Action-Value Derived Evidence for Greedy Affect Control: An fMRI Study

UAMS Psychiatric Research Institute

Keith A. Bush, G. Andrew James, Clinton D. Kilts

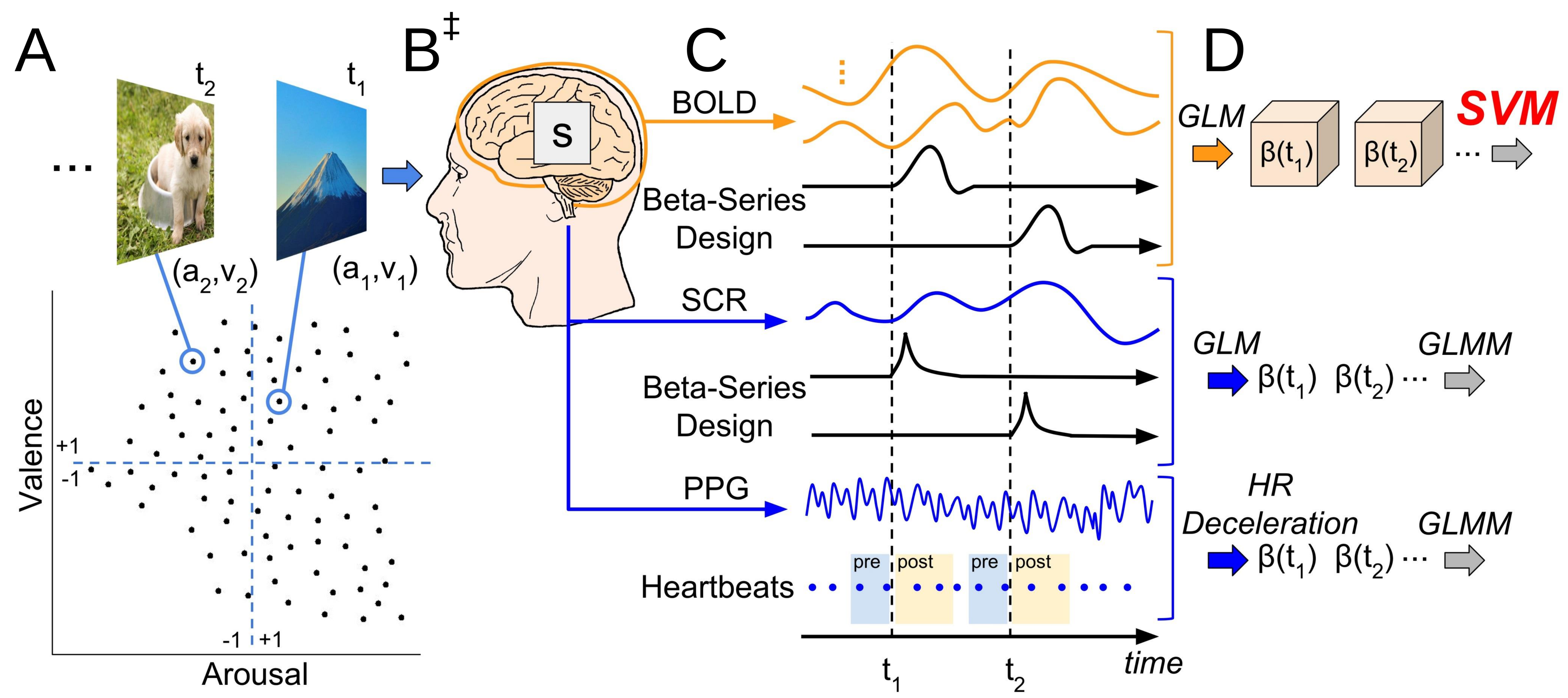
Brain Imaging Research Center, Univ. of Arkansas for Medical Sciences, Little Rock, AR.

