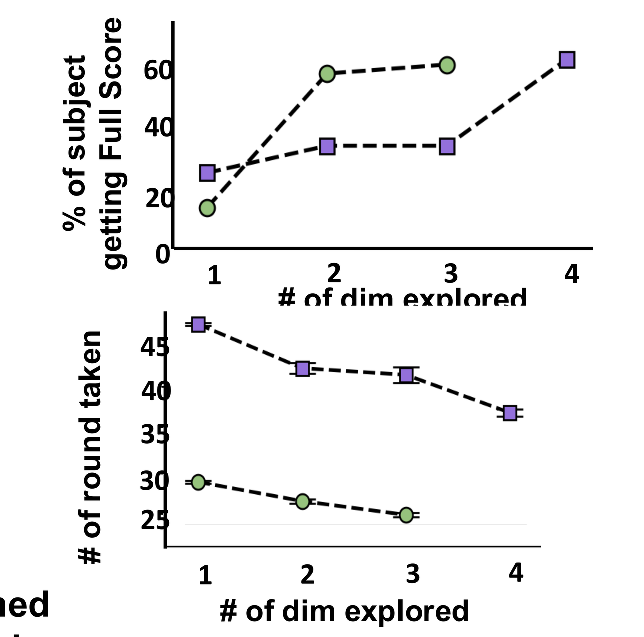
4. Subjects used more 'Attempt' when describing FDS

4. FDS may facilitate efficient learning

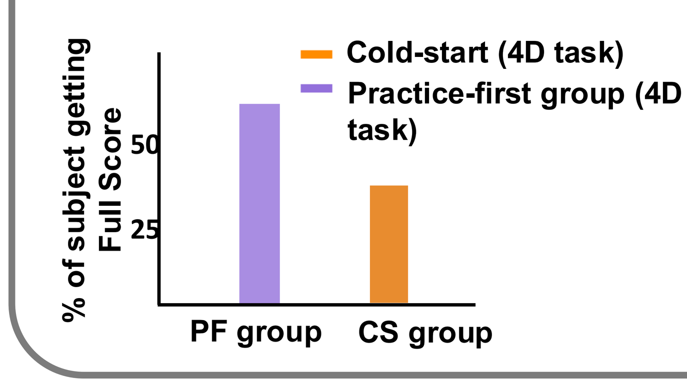
1. Subject who more frequently employed FDS performed better



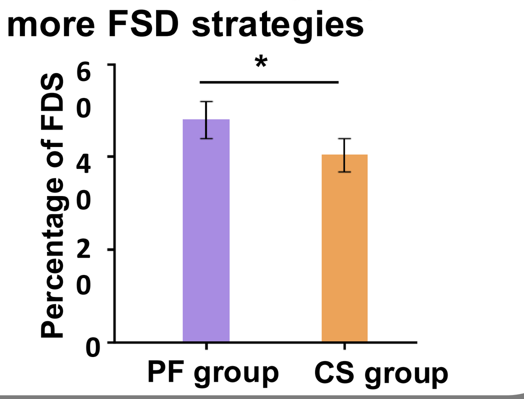
2. Subject applied FDS to more dimensions performed better



3. Practice-first group outperformed the Cold-start group in the 4D Task.



4. Practice-first group used more FSD strategies



5. Existing models failed to capture human-like exploration behavior

Adaptation of previous models (key refs: Wilson and Niv 2012, Niv et al. 2015, Song et al. 2022)

(1) Naïve RL: learns values of combined stimuli with a learning rate (η) and row weights (ω) based on their row positions $I(S)$.

$$V_{t+1}(S) = V_t(S) + \eta \cdot \omega(I(S)) \cdot \left[\frac{R_t}{9} - V_t(S)\right]$$

(2) Feature RL (fRL): updates feature values by calculating how each feature contributes to the reward across all stimuli it appears in

$$W_{t+1}(f) = W_t(f) + \sum_{f\in S} \eta \cdot \omega(I(S)) \cdot \left[\frac{R_t}{9} - V_t(S)\right]$$

