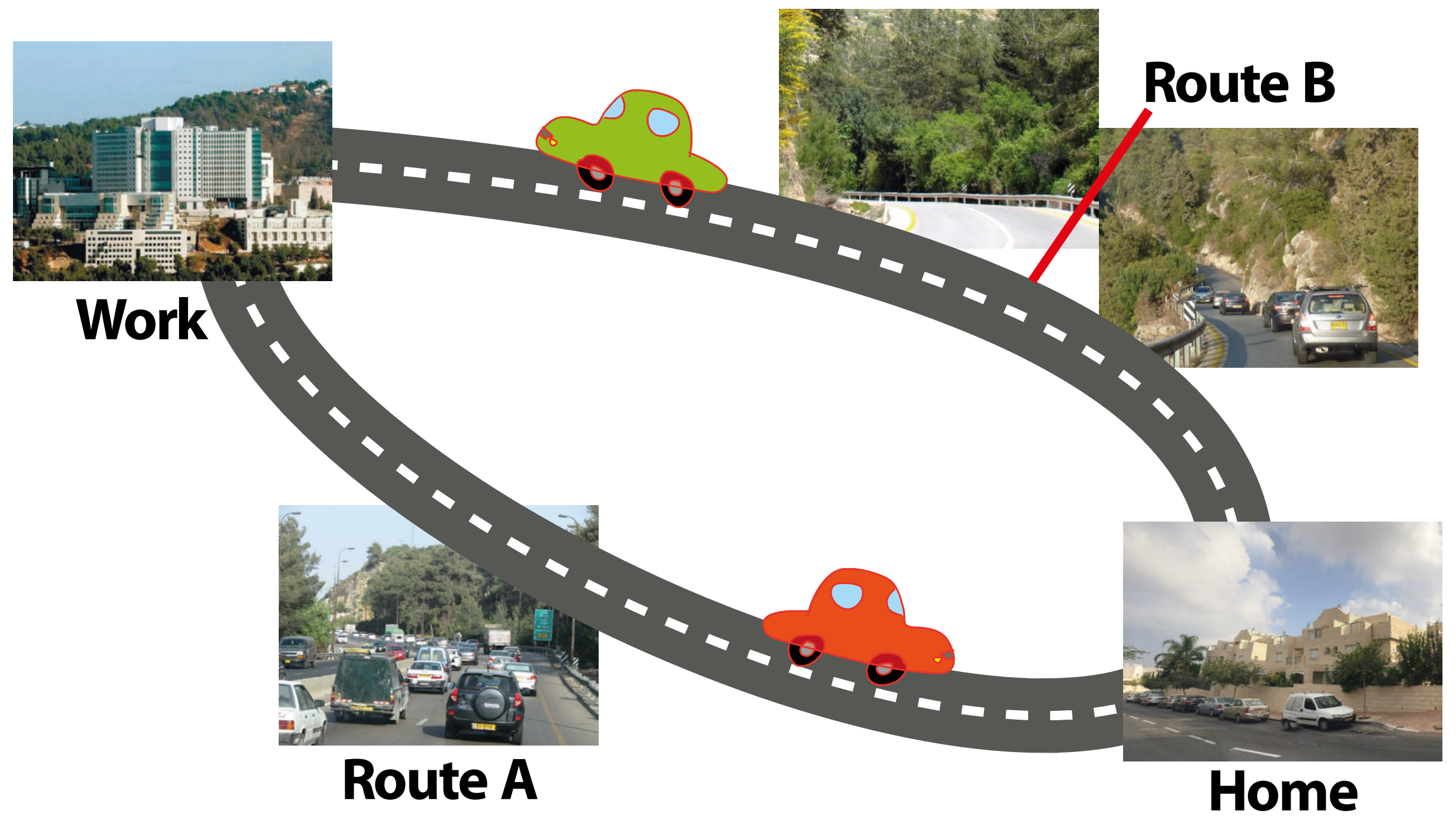
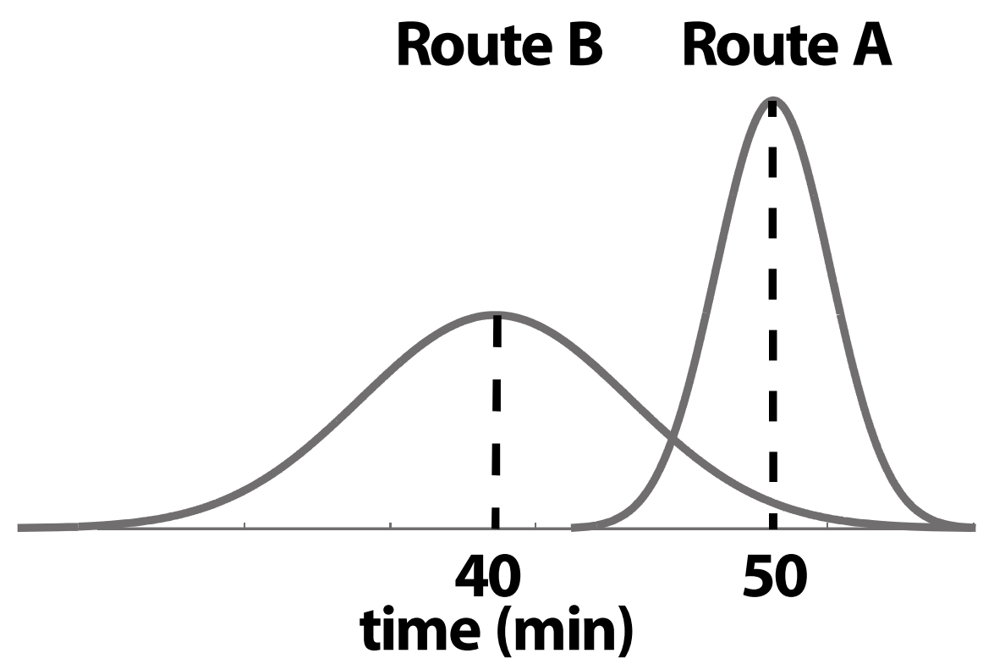
Metacognition and Variance in a Two Armed Bandit Task

Uri Hertz[1], Mehdi Keramati[2], Bahador Bahrami[1]
u.hertz@ucl.ac.uk
1. UCL Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK
2. The Gatsby ComputationalNeuroscience Unit, University College London, WC1N 3BG London, UK

Introduction

You have to get from home to work, and have two routes to choose from. One is constantly jammed, and takes 50 minutes (Route A). The other is free most of the time and takes 35 minutes, but sometimes other cars use it as well and can jam it for more than an hour (Route B). Every morning you have to decide which route to try today.