BIRC BRAIN IMAGING RESEARCH CENTER

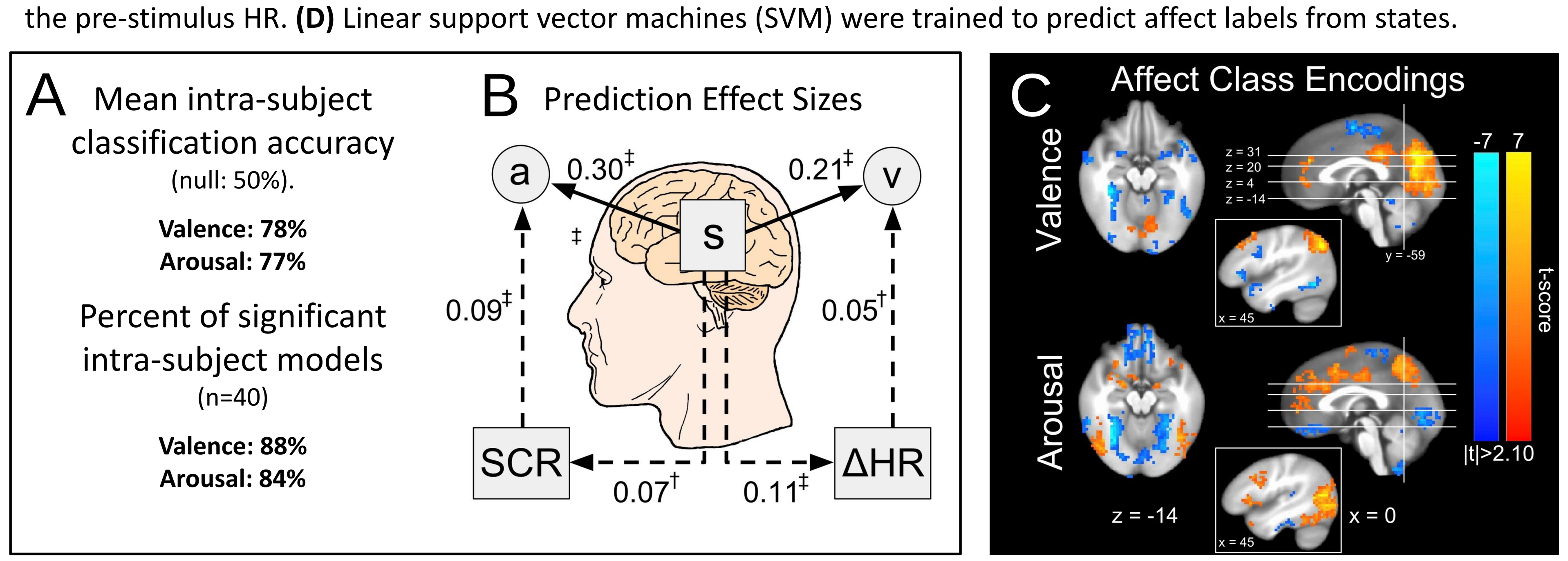
Abstract

We explored whether value-based (specifically action-value or Q-value) cognitive control obtains empirical support from functional magnetic resonance imaging (fMRI) data recorded for (n=40) healthy subjects performing an affect control task. Task trials (n=30 per subject) were comprised of International Affective Picture Set (IAPS) image stimuli (2 s) succeeded by control steps (8 s) in which subjects volitionally re-experienced the perceived affect of the stimuli while observing a fixation symbol. Affect (valence) measurements were predicted by previously reported fMRI-derived machine learning models fit separately to each subject using unique IAPS stimuli. States were defined as mean neural activations within a set of five BrainMap-derived emotion/interoception-involved independent components. Actions were defined as predicted valence differences between successive fMRI volumes discretized into (n=5) bins. Reward was defined as absolute difference between the control valence and stimulus valence in the succeeding volume. For each subject, for each of a set of discount factors (gamma) sampled on the range of [0,1] at 0.1 increments, the Q-function was modeled via random forest implementation of the fitted Q-iteration algorithm. For each discount factor and each subject, we computed: 1) on-policy out-of-sample group median Q-values; 2) random-policy out-of-sample group median Q-values; and, 3) error between on-policy actions and out-of-sample group median optimal actions. We found that on-policy Q-values were significantly greater than random policy Q-values across all discount factors supporting value-based affect control. We also found that error between on-policy actions and optimal actions was lowest for small [0.0-0.1] discount factors supporting a greedy affect control strategy.

