Separately maximizing reward & information in learning

Jack Burgess[1,2,4], Erik Peterson[3], Krista Bond[3,4], Timothy Verstynen[3,4]

[1]Department of Computer Science, Dartmouth College, [2]Department of Psychological & Brain Sciences, Dartmouth College,
[3]Department of Psychology, Carnegie Mellon University, [4]Center for the Neural Basis of Cognition, Carnegie Mellon University & University of Pittsburgh

Motivation: do we intrinsically value information?

- The exploration-exploitation dilemma is considered a fundamental but intractable problem in the learning and decision sciences
- This is because it is typically formulated such that exploration and exploitation share the objective of maximizing reward
- If the problem is reformulated such that there are separate values for reward and information, there is an easy solution (Peterson & Verstynen, 2019):

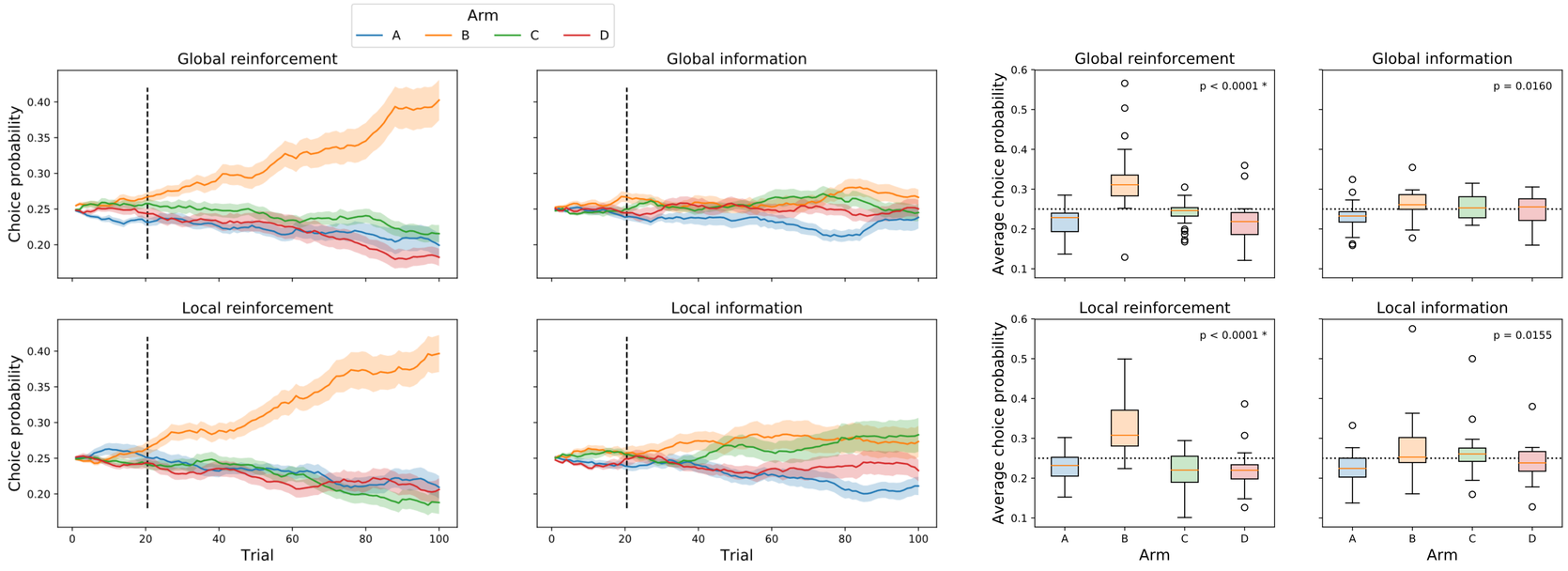
$$\pi_\pi = \begin{cases} \pi_E : E_t - \eta > R_t \\ \pi_R : E_t - \eta \leq R_t \end{cases}$$

Assuming: $\mathbb{E}[R] > 0,\ p(R) < 1,\ E - \eta \geq 0$

- This experiment was designed to test if humans value reward and information separately

Results: when information change is local, arms which change the most are explored the most

- Arms with probability changes (higher information arms) are chosen more in the local condition
- The most reinforced arm is chosen more during reinforcement blocks, as expected

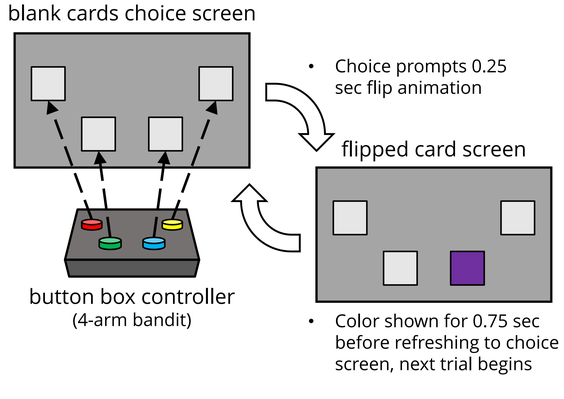


Dashed line denotes point of arm-color probability changes. Shading represents SE bounds. Subject sequential choice probability is estimated by incrementing an arm's relative probability each time it is chosen.

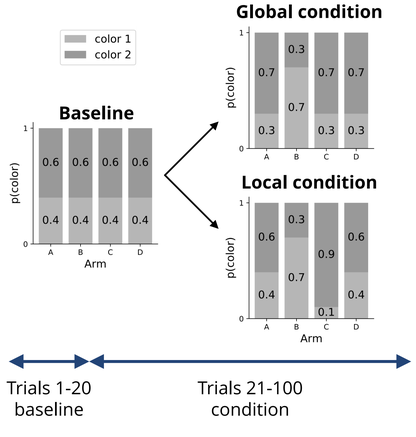
Dotted line denotes random choice probability. Average sequential choice probability taken over condition. P-values from one-way ANOVAS. * Bonferroni-corrected significance threshold of 0.0125.

Methods: testing with reward and information bandits

Trial design: 4-arm bandit



Local vs. global conditions



Reward vs. Information conditions

- Reinforcement (reward) and information blocks use distinct color pairs
- In reinforcement blocks color 1 is rewarded
- In information blocks neither color is rewarded

Conclusion: to be determined

- Our experiment found evidence supportive of the idea that reward and information value can be learned independently
- There could be many variables affecting human behavior, even those we tried to account for like individual color preference
- Future experiments will need to find ways of increasing the effect of information learning

References

Peterson, E. J., & Verstynen, T. D. (2019). A way around the exploration-exploitation dilemma. *BioRxiv*, 671362. https://doi.org/10.1101/671362