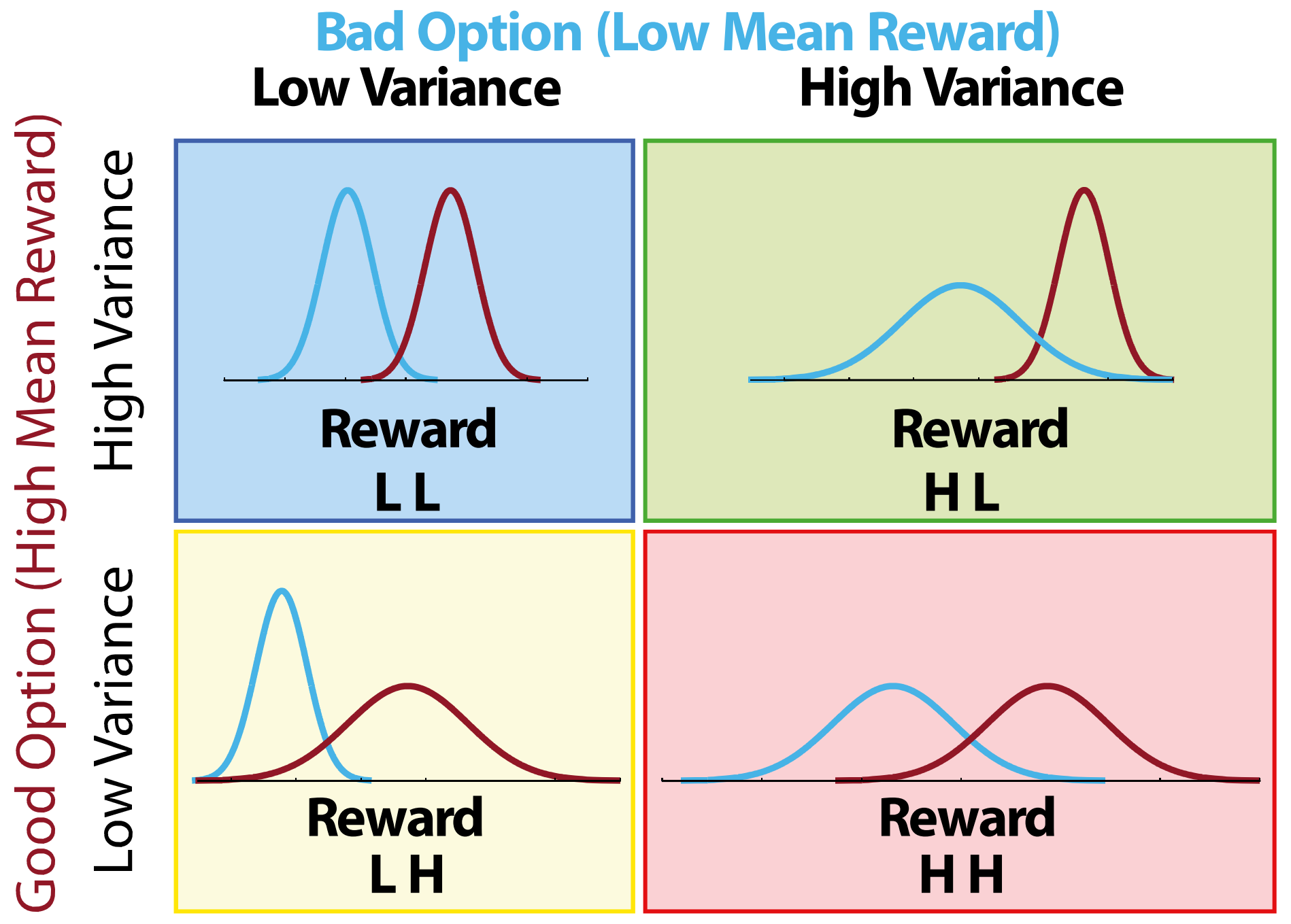


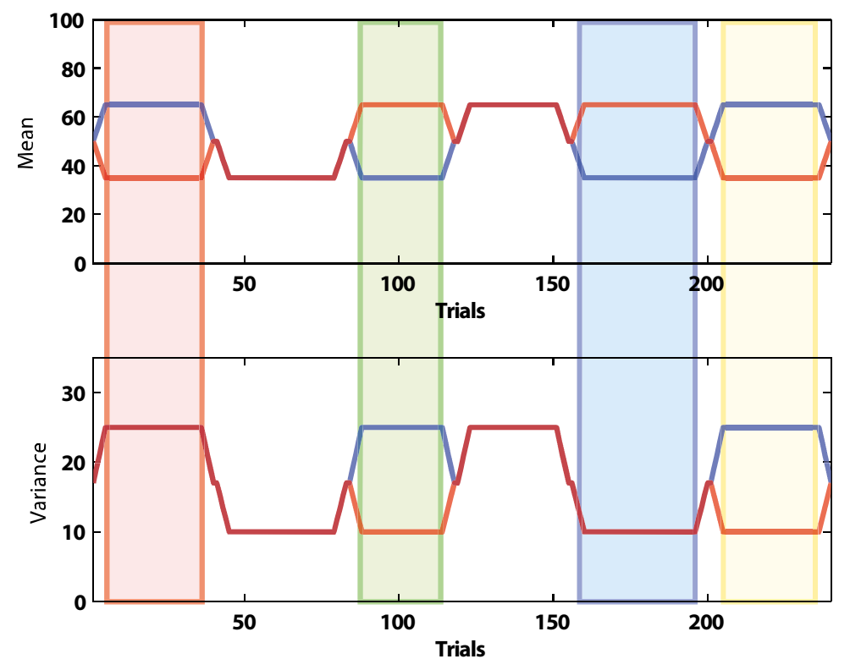


We ask:
1. How does variance of choices' outcomes affect our decisions?
2. Does it affect metacognition, e.g. our confidence in our choices?