Prior Work: Affect Prediction

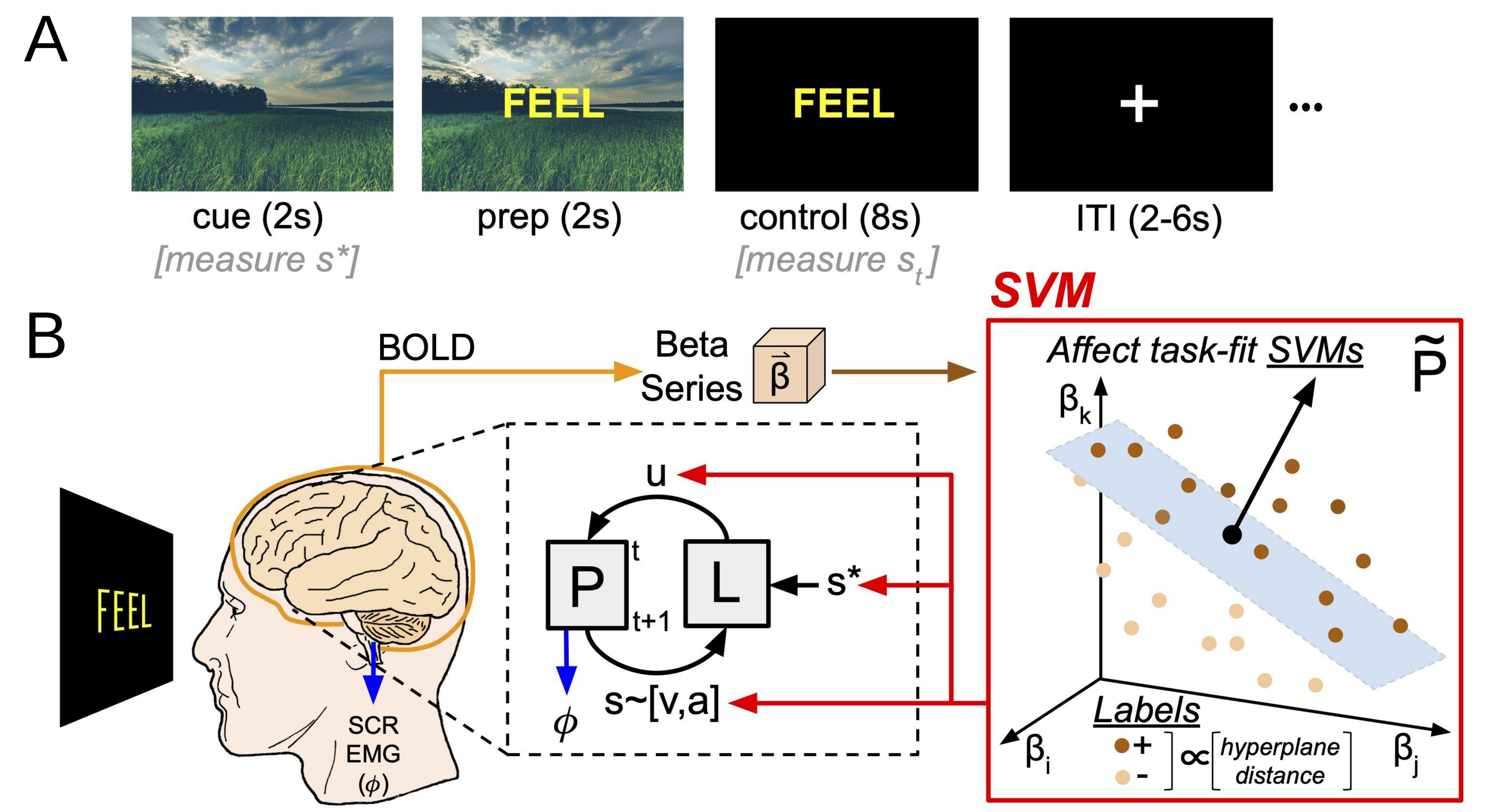


Methodological and Conceptual Overview. **(A)** Ninety IAPS stimuli presented to (n=40, 20 female) subjects, aged 18-65, for 2 s interleaved with random ITI [2–6 s]. **(B)** Concurrent fMRI measurement of the BOLD response and psychophysiology (skin conductance response, SCR, and pulse plethysmography, PPG). **(C)** Neural activation and SCR patterns (i.e., states) were extracted via the beta-series method[1]; heartrate (HR) deceleration was computed relative to the pre-stimulus HR. **(D)** Linear support vector machines (SVM) were trained to predict affect labels from states.



