Lengthening Feedback Intervals Increases Exploitation in Choice Tasks

Darrell A. Worthy, Arthur B. Markman, & W. Todd Maddox

University of Texas at Austin

Introduction

– In experience based choice tasks participants learn rewards given for each option through repeated sampling.

–Must balance *exploitation* of the best option with the need for *exploration* of other options

–Previous research suggests that increased inter-trial intervals (ITIs) may lead to increased exploitation in a rising optimum task where exploitation is suboptimal.[1]

–We examine the effects of altering feedback intervals as well as ITIs in simple choice tasks where exploitation is optimal.

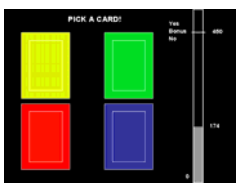
-Longer feedback and inter-trial intervals may decrease exploration of alternative options.

General Methods

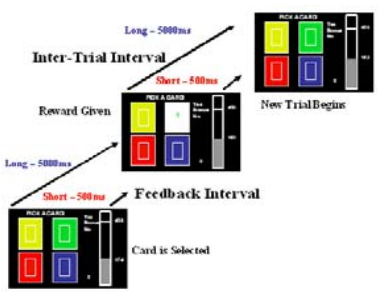
– Participants: Healthy young adults.

– Task: Four choice n-armed bandit task[2,3]

– Could earn monetary bonus



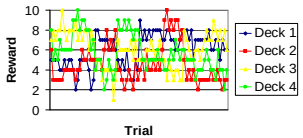
– Manipulated Feedback and Inter-Trial Intervals



Experiment 1

-Average rewards for each deck randomly changed every ten trials

Reward Given on Each Trial



- 2 (Feedback Interval) X 2 (Inter-Trial Interval) between subjects design

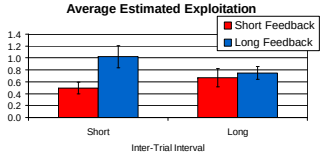
Results

-Used a well-known RL model to estimate the degree of exploitation exhibited by each participant

$$P_{a,t} = \frac{e^{(\gamma EV_t(a))}}{\sum_{b=1}^{n} e^{(\gamma EV_t(b))}}$$

Exploitation Parameter

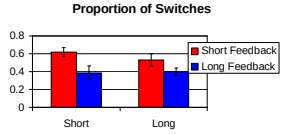
Average Estimated Exploitation



-When given short ITIs, long Feedback Intervals led to greater exploitation of the best option
- Longer ITIs did not produce as strong of an effect

Next examined the proportion of trials where a switch was made from the last deck selected

Proportion of Switches

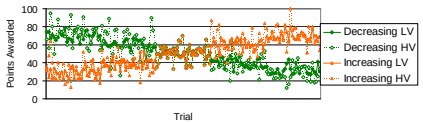


Short Feedback Interval led to more switching between decks for both ITIs

Experiment 2

-Rewards from two decks decreased over time, other two increased.
-Two high variance. two low variance

Rewards Given by Each Deck

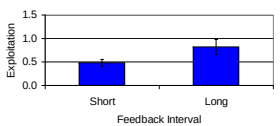


- Examined effects of Feedback Interval length for short ITIs only

Results

-Estimated exploitation parameter values as in Experiment 1

Estimated Exploitation



-Long Feedback Interval led to greater exploitation

-Participants with Long Feedback Intervals earned more points on the task.

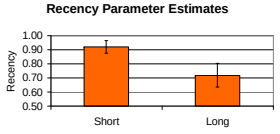
–Also examined the estimated recency parameter values given by the same model

$$EV_{k+1} = EV_k + \alpha[r_{k+1} - EV_k]$$

Recency Parameter

–Higher recency parameter values indicate greater weighting of more recent rewards

Recency Parameter Estimates

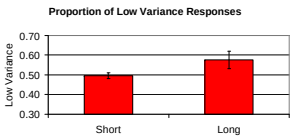


–Participants with short feedback intervals place greater weight on more recent rewards

Results

-Long Feedback Interval led to greater selection of low-variance options

Proportion of Low Variance Responses



-Long Feedback Interval may lead to less risky behavior

Discussion

-When feedback interval is increased exploitation increases.

-Participants with shorter feedback intervals placed greater weight on more recent information and were more likely to explore options with lower expected values.

–Effects were larger for feedback interval than ITI.

–Binding action to feedback may be optimal at longer time-intervals.

–Varying feedback intervals may affect decision making by a different mechanism than the eligibility trace decay mechanism proposed for ITIs.

–Participants were not more optimal with shorter time intervals, but more exploratory.

–Behavior in event based fMRI studies which require feedback intervals and ITIs of a certain length may be altered by the experimental design, not the experimental manipulation

References

References

[1] Bogacz, R., McClure, S.M., Li, J., Cohen, J.D., & Montague, P.R. (2007) Short term memory traces for action bias in human reinforcement learning. *Brain Research, 1153,* 111-121.

[2] Estes, W.K., (1950). Toward a statistical theory of learning. *Psychological Review, 57,* 94-107.

[3] Sutton, R.S., & Barto, A.G. (1998). *Reinforcement Learning: An Introduction* MIT Press, Cambridge Massachusetts.

http://homepage.psy.utexas.edu/homepage/group/MaddoxLAB/

Supported by AFOSR grant FA9550-06-1-0204, and NIMH grant to WTM and ABM, and a supplement to NIMH grant MH077708 to DAW.

worthyda@mail.utexas.edu