The Task

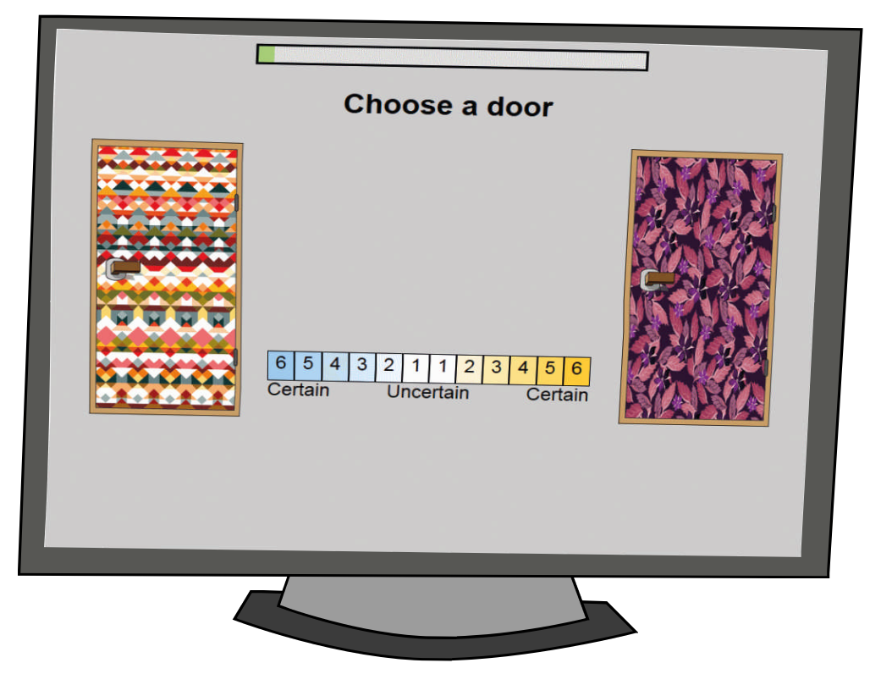
We examined choices and confidence ratings made by participants in a two armed bandit task. Four stable experimental conditions were embedded in a continuous design. In all conditions one option had higher mean rewards. The variance of the reward distributions changed across conditions and could be high (H) or low (L).



