nature neuroscience

Learning the value of information in an uncertain world

Timothy E J Behrens[1,2], Mark W Woolrich[1], Mark E Walton[2] & Matthew F S Rushworth[1,2]

2020.05.11

Heesun Park

Introduction

- We make decisions based on the outcomes of similar decisions in the past.
- Reinforcement Learning Model:

prediction error

$$Q(a_t) \leftarrow Q(a_t) + \alpha\,(r_{t+1} - Q(a_t))\,.$$

learning rate

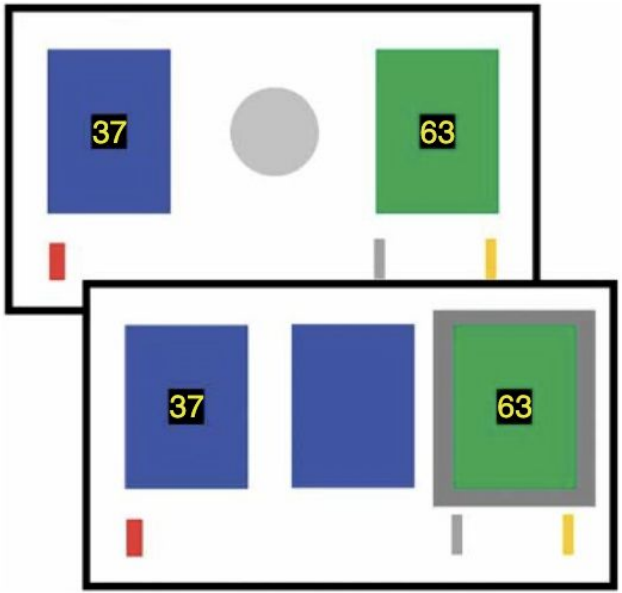
- Goal: maximization of the power to predict future outcomes

- Bayesian accounts of RL
  - Learning rate should depend on the uncertainty
  - Volatile environment: **recent experience**>distant experience → ***large α***
  - Stable environment: salient historical information → ***small α***

Methods

Task

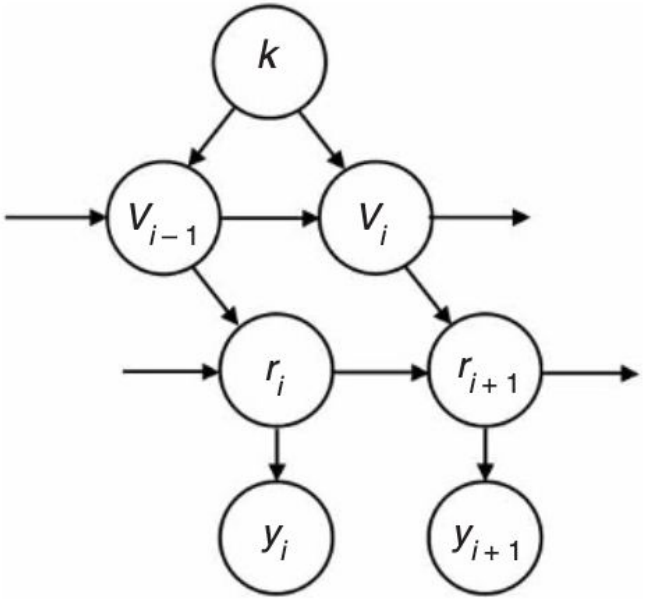
- One-armed bandit task
- Experiment structure (behav, first exp)
    - Stable environment (120 trial)
        - the probability of a blue being a correct color was 75%
    - Volatile environment (170 trial)
        - 80% blue and 80% green every 30 or 40 trials
    - 9: stable → volatile
    - 9: volatile → stable



Methods - First experiment

Behavior - Bayesian Learner

- The Optimal Bayesian Learner
- k: distrust in the constancy of the volatility
- v: volatility
- r: reward probability
- y: data



