The relationship between two routes to adaptive learning in changing environments

Nathan Tardiff, Kathryn N. Graves, & Sharon L. Thompson-Schill
Department of Psychology, University of Pennsylvania

BACKGROUND

- Typical model-free (MF) reinforcement learning (RL) algorithms are computationally efficient but tend to fare poorly in dynamic environments. Two disparate approaches have been proposed as mechanisms by which humans and other animals might flexibly adapt to change.
    - In model-based learning (MB), a "cognitive map" can be used to dynamically update reward values via forward planning [1,2].
    - Humans display adaptive learning rates, modulating their learning rate (LR) in accordance with environmental volatility [3,4].
- To date, the interaction of these two forms of flexibility has not been tested, though neurogenetic evidence suggests that individuals who are more model-based may display less learning rate adaptation [3,5].

QUESTIONS

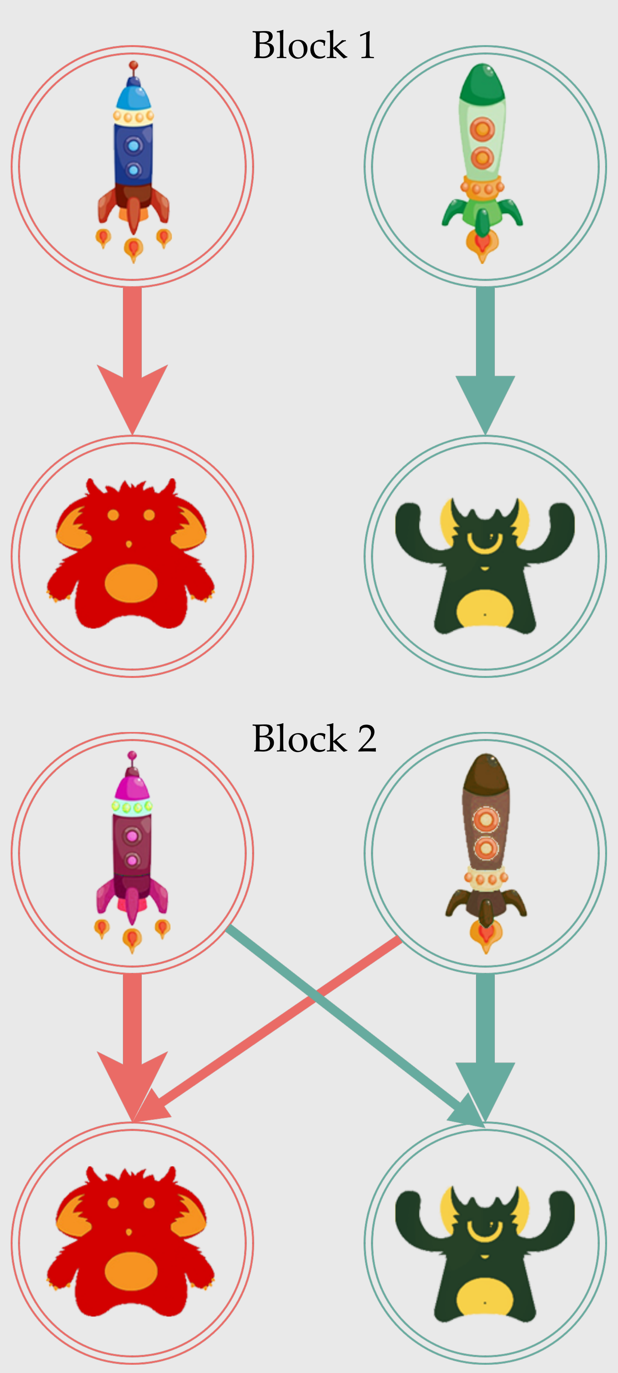
1. **What is the relationship between model-based learning and learning-rate adaptation in human subjects, and do individual differences demonstrate a tradeoff between these capacities?**
2. **How does the use of an adaptive learning rate affect the utility of model based control?**

