Contributions of Reinforcement Learning Variables to the Intracranial Reward Positivity

Rhiannon L. Cowan (rhiannon.cowan@utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Tyler S. Davis (tyler.davis@hsc.utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Bornali Kundu (bornali.kundu@gmail.com)
Department of Neurosurgery, University of Missouri
Columbia, MO 65212, USA

John D. Rolston (jrolston@bwh.harvard.edu)
Department of Neurosurgery, Brigham & Women's Hospital,
Boston, MA 02115, USA

Shervin Rahimpour (shervin.rahimpour@hsc.utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Elliot H. Smith (e.h.smith@utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Abstract:

Reward Positivity (RewP) is frequently studied EEG potential that has been related to prediction error (RPE) signals. We examine how reinforcement learning model variables correlate with intracranial measures of the RewP. We found local field potential (LFP) and high broadband frequency (HFA: 70-150 Hz) signatures of RewP, that when fit with temporal difference variables, highlight canonical structures of reward processing and differences in RewP amplitude for more impulsive choosers. Notably, HFA RewP was predicted by RPE. These results elucidate the intracranial RewP, its role in reward processing, and association with impulsivity.

Keywords: reinforcement learning; reward prediction errors; impulsive choice; reward positivity; electrocorticography

Introduction

The RewP component is an event-related potential (ERP) that occurs 250-350ms after an outcome (Proudfit, 2015). RewP is calculated as the amplitude difference between rewarded and unrewarded trials (Cockburn & Holroyd, 2018; Proudfit, 2015). Scalp EEG studies typically localize RewP over frontocentral contacts (e.g., Fz and Cz) predicting that the canonical signal is derived from anterior cingulate cortex (ACC) (Holroyd & Umemoto, 2016). However, the poor spatial resolution of scalp EEG limits interpretation of cortical RewP origin and network connectivity. Here, we utilize intracranial EEG to study RewP and its correlations with "signed" reward prediction error (RPE) (Heydari & Holroyd, 2022). RPE occurs when there is a mismatch between expected and actual rewards (Preuschoff et al., 2006) and is a fundamental element of reinforcement learning (Schultz et al., 1997). We examine whether the RewP signal is modulated by impulsive choice (IC) a key component of substance use disorders (Huys et al., 2014).

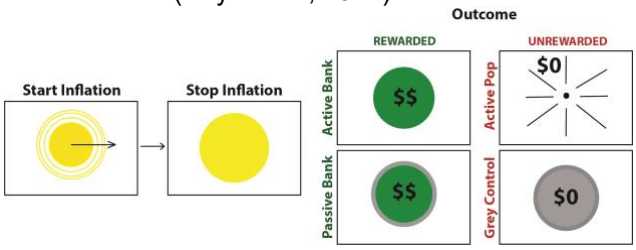

Figure 1: BART task timeline with example of single trial inflation and outcomes of either rewarded or unrewarded trials.

Method

To understand the neural basis of RewP, we fit RewP amplitudes to generalized linear models that included RPE and value expectation (VE) variables calculated from temporal difference (TD) models (Rescorla, 1972; Sutton & Barto, 2018) of the behavior and brain activity of 44 neurosurgical patients performing the Balloon Analog Risk Task (BART; (Lejuez et al., 2002). During BART, subjects inflate and stop artificial balloons to accumulate points. Unrewarded trials include grey control balloons that inflate to an indicated threshold and no points are gained, and popped trials that the patient fails to successfully stop. Rewarded trials include balloons where the subjects stop the balloon from popping and passive trials in which the balloon is inflated to its maximum size (Fig. 1). Subjects averaged 162 ± 25 rewarded trials, 71 ± 16 unrewarded trials. Subject IC level was calculated using the Kullback-Leibler divergence (KLD) between passive and active trial inflation time distributions (Hershey & Olsen, 2007). Subjects were classified as more impulsive (MI, $_{log}$KLD: 0.51, N = 20) or less impulsive (LI, $_{log}$KLD: -0.14, N = 24). RewP was defined for each recorded electrode as the difference between mean rewarded and unrewarded trials between 250 and 350 ms after outcomes for three signals: broadband LFP, HFA, and the broadband spectrogram from 1- 150 Hz (spectral).