Behavior - Delta rule model

- **Predictor** - estimates **the current reward rate** given **past observations**

$$\hat{r}_{i+1} = \hat{r}_i + \alpha\varepsilon_i$$

- **Selector** - generates **actions** based on the estimates of predictor

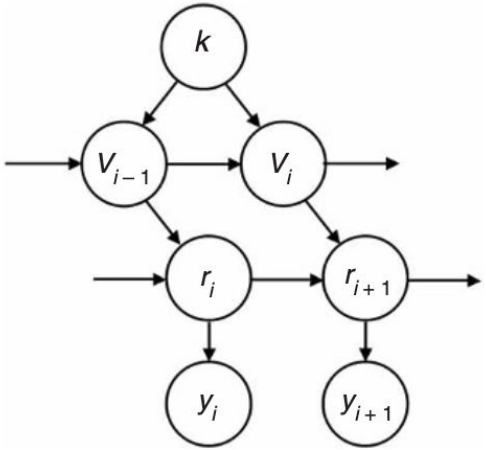
$$g_{\text{blue}\,i+1} = \hat{r}_{i+1} f_{\text{blue}\,i+1}$$
$$g_{\text{green}\,i+1} = (1 - \hat{r}_{i+1}) f_{\text{green}\,i+1}$$

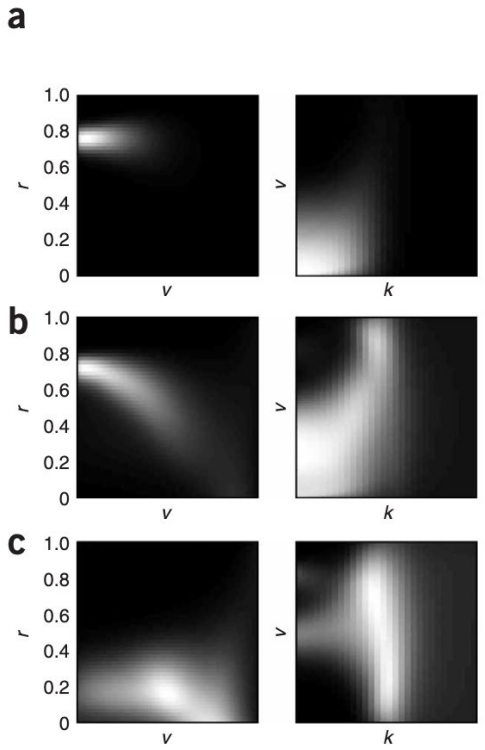
$$g_{\text{blue}\,i+1} = F(\hat{r}_{i+1}, \gamma) f_{\text{blue}\,i+1}$$
$$g_{\text{green}\,i+1} = F(1 - \hat{r}_{i+1}, \gamma) f_{\text{green}\,i+1}$$
$$F(r, \gamma) = \max[\min[(\gamma(r - 0.5) + 0.5), 1], 0]$$

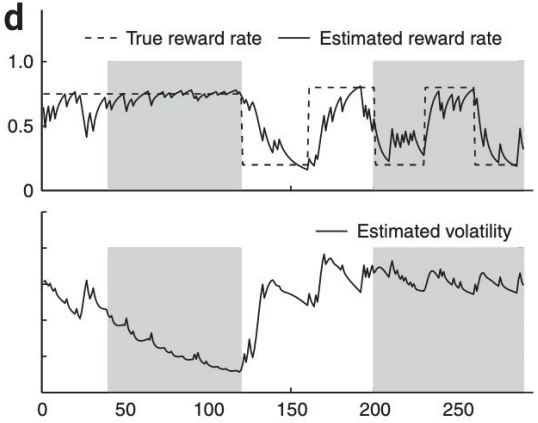
$$P(C = \text{Green}) = \frac{1}{1 + \exp(-\beta(g_{\text{green}} - g_{\text{blue}}))}$$

