# Bayesian Attribution of Incentives Predicts Action-Induced Preference Changes

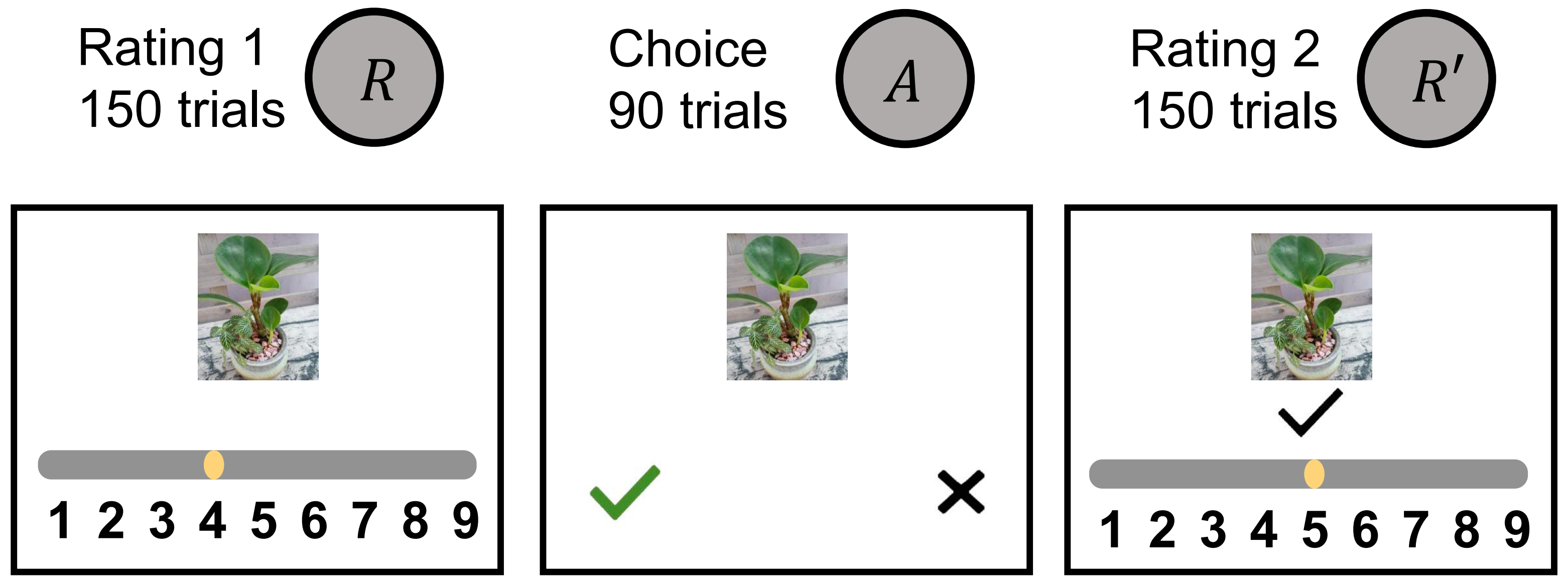
Guihua Yu[1*], Yaomin Jiang[1*], Lusha Zhu[1]

[1] Peking University, Beijing, 100871, China, [*] These authors contribute equally

