# Reinforcement Learning and Rock, Paper, Scissors

G.A. Gagliardi, C.C. Williams, C.D. Hassall, and O.E. Krigolson

Centre for Biomedical Research, University of Victoria

University of Victoria | Theoretical and Applied Neuroscience Laboratory

NSERC CRSNG

## INTRODUCTION

**Reinforcement learning**, particularly of rewards, occurs within a dynamic and complex environment – especially in **adversarial** situations[1]

Thus, our experience of **rewards** is subject to our understanding of the **context** in which those outcomes are embedded[1]

In present study, context is a game of **Rock, Paper, Scissors** played against opponents of varying difficulty

**H1**: Win vs Loss feedback will evoke reward learning related brain activity

**H2**: Neural responses to opponent faces will reflect perceived difficulty

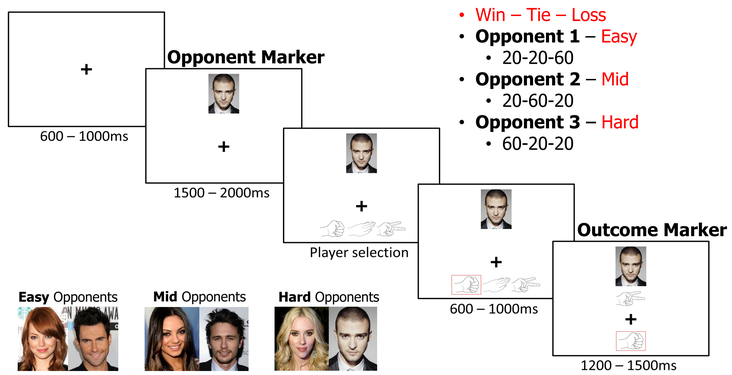
## METHODS

Participants played a game of **Rock, Paper, Scissors** against three simulated opponents

Opponent's skill varied from **easy** to **hard**, based on how often they **lost** or **won**, and were gender-matched to the player

**Event-related potentials** – neural markers for specific brain activity – were extracted from continuous electroencephalograph recordings

Measured at **outcome** & when shown **opponent faces**



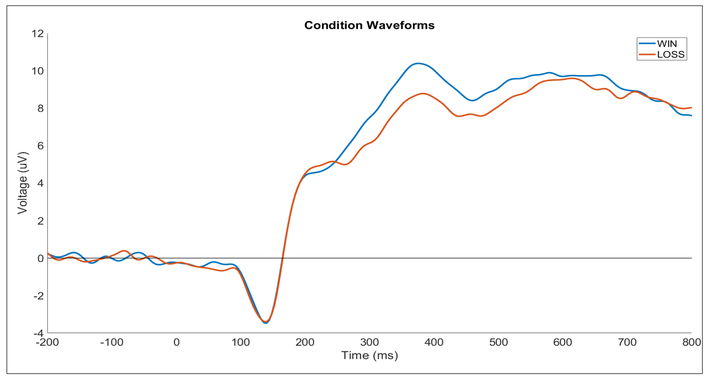




*Figure 1a.* Event-related potential waveforms for Wins (blue) or Losses (red) feedback. ERPs of interest were measured at scalp electrode FCz , where 0 is the point of stimulus presentation – opponent hand in the shape of a rock, paper or scissors. Wins or losses were determined implicitly by the player upon comparing their hand to their opponents'.

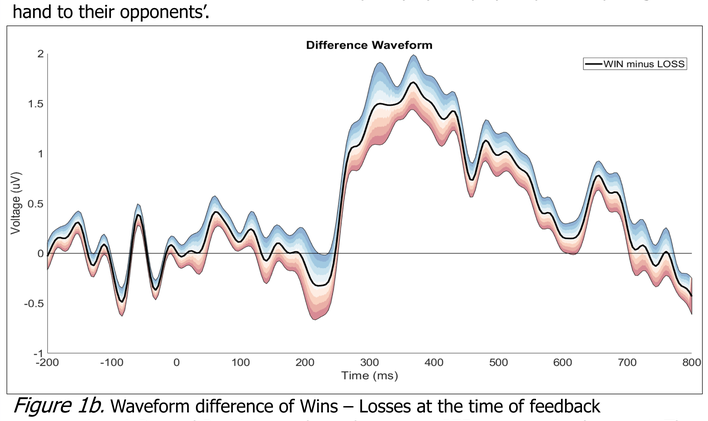



*Figure 1b.* Waveform difference of Wins – Losses at the time of feedback presentation. A Reward Positivity peak can be seen at 375ms post-stimulus onset. The duration of the Reward Positivity lasts from 300 – 450ms. Due to the complexity of the visual stimulus (a human-shaped hand in a symbolic orientation) the maximum peak and duration of this component is delayed by approximately 75ms, when compared to explicit reward responses.[3]