Results. **(A)** Prediction accuracy[2]. **(B)** Induction and convergent validity[3]. **(C)** Anatomical validity[4-5].

Cued-Recall of Affective Stimuli



Cued-Recall Experiment Design & Conceptual Model. **(A)** Subjects were presented with (n=30) cued-recall (CR) formats concurrently with fMRI. **(B)** We hypothesize that the brain will execute control during CR such that the control law, L, manipulates the state, s, of the plant, P, via action, u, to achieve goal, s* according to Q-Learning. Independent measures of physiology will verify affect induction during cued-recall.

Q-Learning

Q-Learning Methods. Subject brain states were dimensionally reduced to 5 emotion/interoception ICA dimensions[6]. Actions, u, were approximated and then binned (group-wise) into 5 discrete bins: {[∞, -2σ], (-2σ,-1σ],(-1σ,1σ),[1σ,2σ),[2σ, ∞]}. We used a random forest implementation of fitted Q-iteration[7] to estimation the Q-function. We fit Q-functions for discount factors, γ ϵ [0,1], at intervals of 0.1 for each subject.

Q-Learning Validation. Q-values at each state-action were estimated by ensemble estimation (out-of-subject). Additionally, the optimal action and expected value of random action were estimated for each state for each subject.

Policy

$$u_t = \pi^*(s_t) = \mathrm{argmax}_u Q^*(s_t,u)$$

Q-function

$$Q(s_t,u_t) = r(s_{t+1}) + \gamma \cdot \mathrm{argmax}_u Q(s_{t+1},u)$$

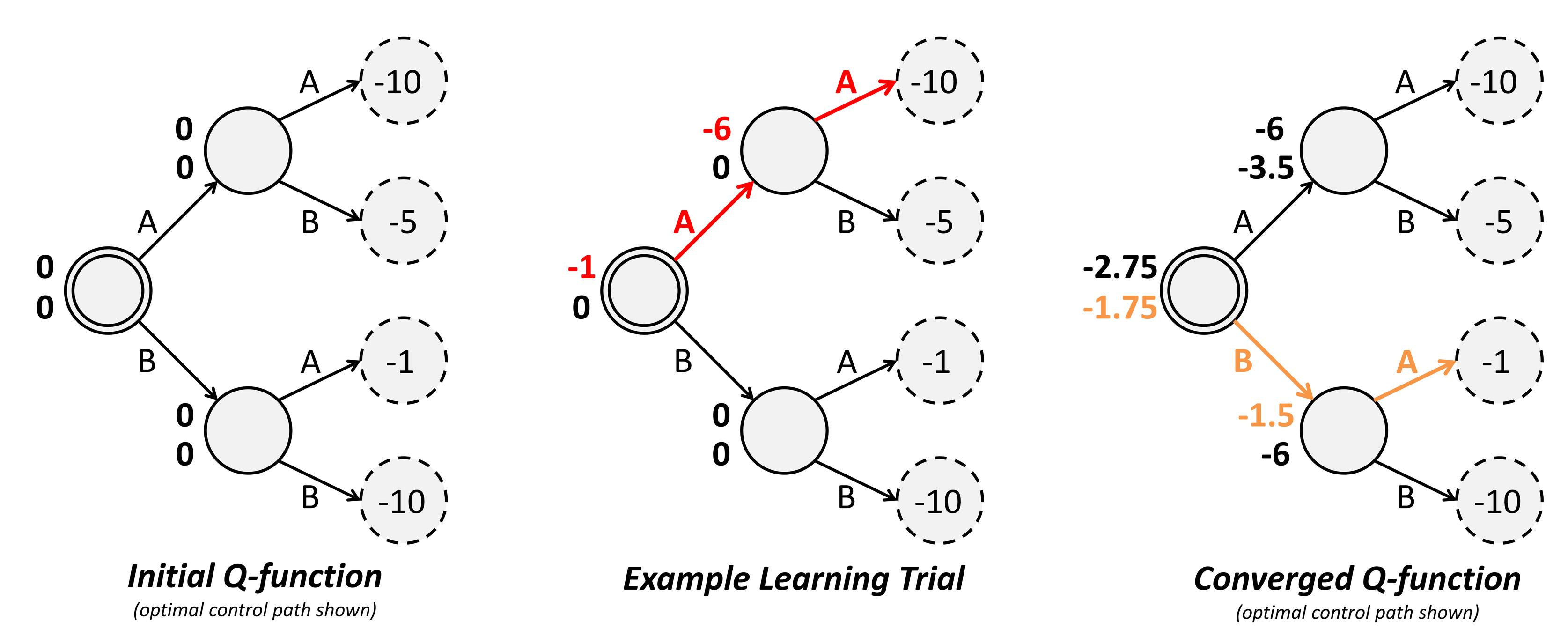
Action Approximation

$$u_t = s_{t+1} - s_t$$

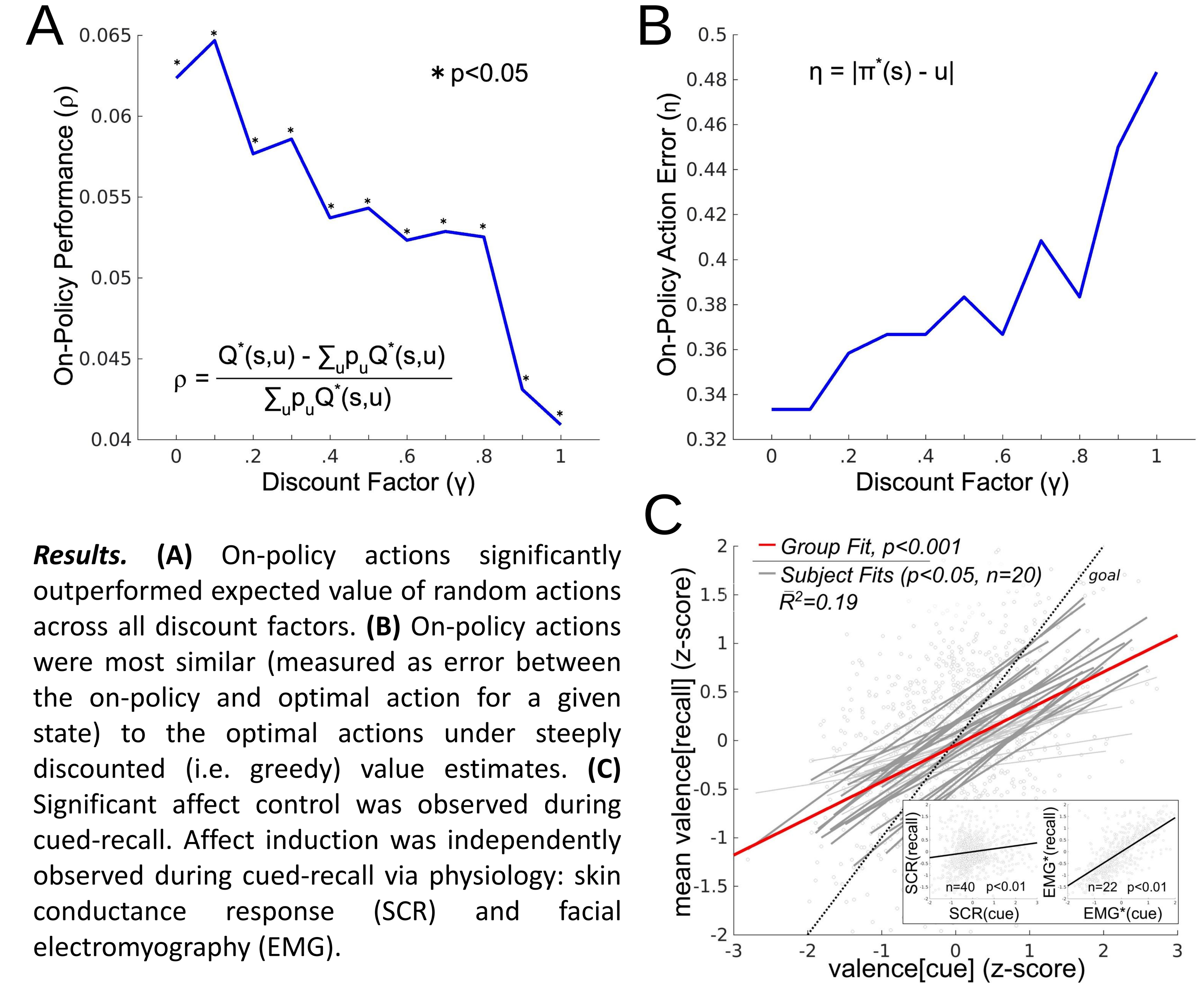
Reward Function

$$r(s_{t+1}) = |s_{t+1} - s^*|$$