## Introduction

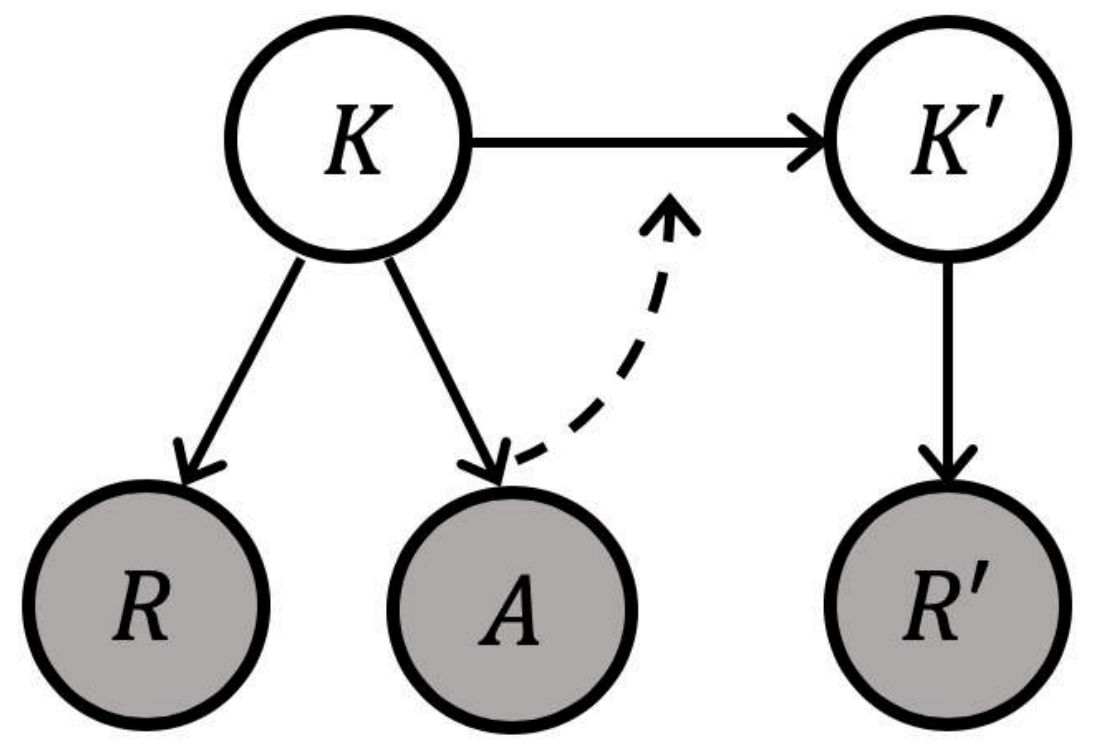
Classical decision theories usually assume that choices reflect, but do not modify, preferences[1-2]. By contrast, extant data from behavioral and neuroimaging studies[3] have suggested that humans sometimes change their preferences of certain things in a seemingly irrational manner, after making decisions about the stimuli—with when, how, and why such action-induced preference changes occur largely unknown. Here, we sought to establish a computational account for predicting such preferences changes in a classic decision making setup widely used for evaluating internal valuations underlying goal-directed behavior.

## Experimental design


Rating 1
150 trials
Choice
90 trials
Rating 2
150 trials

Participants encountered a sequence of unfamiliar pictures and needed to rate these pictures before and after making decisions about whether or not to bring home a particular picture printed on a postcard.

## Bayesian modeling



### Internal evaluation

$$P(K_{ij}|A) = \frac{P(A|K_{ij})P(K_{ij})}{\sum_{K_{ij}} P(A|K_{ij})P(K_{ij})}$$

### Rating-preference transition

$$P(K_{i,j}|R_{i,j}) = \frac{P(R_{i,j}|K_{i,j})P(K_{i,j})}{\sum_{K_{i,j}=1}^{9} P(R_{i,j}|K_{i,j})P(K_{i,j})}$$