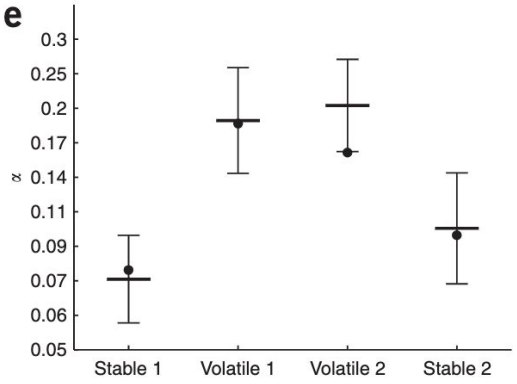
Results - First experiment

Bayesian learner and human behavior









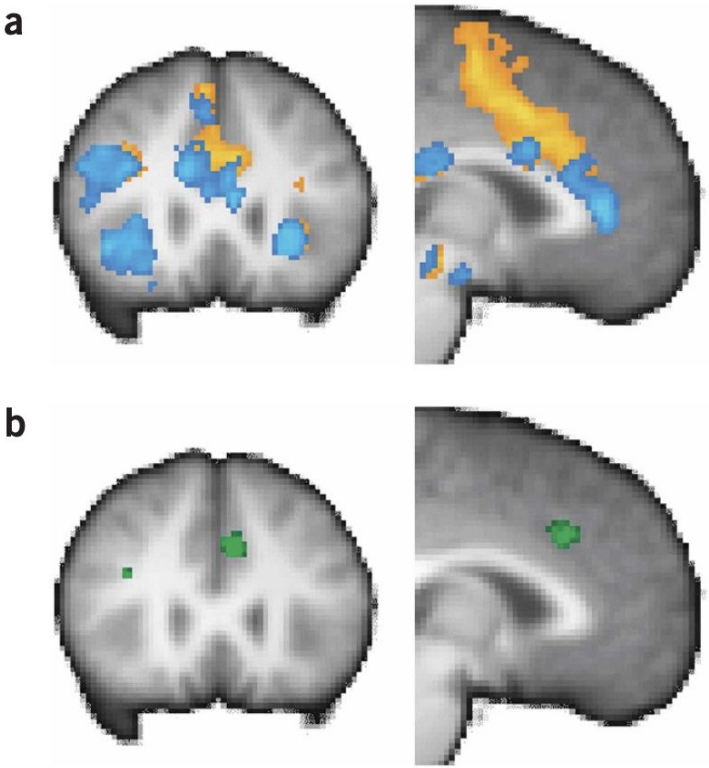
fMRI - Second experiment

- The same task as in the behavioral experiment (18 sub)
- Experiment structure
  - Stable: 60 trials (75% blue)
  - Volatile: 60 trials (80% blue and 80% green every 20 trials)
- DECIDE → INTERVAL → MONITOR: 3 phases in a trial
- The Anterior Cingulate Cortex (ACC)