Toy Example. γ=0.5, r = -1 + <terminal>



Results



Results. **(A)** On-policy actions significantly outperformed expected value of random actions across all discount factors. **(B)** On-policy actions were most similar (measured as error between the on-policy and optimal action for a given state) to the optimal actions under steeply discounted (i.e. greedy) value estimates. **(C)** Significant affect control was observed during cued-recall. Affect induction was independently observed during cued-recall via physiology: skin conductance response (SCR) and facial electromyography (EMG).

Discussion and Future Work

This study demonstrated that fMRI BOLD signal captured during the cued-recall task is sufficiently Markovian to permit approximately optimal control values, fit via Q-learning, to measure the performance of on-policy actions in comparison to random actions across a range of action-value function parameterizations. Moreover, the most likely parameterization indicated that subjects employed greedy action selection. One limitation of this work is that the reward function did not include a penalty for action (cognitive load)[8], which could alter our findings. Ongoing work in our lab incorporates action penalties into this analysis.

Acknowledgements

The authors would like to thank Kayla Wilson, Anthony Privratsky, Bradford Martins, Jennifer Payne, Natalie Morris, Nathan Jones, and Laura Spell for their efforts in acquiring data as well as Sonet Smitherman and Favrin Smith for their assistance in attaining protocol approval and maintaining human subject research compliance.

References