### Assumptions

$K_{i,j} \in \{1,2,3,\dots,9\}$

$P(K_{i,j}) \propto N(\mu_\mathrm{i}, \sigma_\mathrm{i})$

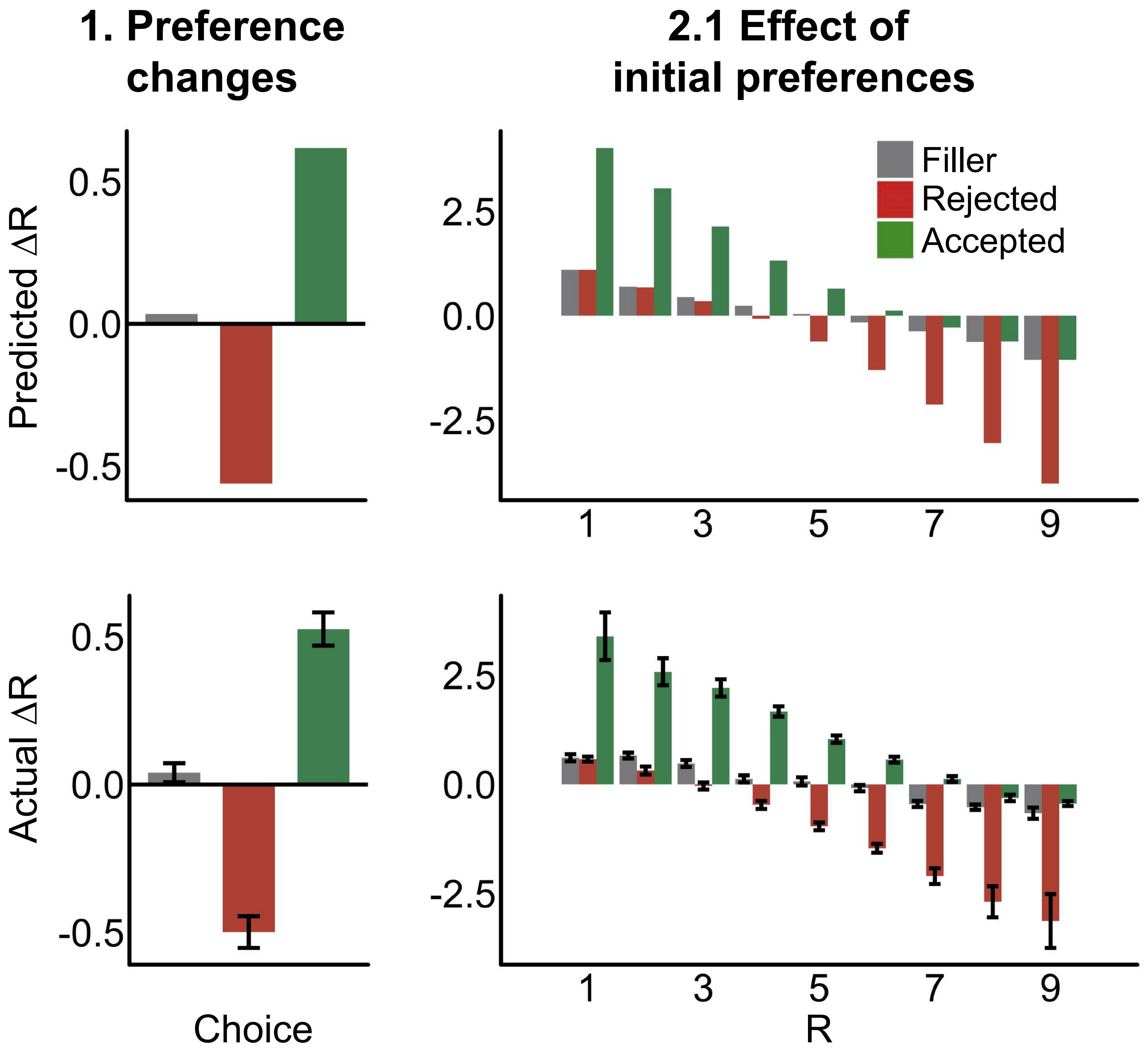
$P(R_{i,j}|K_{i,j}) \propto N(K_\mathrm{i}, \delta_\mathrm{i})$