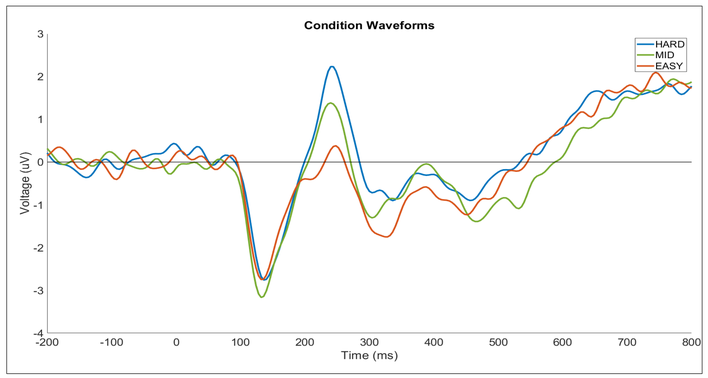



*Figure 2a.* Event-related potential waveforms for Hard (blue), Medium (green) and Easy (red) opponents measured at scalp electrode FCz. Participants had to judge from experience which face was the most difficult versus easiest to play against. The amplitude of these waveforms can be seen to scale with the level of difficulty of the three opponents.

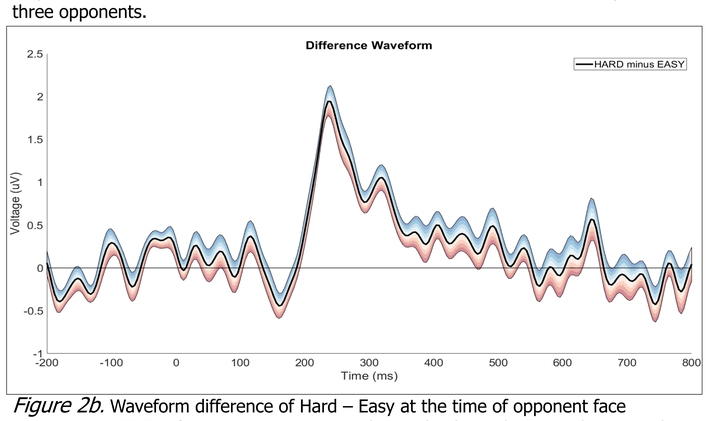
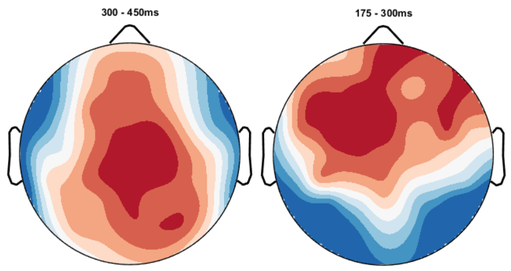



*Figure 2b.* Waveform difference of Hard – Easy at the time of opponent face presentation. ERPs of interest were measured at scalp electrode FCz , where 0 is the point of stimulus presentation – either the Hard or Easy opponent face image. A P200 peak can be seen from 175 – 300ms with a maximum peak at 245ms post-stimulus onset.

## RESULTS

*Figure 4.* Topographic maps of the defined peak regions from *Figures 1b* and *2b*. The reward positivity map (left) shows the mean peak amplitude across the scalp between 300 - 450ms, with the highest activity at fronto- and parieto-central electrodes. The P200 (right) ERP component shows the mean peak amplitudes across the scalp between 175 – 300ms, with the highest activity measured at frontocentral electrode sites.



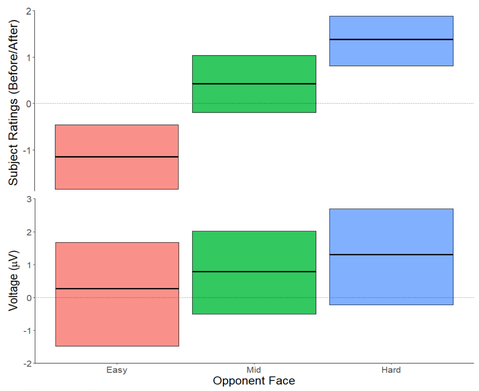




*Figure 3.* Bar plots of scaled responses to opponent faces. Participants were surveyed once at start of task for a baseline measure of their perceived difficulty of opponent faces and then again after the completion of the game. The difference between these ratings across faces is plotted above (top). Mean peak amplitudes by opponent faces is also plotted (bottom).

## CONCLUSIONS

**Outcome feedback** showed expected neural responses to reward

- **Reward positivity** ERP for Win vs Loss feedback
- Inflated latency & duration of waveform for processing implicit rewards in **context** of game environment[2]

**Opponent faces** showed scaled response for Hard>Mid>Easy

- Higher P200 component related to **emotional conflict**[4]

Subjective ratings of opponent difficulty linearly scaled to ERPs

**References**

Hassall, C.D., Hajcak, G., & Krigolson, O.E. (2019). The importance of agency in human reward processing. *Cognitive, Affective, & Behavioural Neuroscience,* 1-9.
Krigolson, O.E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. *International Journal of Psychophysiology, 132,* 175-183.
Proudfit, G.H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology, 52,* 449-459.
Zinchenko, A., Obermeier, C., Kanske, P., Schröger, E., & Kotz, S.A. (2017). Positive emotion impedes emotional but not cognitive conflict processing. *Cognitive, Affective, & Behavioural Neuroscience, 17,* 665-677.