METHODS

Subjects

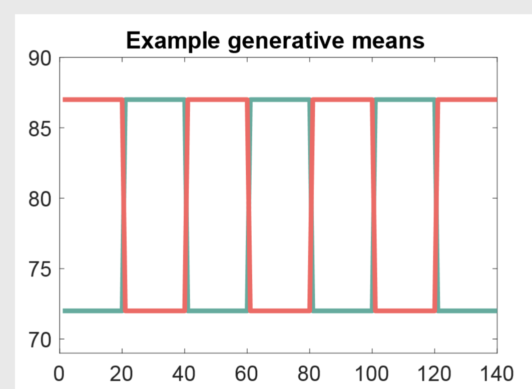
N = 200 completed two blocks of a reinforcement learning task (final N = 173 after performance cutoffs)

Two-step task



Task design

- Assesses MB control
- Two first-stage options, leading to one of two second-stage states
- Each second stage state has one choice, leading to reward
- Rewards: real-valued, gaussian (SD = 6), with a distance of 15 points between second-stage generative means
- Means of the options reverse every 20 trials
- Reversals/block: 7
- Trials/block: 140
- Block 1: reversal learning (deterministic transitions)
- Block 2: MB learning (stochastic transitions: P(common)=0.8)

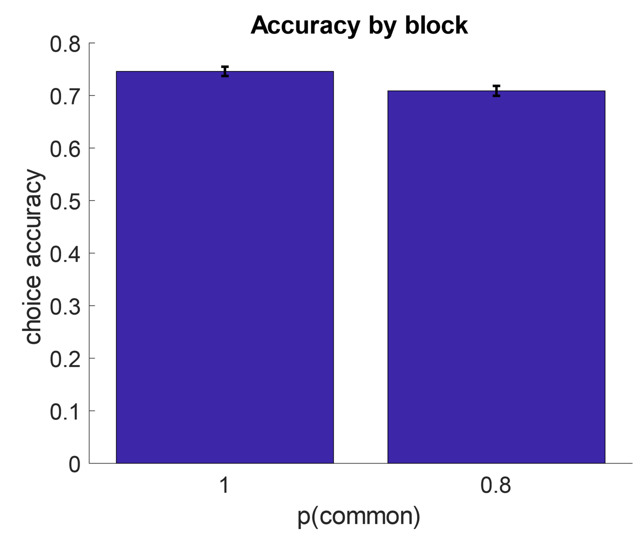


RESULTS

Behavioral Performance

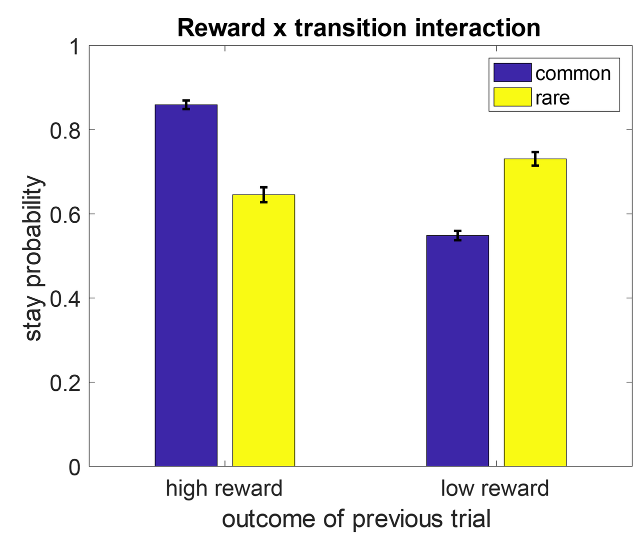
Overall accuracy and reversal learning

- Subjects learn the task, performing better in the deterministic condition ($t(172) = 3.87$, $p < .0002$).
- Subjects demonstrate variability in reversal threshold (defined as number of trials to 5/5 correct choices) and the extent to which they are model-based (see below).