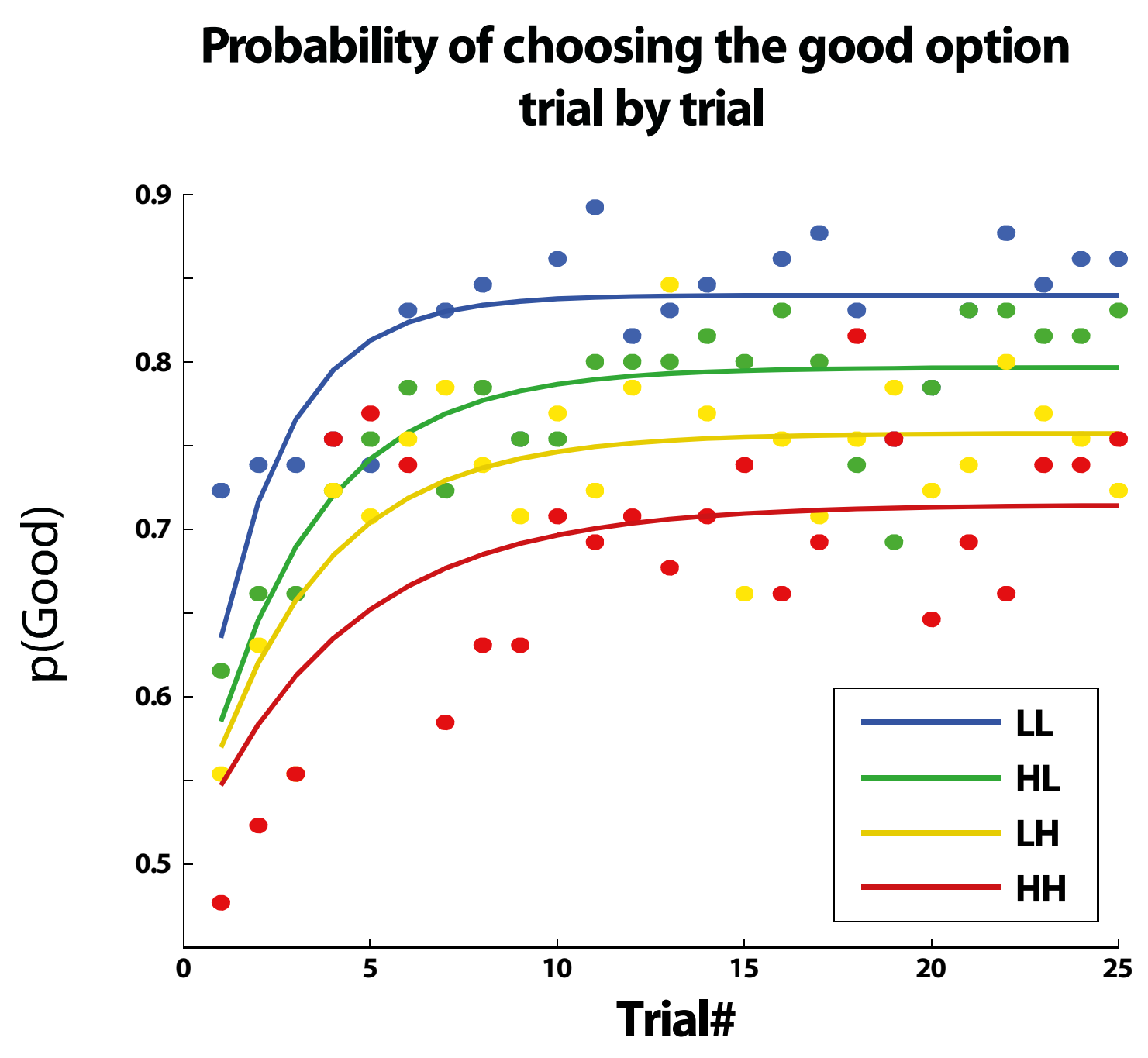


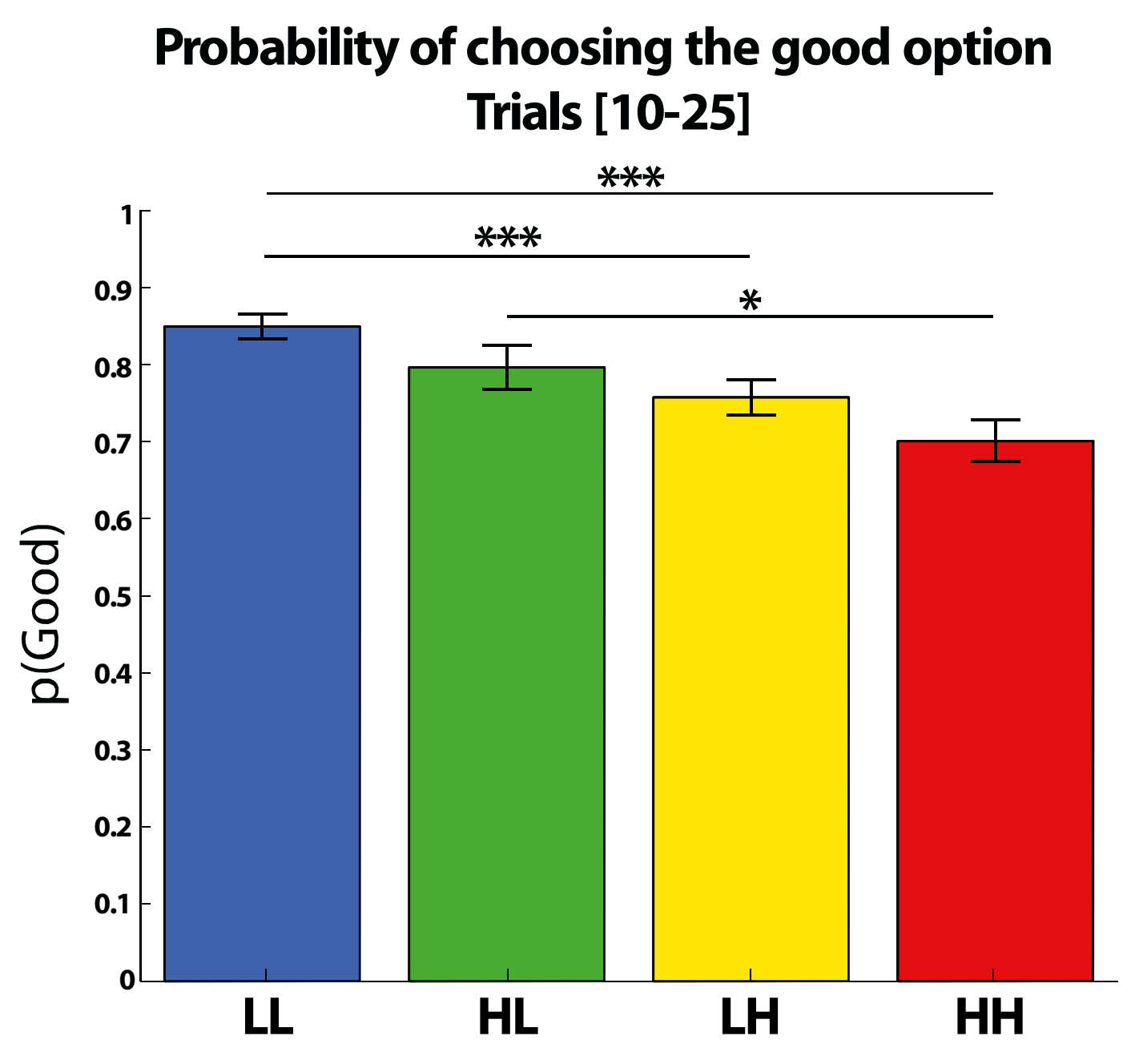
The experiment lasted 240 trials, with 6 periods of stable rewards' distribution parameters, 4 of which were the conditions described above (3 different designs were used). In each trial participants had to choose between two doors (doors location varied between trials). Choice was made using a confidence scale (-6 to 6).

Eighty participants were recruited using Amazon M-Turk, from which 15 were excluded.



