Intracranial recordings reveal neural encoding of attention-modulated reinforcement learning in humans

Christina Maher (christina.maher@icahn.mssm.edu)
Friedman Brain Institute
Icahn School of Medicine at Mount Sinai, New York, NY

Salman Qasim (salman.qasim@mssm.edu)
Friedman Brain Institute
Icahn School of Medicine at Mount Sinai, New York, NY

Lizbeth Nuñez Martinez (lizbeth.nunezmartinez@mssm.edu)
Departments of Neuroscience, Neurosurgery, Neurology
Icahn School of Medicine at Mount Sinai, New York, NY

Ignacio Saez (ignacio.saez@mssm.edu)
Departments of Neuroscience, Neurosurgery, Neurology
Icahn School of Medicine at Mount Sinai, New York, NY

Angela Radulescu (angela.radulescu@mssm.edu)
Departments of Psychiatry and Neuroscience
Icahn School of Medicine at Mount Sinai, New York, NY

Abstract: Reinforcement learning (RL) is tractable in multidimensional environments when agents maintain efficient state representations, or mental models of relevant information. Attention supports state representations in service of RL by constraining learning to relevant dimensions. However, the physiological processes supporting value updating and attentional control are unknown. To investigate the neural mechanism supporting these processes we relate attention-modulated RL models to neuronal activity recorded directly from the prefrontal cortex of neurosurgical patients playing a multidimensional decision-making task. These models revealed that participants deploy selective attention during RL. Model-estimated expected value of the chosen stimulus correlated with neuronal activity in the orbitofrontal (OFC) and lateral prefrontal cortex (LFPC), though value signals in the LPFC were additionally biased by model-estimated attention. In sum, these results provide mechanistic insight into the neuronal implementation of the computations involved in attention-modulated RL.

Keywords: reinforcement learning; attention; intracranial electrophysiology; human prefrontal cortex

Introduction

Attention supports real-world RL by constraining available information in multidimensional environments (Niv, 2019). In doing so, attention facilitates the maintenance of state representations, or mental models of the environment which include relevant information in service of RL. Previous work proposes an algorithmic interaction of value-based learning and attention (Leong et al., 2017; Niv et al., 2015; Wilson & Niv, 2012), processes associated with the OFC (Saez et al., 2018) and LPFC (Buschman & Miller, 2007), respectively. However, how these regions interact to support multidimensional learning is not well understood. Combining intracranial electrophysiology (iEEG) and behavioral modeling we hypothesized that: (1) participants deploy selective attention during RL, (2) OFC and LPFC encode attention-modulated expected value, (3) attention biases neural value signals. We reveal a neural mechanism by which model-based computations are implemented in the OFC and LPFC.

Methods

Neurosurgical epilepsy patients (N=20) completed a multidimensional decision-making task in which they chose between stimuli varying along two dimensions: shape and color (Fig. 1A). In each block, participants were instructed which dimension was relevant (i.e., "shape"). Participants' selectively attended to the relevant dimension and learned which feature (i.e., "circle") was most rewarding. All participants performed well (Fig 1B). Gem Hunters captures naturalistic learning dynamics, as in the real-world only a subset of available information is relevant. Instructing participants of the relevant dimension allowed us to investigate efficient state representation in service of RL.

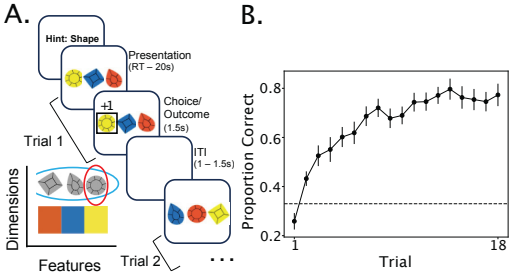

Figure 1: A. Gem Hunters task (6 blocks; 18 trials per block). **B.** Accuracy increased across trials (N=20). Dashed line = chance. Error bars = SEM.