$P(A_{i,j}|K_{i,j}) = 1/(1 + e^{-\alpha_i+\beta_i K_{i,j}})$

## Results

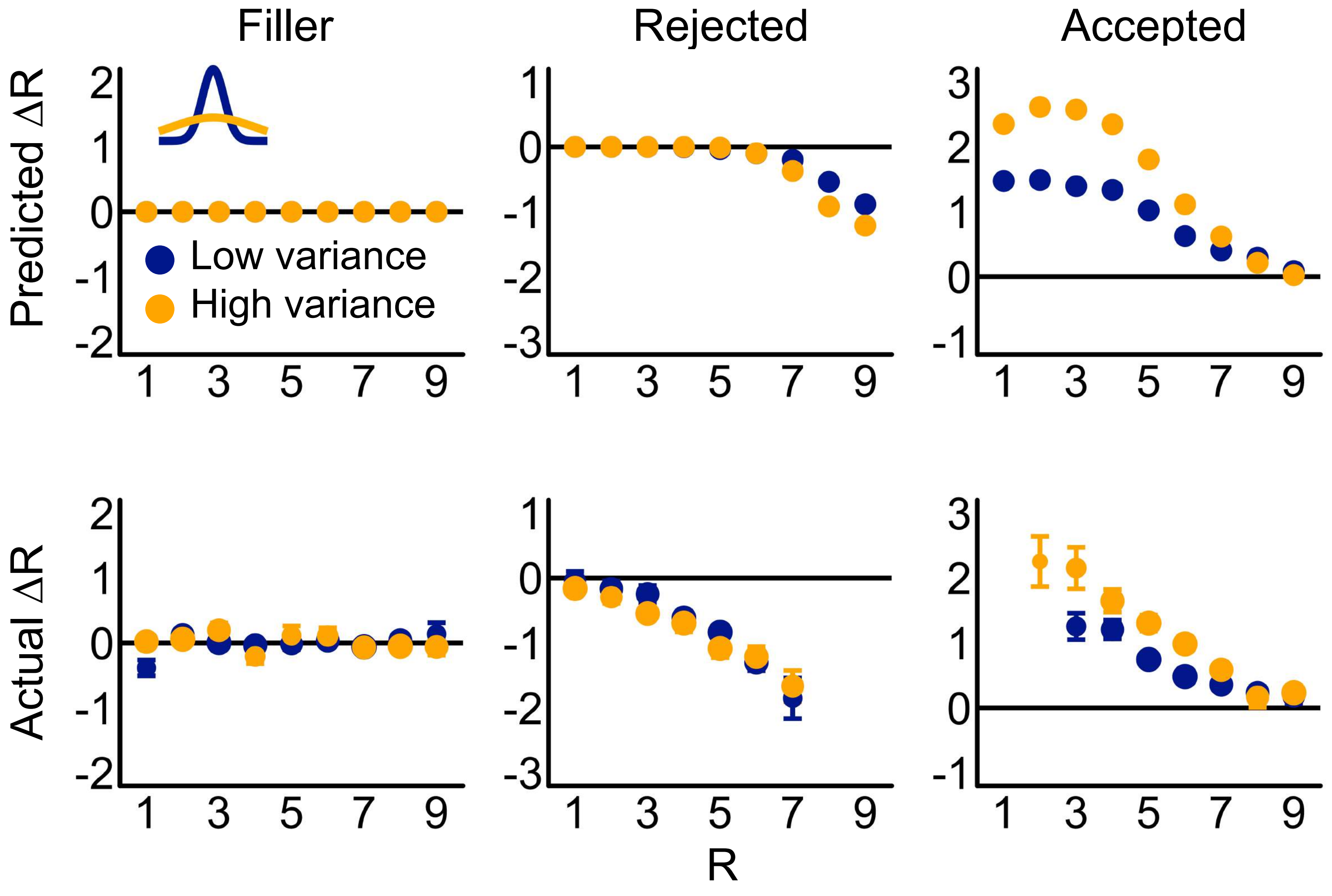
### Overview

1. Action-induced preference changes.
2. Model-derived predictions on when and how such effects occur.
3. Alternative explanations.