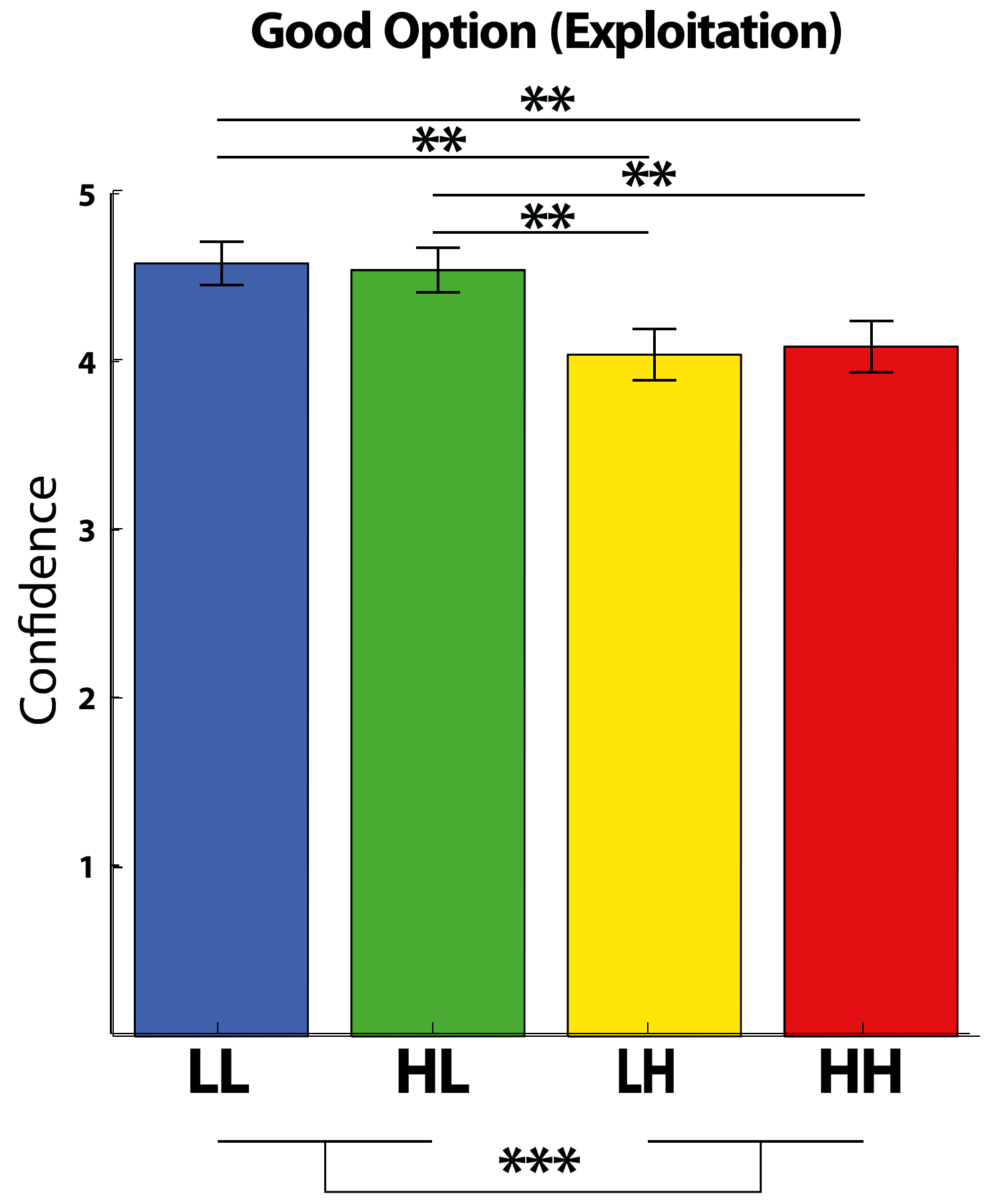
Results

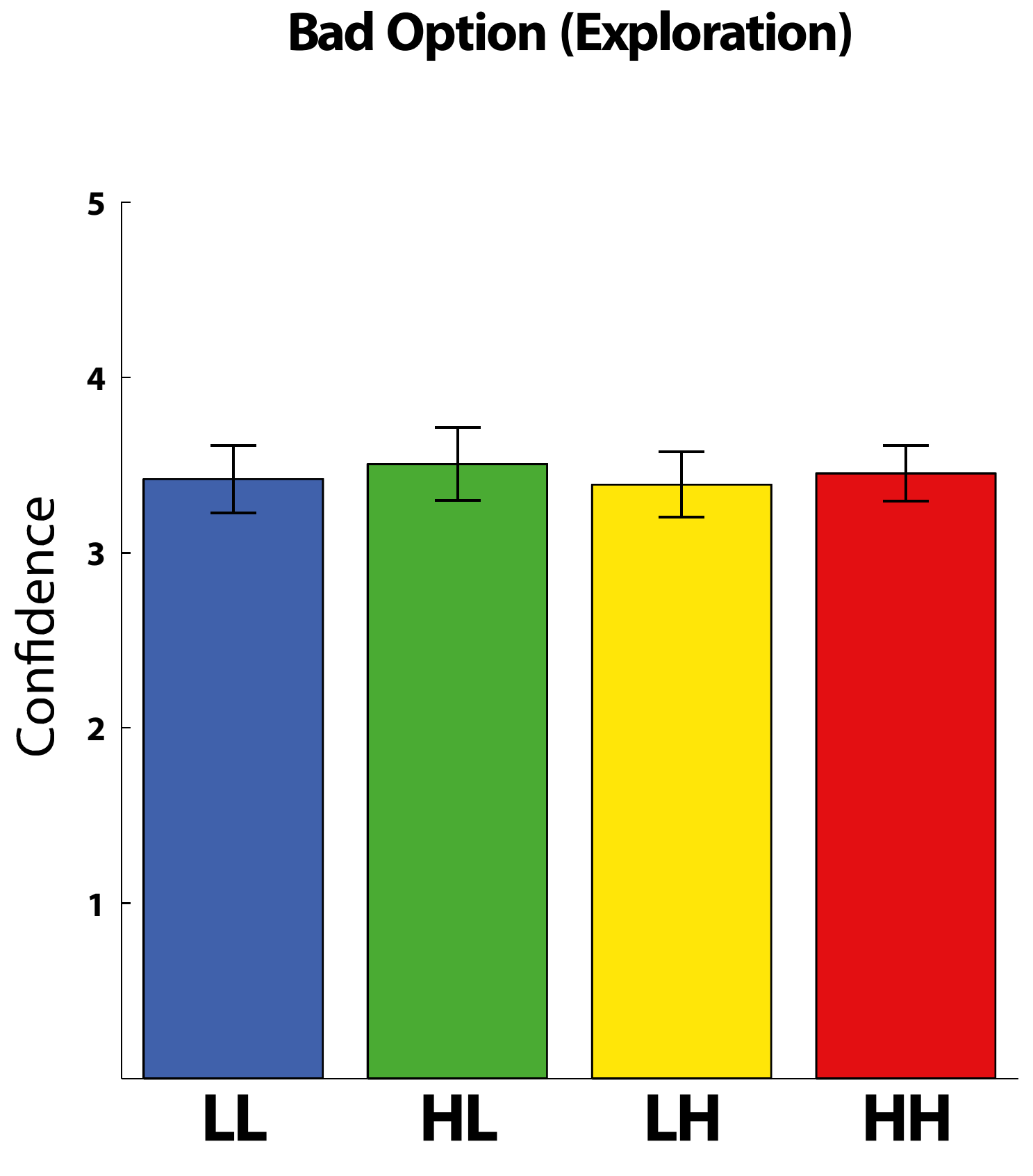




Probability of Choosing the Good Option

Probability of choosing the good option in trials 10-25 of each block was calculated for each participant. These probabilities were analysed using a mixed effect ANOVA, with variance of good option and variance of bad option as between subjects factors. Both variance factors were significant (Good Option Variance: F(1,259) = 22.24, p = 1e-05 , Bad Option Variance: F(1,259) = 5.2, p = 0.026).
(* p < 0.01, *** p < 0.0001)





Confidence Ratings

We examined the confidence ratings during exploration (choosing the bad option) and exploitation (choosing the good option) separately. Confidence ratings were averaged between trials 10 and 25 of the four experimental conditions. Exploration confidence ratings were overall lower than exploitation ratings (paired t-test t(64) = 8.3, p = 9e-12).
(** p<0.005, *** p<1e-06)

Models

Means Tracking ●

It could be that sampling alone (e.g. less sampling of high variance options or low mean options) led to the observed reduction in probability of choice. We used a time difference model which tracks only the mean rewards:

$$\begin{cases} Q_a(t+1) = Q_a(t) + \alpha \cdot \left(R(t) - Q_a(t)\right) \\ Q_b(t+1) = Q_b(t) \end{cases}$$

$$p(a) = \frac{\exp\left(\beta \cdot Q_a(t)\right)}{\exp\left(\beta \cdot Q_a(t)\right) + \exp\left(\ \beta Q_b(t)\right)}$$

Beta Modulation ●

As probability of choosing the good option seem to declince as the variance of rewards increased , we used another model that tracks both variance and mean of rewards, and uses the mean variances to modulate β.

$$\begin{cases} Q_a(t+1) = Q_a(t) + \alpha \cdot \left(R(t) - Q_a(t)\right) \\ Q_b(t+1) = Q_b(t) \end{cases}$$

$$\begin{cases} V_a(t+1) = V_a(t) + \gamma \cdot \left(\left(R(t) - Q_a(t)\right)^2 - V_a(t)\right) \\ V_b(t+1) = V_b(t) \end{cases}$$

$$\beta(t) = \frac{\beta_0}{1 + C \cdot \left(V_a(t) + V_b(t)\right)/2}$$

$$p(a) = \frac{\exp\left(\beta(t) \cdot Q_a(t)\right)}{\exp\left(\beta(t) \cdot Q_a(t)\right) + \exp\left(\beta(t) \cdot Q_b(t)\right)}$$

Threshold Comparison ●

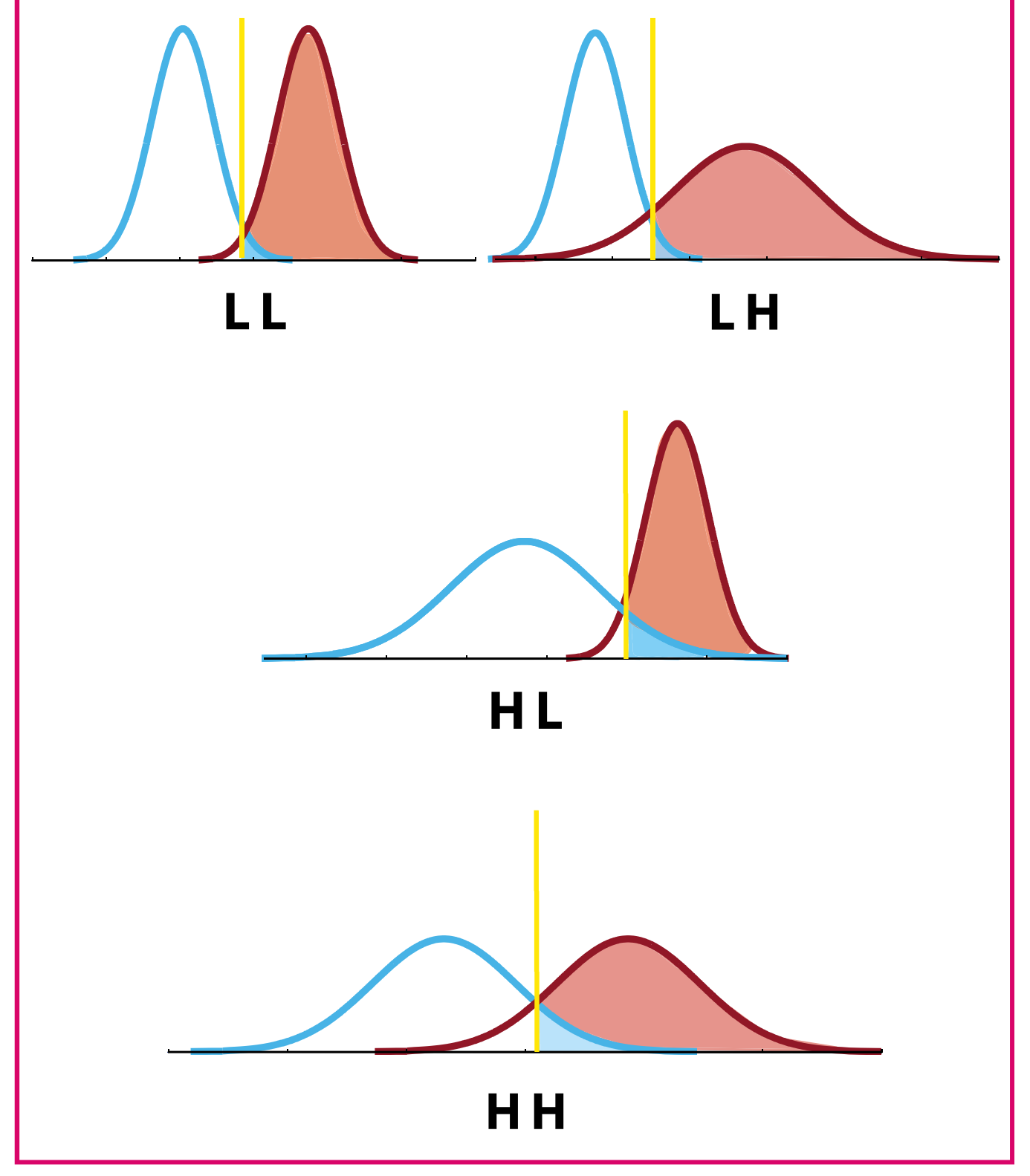
In this model the participant compares the probability of two options to be higher than a threshold.