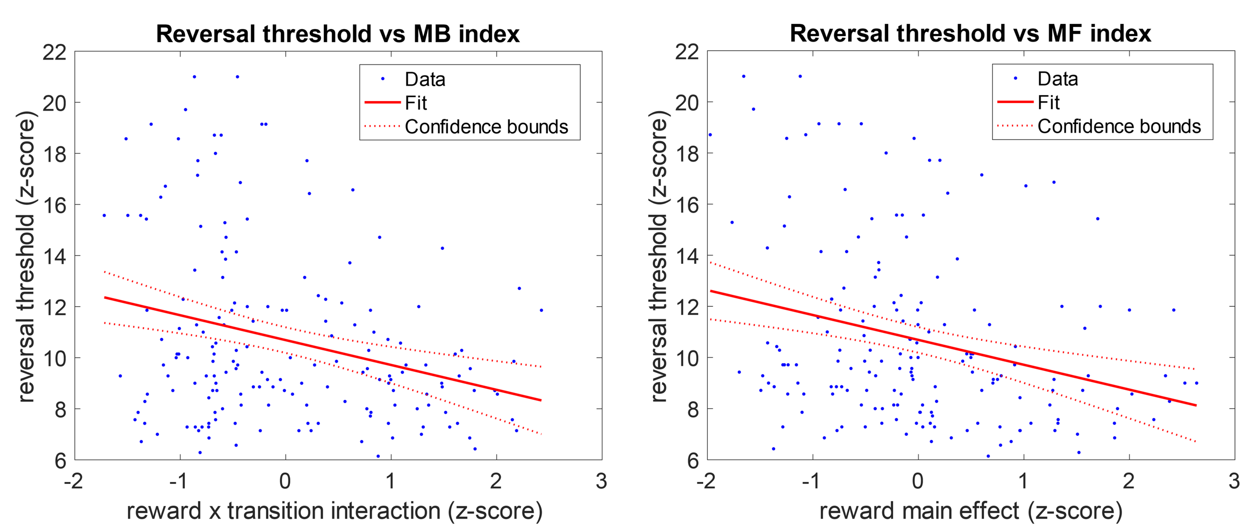


Model-based learning

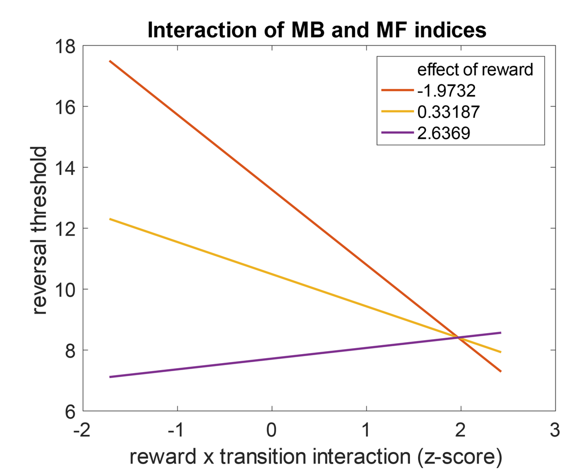
- We replicate the canonical two-step hybrid MB/MF pattern. A mixed effects regression confirms a significant main effect of reward ($\beta = 0.44$, $z = 10.05$, $p < .0001$) and a significant reward x transition interaction ($\beta = 0.46$, $z = 8.06$, $p < .0001$). Effects of transition and previous correct choice also significant (transition: $\beta = 0.19$, $z = 5.81$, $p < .0001$; correct: $\beta = 0.41$, $z = 12.01$, $p < .0001$).



Predictors of reversal performance



- Subject-specific reward x transition interaction (MB index) and reward (MF index) beta weights were extracted from the mixed effects model.
- Contrary to the hypothesis, both indices predict better reversal performance in block 1, even controlling for the effect of the other (MB: $\beta = -1.26$, $t(169) = -4.80$, $p < .0001$; MF: $\beta = -1.20$, $t(169) = -4.40$, $p < .0001$).
- There was also a significant interaction (MBxMF: : $\beta = 0.61$, $t(169) = 2.09$, $p = .038$).



ACKNOWLEDGEMENTS

Funding for this research was provided by NIH grant #R01DC009209. Task code, simulation and fitting code, and stimuli adapted from [2].