RL models

We evaluated two RL models: Uniform Attention (UA) and Attention at Choice and Learning (ACL). Both models are based on Rescorla-Wagner learning rule. UA model implements uniform attention to both dimensions of each stimulus, whereas ACL model implements selective attention to the instructed relevant dimension. We assume participants choose between available stimuli based on their expected value (EV):

$$V_{(t)}(S_j) = \sum_d \phi_d \cdot v_t(d, S_i) \quad \text{(Eq. 1)}$$

$V_{(t)}(S_j)$ is the value of stimulus j on trial t, ϕ is the attention weight on dimension d, and $v_t(d, S_i)$ denotes the value of the feature in dimension d of stimulus S_i. Following feedback, a reward prediction error (RPE) is calculated:

$$\delta_t = r_t - V_t(S_c) \quad \text{(Eq. 2)}$$

where $V_t(S_c)$ is the chosen stimulus' EV. The RPE updates the chosen stimuli's associated feature values:

$$v_{t+1}(d, S_c) = v_t(d, S_c) + \eta \cdot \phi_d \cdot \delta_t \quad \text{(Eq. 3)}$$

The update is scaled by learning rate η. Choice probability was computed using a softmax action selection rule. The ACL model's ϕ_d was a free parameter implementing selective attention to favor the relevant dimension (Eq.1/3). The UA model's ϕ_d was fixed at 0.50 for both dimensions.

iEEG

Local field potentials were recorded from OFC (N= 144 electrodes) and LPFC (N=124 electrodes; Fig 3A). We leveraged iEEG's high spatiotemporal resolution to measure region-specific fluctuations in neuronal activity in response to model-based parameters. As our hypotheses involve local information encoding, we focused analyses on high gamma activity (70-200 Hz; HGA) because this signal captures population-level spatiotemporal dynamics and is correlated with fMRI

BOLD signal and single-unit spiking (Nir et al., 2007). Oscillatory power was z-scored to a baseline ITI.

Results and Discussion

Selective attention modulates RL. We used a leave-one-game-out cross validation procedure for maximum likelihood estimation. The ACL model best explained participants' behavior ($t(19) = 2.32$, $p < 0.05$; Fig 2A). This finding confirms our hypothesis that participants deploy selective attention to maintain efficient representations of relevant information during RL. Participants' fitted attention weight (ϕ_d) was positively correlated with task performance ($\rho(18) = 0.42$, $p < 0.05$; Fig 2B), demonstrating that even with instruction, sustained selective attention is necessary for successful RL in multidimensional environments.

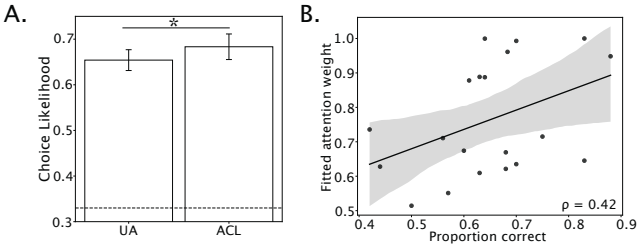

Figure 2: A. Average choice likelihood per trial shows ACL model predicted behavior significantly better than UA model ($p < 0.05$). **B.** Correlation between fitted attention weight and task performance shows attention is necessary for successful RL ($p < 0.05$).

OFC and LPFC encode attention-modulated value signals. As hypothesized, we observed a significant effect of attention-modulated EV for the chosen stimulus (ΦEV; Eq.1) on OFC and LPFC HGA power. A linear mixed effects model nested within subjects was conducted within region to estimate how strongly ΦEV was represented in OFC and LPFC HGA power while controlling for reward, chosen features, and relevant dimension. ΦEV was represented significantly in both the OFC ($\beta = -0.01$, $z = -3.78$, $p < 0.001$) and LPFC ($\beta = -0.02$, $z = -3.14$, $p < 0.01$; Fig 3B).

LPFC value signals are biased by attention. We found participants' selectively attend to relevant information to guide RL (Fig 2A). Further, individual differences in selective attention were related to performance (Fig 2B). Therefore, we hypothesized neural encoding of value signals will reflect an attentional bias. To test this hypothesis, subject-level estimates of ΦEV encoding (β coefficient) within region were correlated with participants' fitted attention weight (ϕ_d). There was a significant negative correlation between LPFC ΦEV encoding and ϕ_d ($\rho(18) = -0.69$, $p < 0.001$), demonstrating that greater selective attention to the relevant dimension is associated with stronger ΦEV encoding in the LPFC. This finding suggests the neural mechanics of attention and RL are overlapping which is supported by findings in nonhuman primates (Chiang et al., 2022; Jahn et al., 2024; Wallis et al., 2001; Wallis & Miller, 2003). This finding was region specific (OFC: $\rho(17) = -0.13$, $p = 0.59$), suggesting specialized roles for the OFC and LPFC in RL wherein the LPFC directs attention to relevant information while the OFC tracks values for relevant states (Schuck et al., 2016; Wilson et al., 2014).