$$\begin{cases} Q_a(t+1) = Q_a(t) + \alpha \cdot \left(R(t) - Q_a(t)\right) \\ Q_b(t+1) = Q_b(t) \end{cases}$$

$$\begin{cases} V_a(t+1) = V_a(t) + \gamma \cdot \left(\left(R(t) - Q_a(t)\right)^2 - V_a(t)\right) \\ V_b(t+1) = V_b(t) \end{cases}$$

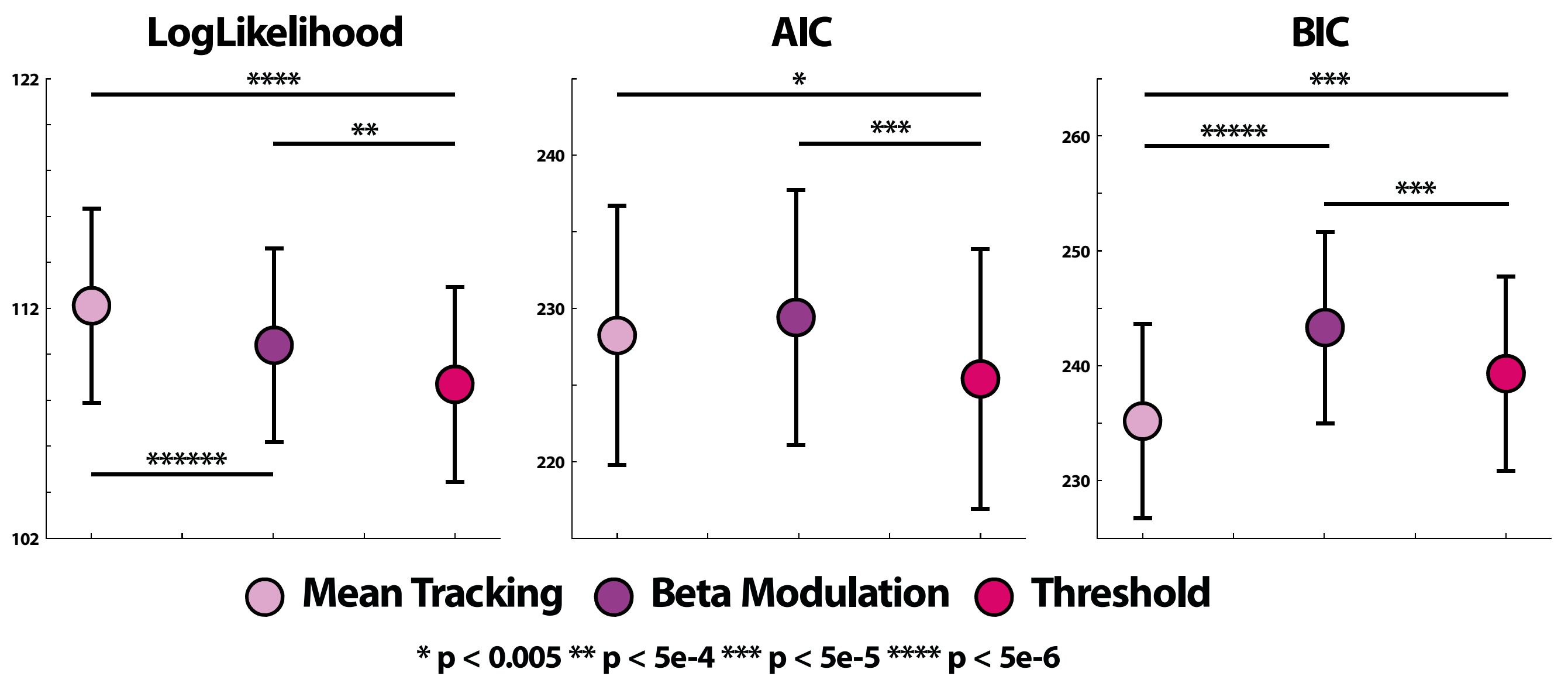
$$T_a(t) = \frac{1}{\sqrt{2\pi \cdot V_a(t)}} \int_{Threshold}^{\infty} \exp\left(-\frac{\left(x - Q_a(t)\right)^2}{2 \cdot V_a(t)}\right) dx$$

$$p(a) = \frac{\exp\left(\beta \cdot T_a(t)\right)}{\exp\left(\beta \cdot T_a(t)\right) + \exp\left(\ \beta T_b(t)\right)}$$