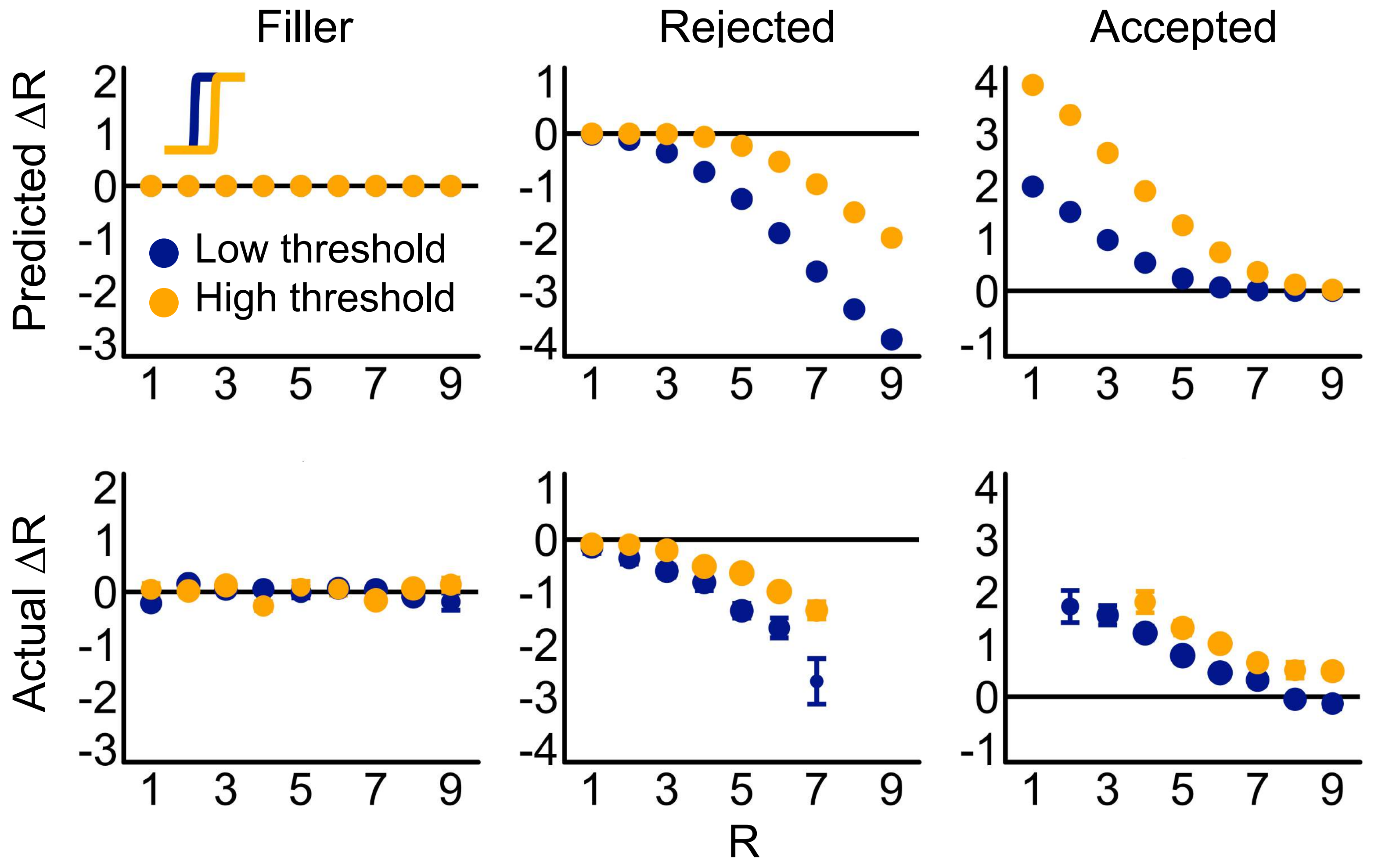

1. Preference changes

2.1 Effect of initial preferences

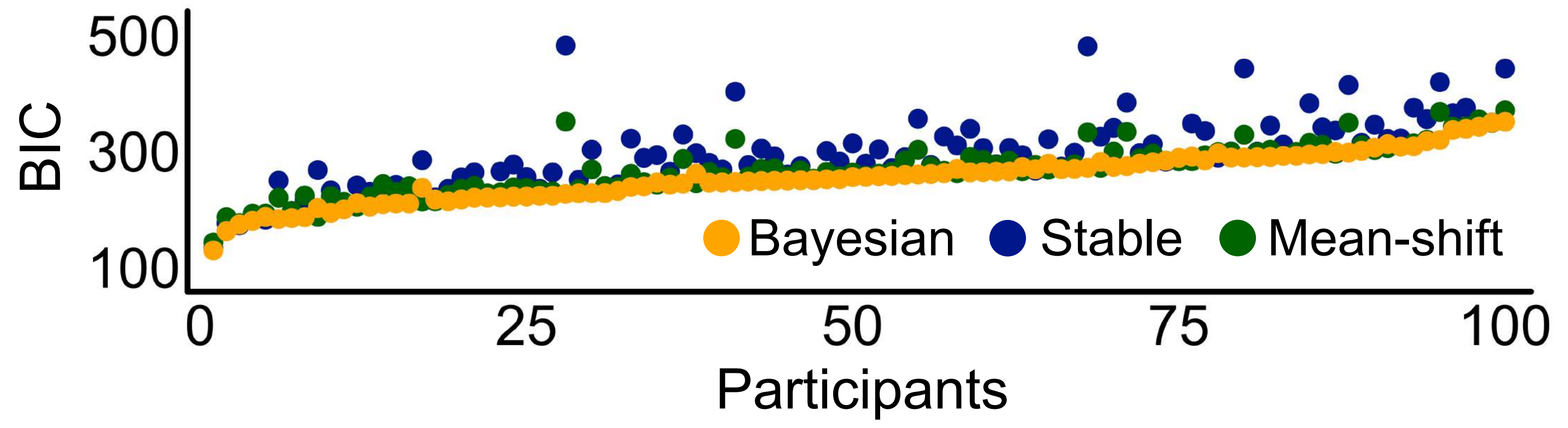
### 2.2 Effect of the level of uncertainty in initial preferences



### 2.3 Effect of the choice sensitivity on preferences



### 3. Bayesian attribution vs. alternatives



## Discussion

Our results, although preliminary, hint on a possible mechanism by which internal incentives are appraised flexibly and dynamically, reflecting subtle differences in a decision maker's experience and cognitive states. Next, we will experimentally manipulate the key modulators identified by our model (e.g., initial preferences, uncertainty, choice sensitivity, etc.) and test for model predictions; we will also investigate at the neural level how past decisions are retrieved and evaluated in the brain and how such processing interacts with value signals in service of internal reward assessment.

## Reference

1. Samuelson,P., *Economica*.(1938)
2. Ariely, D. & Norton,M., *Trends Cogn Sci*.(2008)
3. Izuma,K., & Murayama,K., *Front Psychol*.(2013)

This work was supported by CNSF 31671171.

**Contact**: guihuayu@pku.edu.cn, yaominjiang@pku.edu.cn, or lushazhu@pku.edu.cn.