CONTACT

Thompson-Schill Lab >>
ntardiff@sas.upenn.edu

Simulation and Model Fits

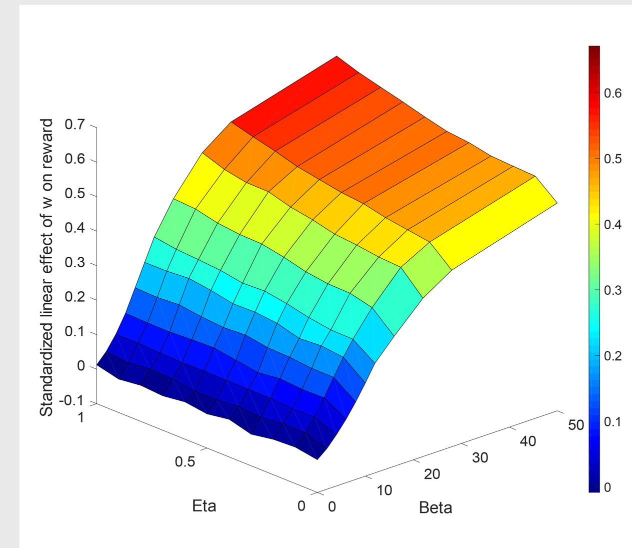
The model

- Hybrid MB/MF [2]
- MF component has variable learning rate (Pearce-Hall) [6]

Parameters					
α_{init}	β	λ	w	η	ψ
initial learning rate	inverse temp	eligibility trace	mixing weight	learning rate decay	PE scaling

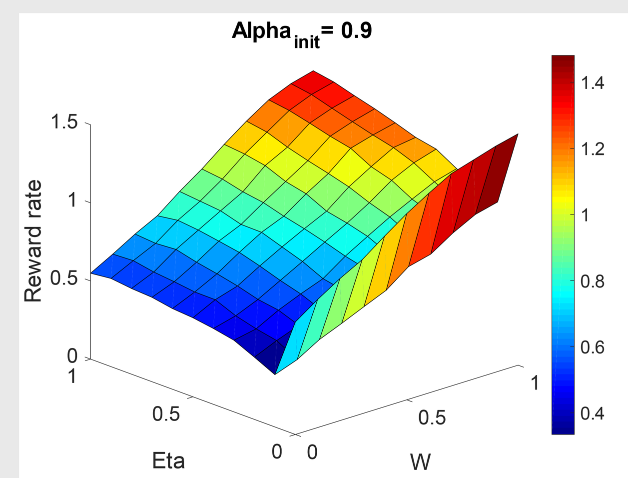
Utility of MB control

- Utility is measured as the linear effect across w for each combination of the other parameters (500 iterations).
- Averaged over α_{init}, increasing learning rate variability (η) increases the utility of MB control (tested at $\lambda = 0.5$, $\psi = 4$*SD).



Optimal parameters

- Though increasing η increases the utility of MB control, reward rate is maximized using a fixed LR ($\eta = 0$, $\alpha = 0.9$, $w = 1.0$; averaged over β; tested at $\lambda = 0.5$, $\psi = 4$*SD).
- However, a highly variable learning rate also performs well ($\eta = 1.0$, $w = 1.0$), suggesting multiple strategies for successful performance.



Model fits

- Model evidence suggests most subjects were better fit with a fixed learning rate.
- Notably, learning rates were high (fixed LR model: median = 0.81 [IQR = 0.34]).
- Corroborating the regression analysis, subjects displayed a range of w values (median = 0.54 [IQR = 0.82]).
- In the full model η and w were uncorrelated (*rho* = .097, $p = 0.20$).

Model fits				
model	LL	AIC	BIC	# best fit (AIC)
full	-10179	**22434**	25488	72
fixed LR	-10611	22605	**24641**	**101**

SUMMARY & CONCLUSIONS

- Our revised version of the two-step task successfully elicits MB control.
- We failed to find the hypothesized relationship between adaptive learning rates and MB control. However, our task design did not elicit clear evidence of adaptive learning rates.
- On average, subjects demonstrated near-optimal learning rates for the task.
- If task design is not to blame, genetic differences may not be representative of overall individual differences.
- The positive relationship between MB index and reversal performance suggests it taps learning characteristics other than the use of MB control.

REFERENCES

1. Dolan & Dayan (2013). *Neuron.*
2. Kool, Cushman, & Gershman. (2016). *PLOS Comput. Biol.*
3. Krugel et al. (2009). *PNAS.*
4. Nassar et al. (2010). *J. Neurosci.*
5. Doll et al. (2016). *J. Neurosci.*
6. Diederen & Schultz. (2015). *J. Neurophysiol.*