Results - Second experiment

Volatility related activity in the ACC

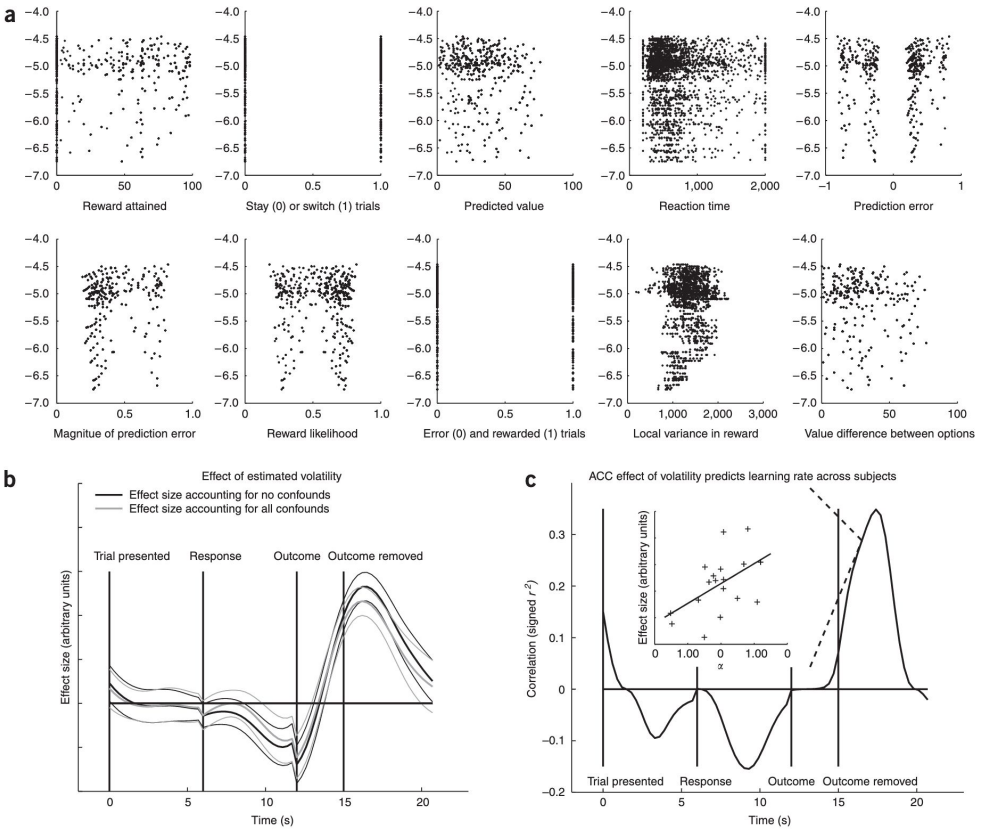
- DECIDE (orange)
- MONITOR (blue)
- Volatility x MONITOR (green)



Results - Second experiment

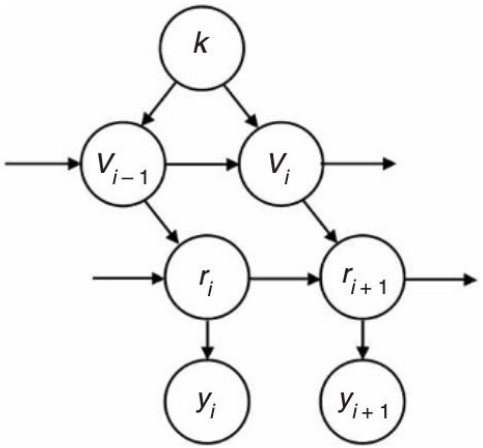
Confounding factors

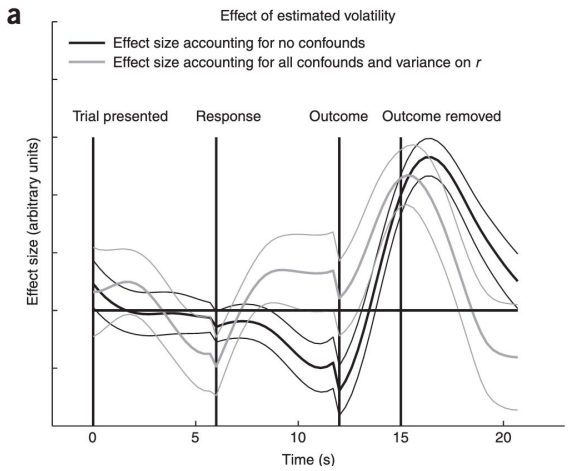
1. Reward attained by the subject
2. Switch trials
3. Predicted value of the chosen option
4. Reaction time
5. Prediction error
6. Magnitude of prediction error
7. Predicted reward likelihood
8. Error trials
9. Local variance in reward attained
10. The difference in value between the two options presented at the trial