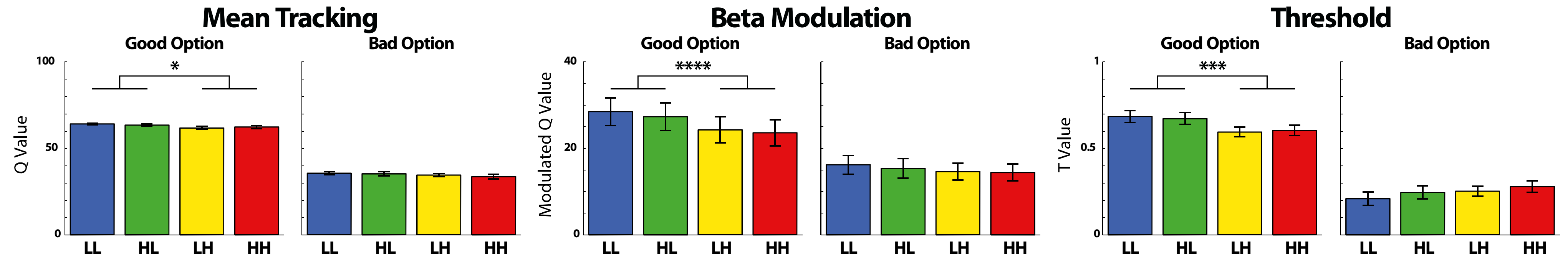
Fitting Models to Choices

We fitted the three models to the individual participants' choices to obtain individual log likelihood estimates.

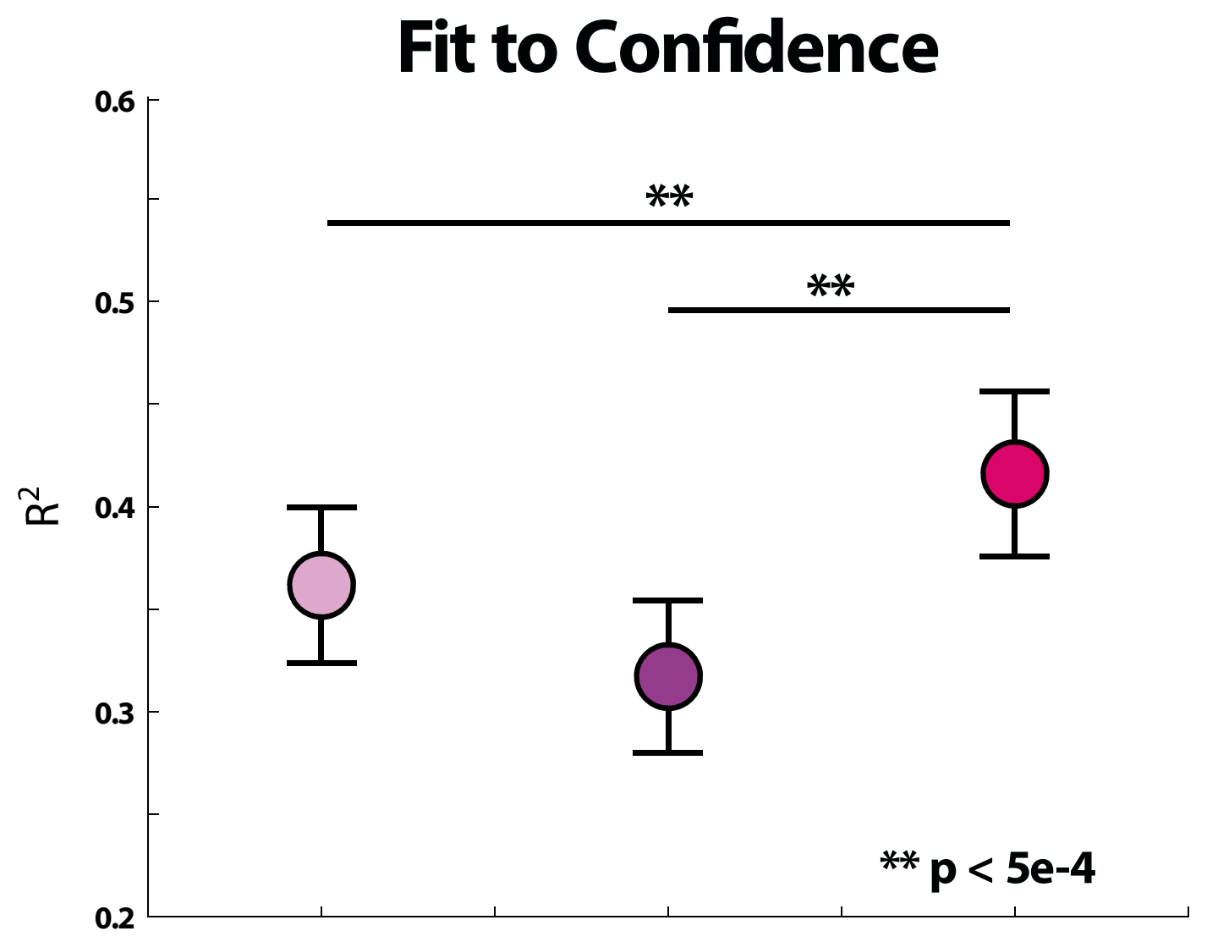


Models Estimates and Confidence Ratings

We examined if the values assigned by each model to the different options corresponds to the confidence ratings. These values were the Q values from the Means Tracking model, modulated Q values from the Beta Modulation model and T values from the Threshold Comparison model.



In order to examine the differences in predicting confidence ratings each participants mean confidence in all conditions (good option and bad option) was regressed against the model's estimatedvalues. The goodness of fit (R^2) for participants was compared across models.



Summary

Participants chose the good option less frequently as the variance of options increased.
Metacognitive reports showed dependency on the variance of the chosen option only during exploitation.
These observations can be explained if confidence ratings reflect the probability of our choice outcome being higher than a threshold.