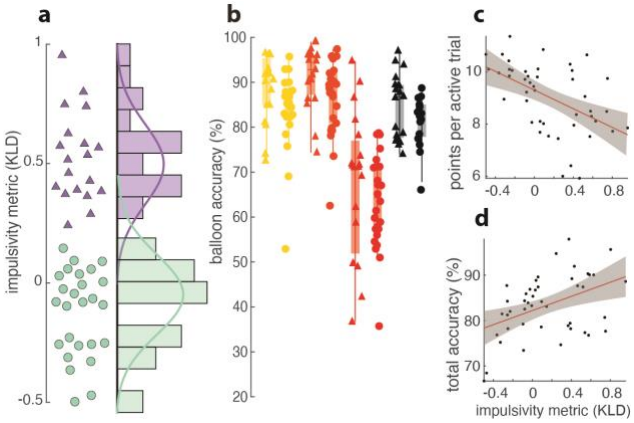

Figure 2: BART behavior related to IC (circles = LI, triangles = MI). (a) IC levels for 44 subjects. (b) accuracy by balloon color. (c-d), significant continuous relationships between IC and points per trial (c) and accuracy (d).

Intracranial correlates of RewP

In total, we examined a total of 3211 intracranial electrodes (M = 72.98 ±19.08 per subject). Rank

sum tests revealed 315 (9.81%) LFP, 297 (9.25%) HFA, and 1251 (38.96%) spectral contacts to encode RewP. Across all electrodes, each subject's outcome-aligned HFA (correlated with population neuronal firing near the electrode; (Manning et al., 2009; Miller, 2010) and LFP was modeled as a linear combination of the RewP amplitude, reward value expectation (VE), RPE, and the interaction VE, RPE, and RewP while controlling for the variance introduced by patient and electrodes as random effects in a mixed effects model. We estimated an optimal learning rate using maximum likelihood estimation (Daw, 2009). We measured whether the optimal learning rate for each participant related to impulsivity scores and depended on rewarded vs. unrewarded outcomes.

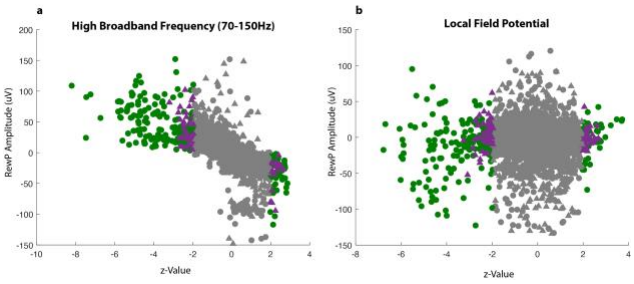

Figure 3: Scatterplots of HFA and LFP signals. a, HFA RewP difference amplitudes against z-values from ranksum tests across all contacts. b, HFA RewP difference amplitudes against z-values from ranksum tests across all contacts. Significant MI z-values ($p < 0.05$) highlighted in green. Significant LI z-values highlighted in purple.

Results

Increased impulsivity predicted greater task accuracy ($R^2 = 0.186$, $F(2, 45) = 9.83$, $p = 0.0031$), yet lower points scored during active trials ($R^2 = 0.258$, $F(2, 45) = 14.9$, $p < 0.001$) (Fig 1).

Across all electrodes, we predicted that RewP would correlate with RPE and VE. For our LFP model, we did not see that RewP difference amplitude was predicted by VE ($t = -1.91$, $p = 0.057$), RPE ($t = 0.984$, $p = 0.325$), or actual reward (points scored) ($t = 0.427$, p = 0.669). However, for the HFA model we saw both RPE ($t = -3.887$, $p = 0.0001$) and actual reward ($t = 3.532$, p = 0.0004) to be significantly predicted by RewP amplitude.