1. Rissman, J., Gazzaley, A. & D'Esposito, M. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* 23, 752–763 (2004).
2. Bush KA, Gardner J, Privratsky A, Chung M-H, James GA, Kilts CD. Brain States that Encode Perceived Emotion are Reproducible Across Studies but their Classification Performance is Stimulus Dependent. *Front. Human Neuroscience*, 2018; 12:262.
3. Wilson, KA, James GA, Kilts CD, Bush KA. Functional Brain States Outperform Heart Rate Variability in the Encoding of Affective Valence. *Neuropsychologia* (major revision pending, NSY-D-18-00492).
4. Bush KA, Privratsky AA, Gardner J, Zielinski MJ, Kilts CD. Common Functional Brain States Encode both Perceived Emotion and the Psychophysiological Response to Affective Stimuli, *Scientific Reports,* 2018; 8:15444. DOI:10.1038/s441598-018-33621-6.
5. Haufe, S. *et al.* On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* 87, 96–110 (2014).
6. Ray, K. L. *et al.* ICA model order selection of task co-activation networks. *Front. Neurosci.* 7, (2013).
7. Ernst, D., Geurts, P. & Wehenkel, L. Tree-based batch mode reinforcement learning. J. Mach. Learn. Res. 6, 503–556 (2005).
8. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron 79, 217–240 (2013).

Funding

Primary funding for this work was provided by the National Science Foundation (BCS-1735820). Other funding for this work was provided in part by the Arkansas Science and Technology Authority (15-B-3), the Department of Psychiatry of the University of Arkansas for Medical Sciences, and the National Institute on Drug Abuse (1R01DA036360 and 1T32DA022981).