The value of S is the sum of the values of its contained features:

$$V(S) = \sum_{f\in S} W(f)$$

(3) Bayesian rule learning: performs Bayesian inference on all hypotheses

$$P(h \mid c_{1:t}, r_{1:t}) \propto P(r_t \mid h, c_t) P(h \mid c_{1:t-1}, r_{1:t-1})$$

Expected reward (ER) for each choice:

$$ER(c_{t+1}) = \sum_h P(h \mid c_{1:t}, r_{1:t}) P(r_{t+1} \mid h, c_{t+1})$$

(4) Value-based Serial hypothesis (SH):

- Updates the ER of h_{t+1} and values of features under fRL framework:

$$ER(h_{t+1}) = \sum_j \sum_i V(f_{i,j})$$

$$V_{t+1}(f_{i,j}) = V_t(f_{i,j}) + \eta(r_{t+1} - ER(h_{t+1}))$$

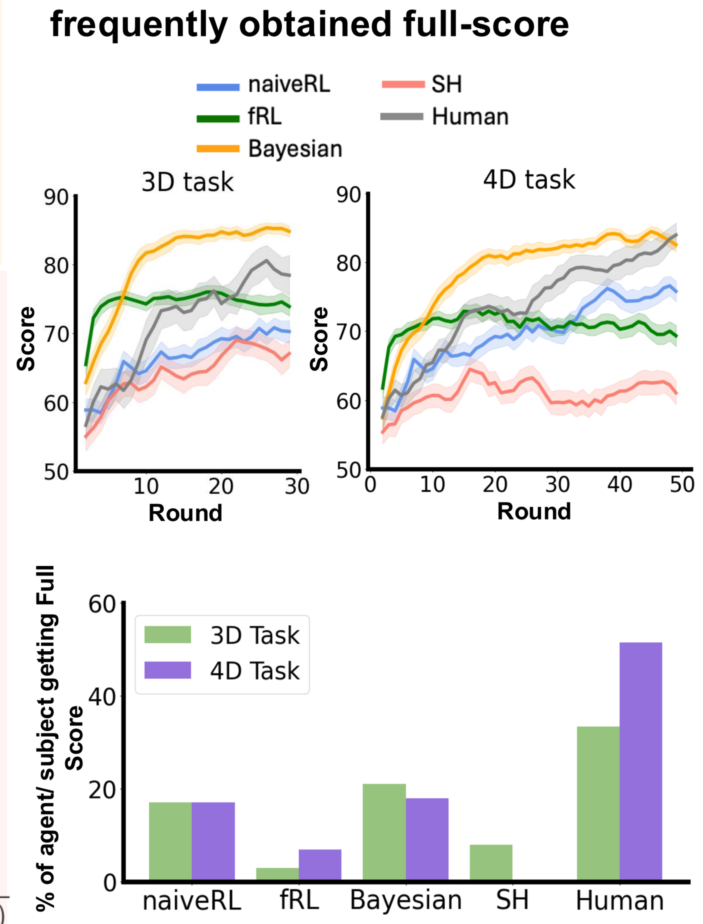
- Determine the probability of staying on the same h:

$$Pr(stay|h) = \frac{1}{1 + exp(-\beta_{stay}(log\frac{P(h)}{1-P(h)} - \theta)))}$$