We were also interested if the continuous impulsivity level of each participant modulated the association of RewP to VE and RPE variables. For our LFP model, we saw that RewP difference amplitude was predicted by an interaction between impulsivity and VE ($t = -2.04$, p = 0.042) and a significant three-way interaction between impulsivity, VE, and actual reward ($t = 2.547$, p = 0.011). For the HFA model there were no significant interactions with impulsivity.

Anatomical regions that significantly encoded LFP RewP include middle temporal gyrus (MTG: 9.3%), middle front gyrus (MFG: 9.3%), white matter (WM: 10.8%), anterior cingulate gyrus (ACgG: 12.3%), medial orbital gyrus (15.0%) and hippocampus (HIPP: 6.1%). HFA RewP was primarily encoded in middle cingulate gyrus (16.6%), MTG (11.3%), MFG (8%), CWM (10.2%), ACgG (8.3%), HIPP (5.2%), and amygdala (5.3%).

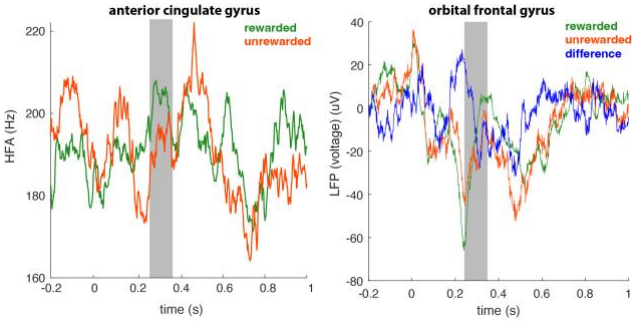

Figure 4: Examples of LFP RewP signal (grey window 250-350ms) in anterior cingulate gyrus, and orbital frontal gyrus.

Discussion

We examined the neural underpinnings of reward positivity and impulsive choice using intracranial LFP and HFA signals, correlated with TD variables value and prediction error. We observed a performance trade-off between reward and accuracy, with MI subjects opting for smaller, more immediate rewards and LI subjects opting for riskier, larger rewards but overall gaining more points. Neurally, we observed that HFA-related RewP correlated with RPE signals and actual reward. Interestingly, MI subjects tended to have smaller RewP amplitudes compared to LI subjects. These findings have implications for reinforcement learning, psychiatric disorders, and understanding the intracranial RewP.

Acknowledgments

This research was supported by funding: R01MH128187

References

Cockburn, J., & Holroyd, C. B. (2018). Feedback information and the reward positivity. *International Journal of Psychophysiology*, *132*, 243–251. https://doi.org/10.1016/j.ijpsycho.2017.11.017

Daw, N. D. (2009). Trial-by-trial data analysis using computational models. *Attention and Performance.*, 26. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Heydari, S., & Holroyd, C. B. (2022). Pain feedback interferes with reward positivity production. *Psychophysiology*, *59*(6), e14004. https://doi.org/10.1111/psyp.14004

Holroyd, C. B., & Umemoto, A. (2016). The research domain criteria framework: The case for anterior cingulate cortex. *Neuroscience and Biobehavioral Reviews*, *71*, 418–443. https://doi.org/10.1016/j.neubiorev.2016.09.021

Huys, Q. J. M., Tobler, P. N., Hasler, G., & Flagel, S. B. (2014). The role of learning-related dopamine signals in addiction vulnerability. In *Progress in Brain Research* (Vol. 211, pp. 31–77). Elsevier. https://doi.org/10.1016/B978-0-444-63425-2.00003-9

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), Article 2. https://doi.org/10.1037//1076-898X.8.2.75

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *Neuron*, *51*(3), Article 3. https://doi.org/10.1016/j.neuron.2006.06.024

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459. https://doi.org/10.1111/psyp.12370

Rescorla, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current Research and Theory*, 64–99.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.