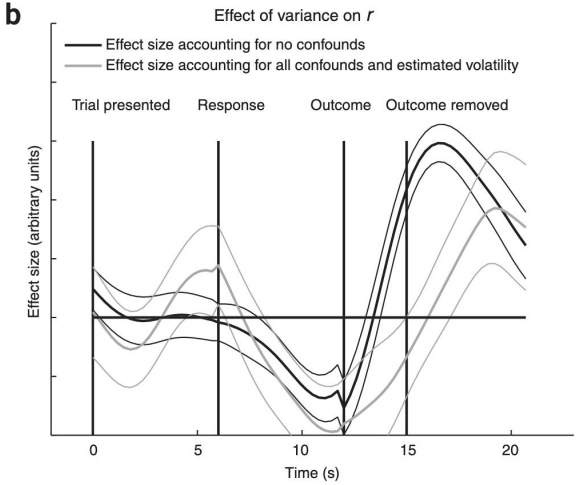


Results - Second experiment

Estimated volatility and variance in r

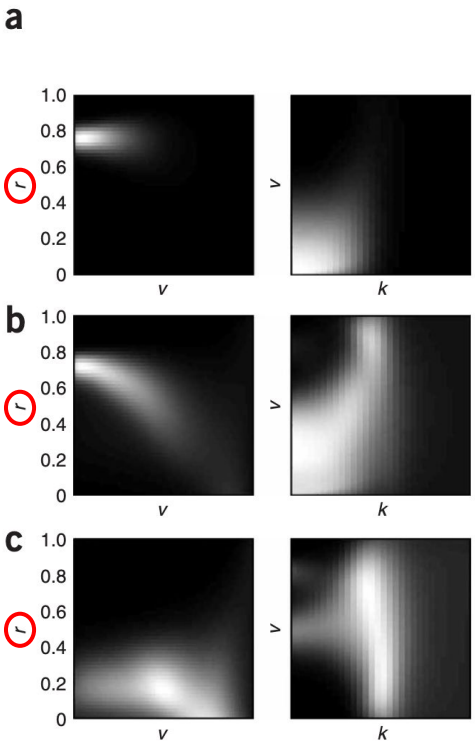


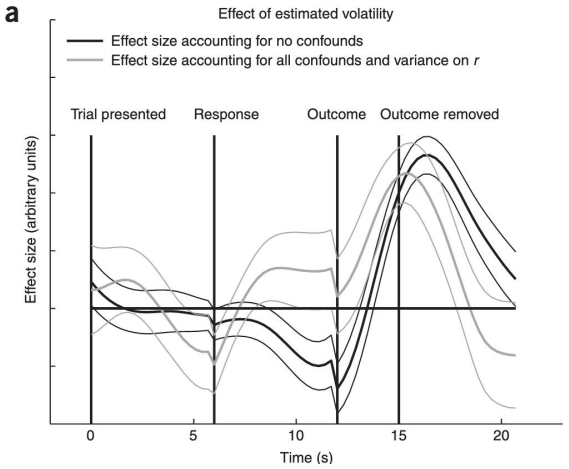


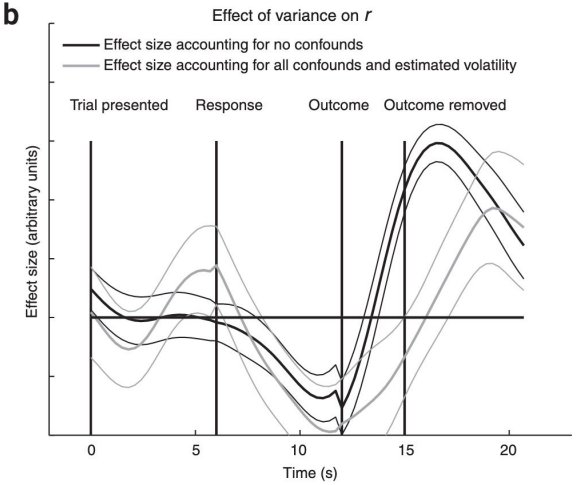


Results - Second experiment

Estimated volatility and variance in r







Discussion

- Implication
  - The role of the ACC
- Alternative interpretation
  - ACC related to subject arousal and changes in attention by response conflict?
- Ignorance of task structure and assumption on continuous outcome probability