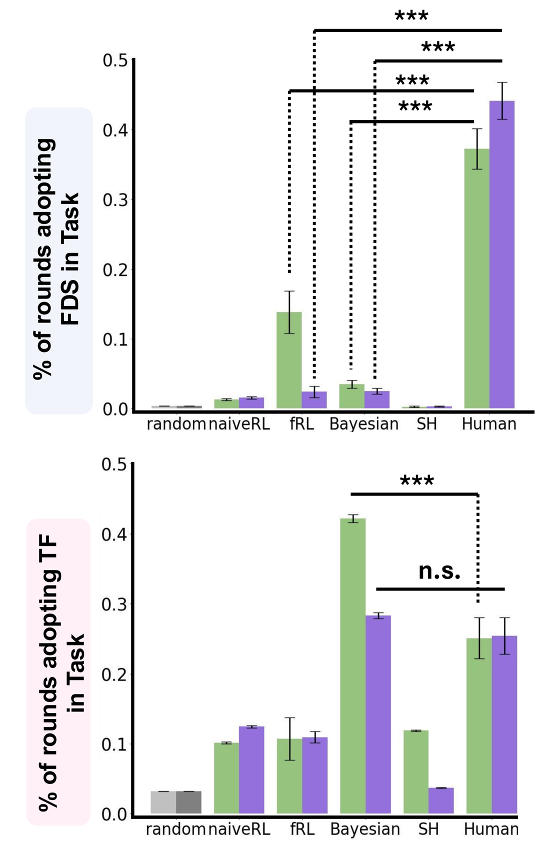
- If decide to switch to a new h, compute the probability of switch to the hypothesis h_{t+1} as:

$$P(h_{t+1}) = (1 - Pr(stay))\frac{exp(\beta_{switch} - ER(h_{t+1}))}{\sum_{h'\neq h_t} exp(\beta_{switch} - ER(h'))}$$

1. Bayesian model yielded highest performance, though humans more frequently obtained full-score

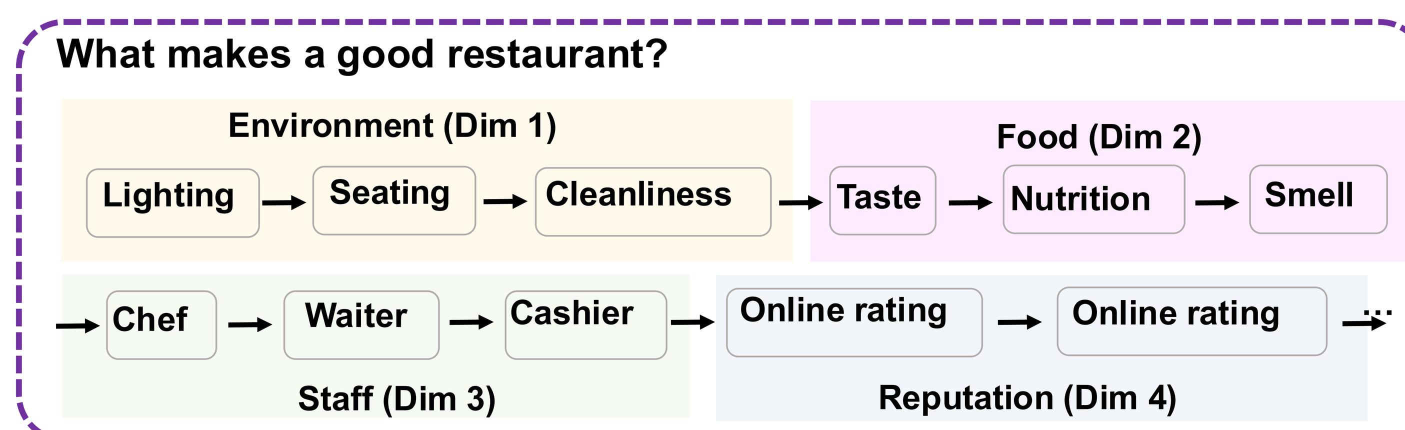


2. FDS was rarely employed by model agents, though TF was more prevalently used



Conclusion and Future Plan

- Human develop dimension-guided strategy to explore in multi-dimensional and unknown environment.
- Dimension-guided exploration facilitated efficient learning in complex environment.



Ongoing work:

- Modeling to understand the cognitive mechanism behind dimension-guided exploration
- Extending the task to high-dimensional, realistic environments