Conclusion

We leveraged behavioral modeling's parameterization of latent cognitive processes and access to direct-brain recordings in humans to identify the neural architecture that supports the computational processes underlying adaptive decision-making. By integrating attention and RL, we address the complexity of value-based learning in multidimensional environments and relate this computational solution to a biologically plausible neural mechanism. Our behavioral results suggest humans selectively attend to reward-relevant information, thus maintaining efficient state representations to guide RL. Neural results reveal OFC and LPFC HGA encodes ΦEV. This encoding is biased by attention in the LPFC. Together our results provide neurocomputational correlates of flexible learning and decision-making.

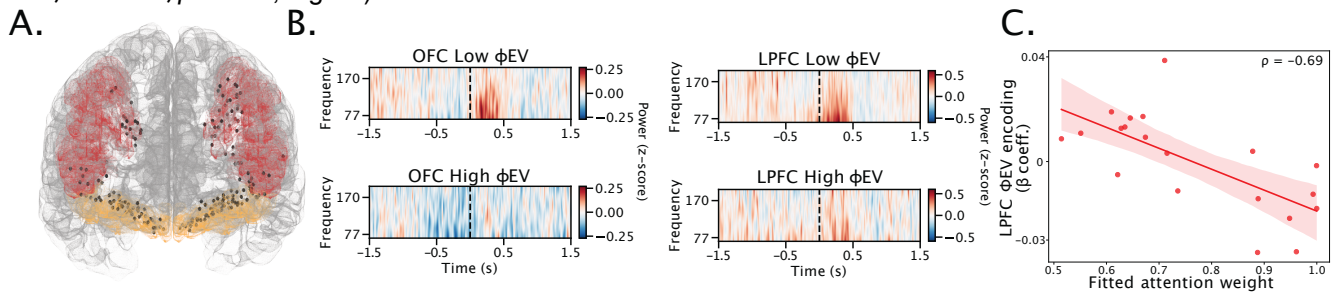

Figure 3: A. Electrodes (black) in OFC (orange; 144 electrodes) and LPFC (red; 124 electrodes). **B.** Z-scored HGA power for low/high ΦEV in two exemplar patients (OFC=11 electrodes; LPFC=5 electrodes). Dashed line = choice/reward. **C.** Correlation between fitted attention weight and LPFC attention-modulated EV encoding reveals an LPFC-specific interaction of attention and value-learning ($p < 0.001$).

References

Buschman, T. J., & Miller, E. K. (2007). Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science*, *315*(5820), 1860–1862.

Chiang, F.-K., Wallis, J. D., & Rich, E. L. (2022). Cognitive strategies shift information from single neurons to populations in prefrontal cortex. *Neuron*, *110*(4), 709-721.e4.

Jahn, C. I., Markov, N. T., Morea, B., Daw, N. D., Ebitz, R. B., & Buschman, T. J. (2024). Learning attentional templates for value-based decision-making. *Cell*, *187*(6), 1476-1489.e21.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463.

Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., & Malach, R. (2007). Coupling between Neuronal Firing Rate, Gamma LFP, and BOLD fMRI Is Related to Interneuronal Correlations. *Current Biology*, *17*(15), 1275–1285.

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157.

Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M. (2018). Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency Activity in Human OFC. *Current Biology*, *28*(18), 2889-2899.e3.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412.

Wallis, J. D., Dias, R., Robbins, T. W., & Roberts, A. C. (2001). Dissociable contributions of the orbitofrontal and lateral prefrontal cortex of the marmoset to performance on a detour reaching task. *European Journal of Neuroscience*, *13*(9), 1797–1808.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience*, *18*(7), 2069–2081.

Wilson, R. C., & Niv, Y. (2012). Inferring Relevance in a Changing World. *Frontiers in Human Neuroscience*, *5*.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279.