Technical University of Munich

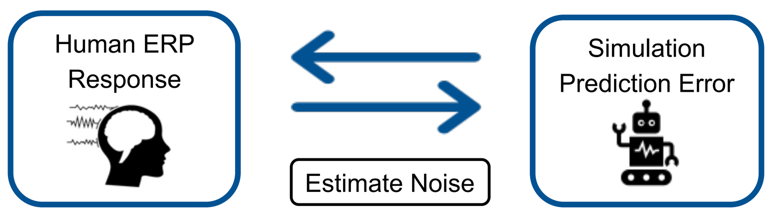
Event-related potential magnitude corresponds to prediction error from optimal policy agent observation in a gridworld experiment

Nick Tacca[1], Stefan Ehrlich[2], Gordon Cheng[2]

1. Elite Master of Science in Neuroengineering | TUM Department of Electrical and Computer Engineering | Technical University of Munich
2. Chair for Cognitive Systems | TUM Department of Electrical and Computer Engineering | Technical University of Munich

Problem Statement

The primary goal of this study is to determine the relationship between the **human event-related potential (ERP) response** and **simulation prediction error** generated from observation of a passive gridworld experiment



Motivation

- Provide a method to understand the human policy better in passive BCI scenarios through generation of simulation data
- Develop a reward structure in line with the human policy for closed-loop paradigms
- Generalize to more complex BCI scenarios that reflect the human's intentions better

Gridworld Environment

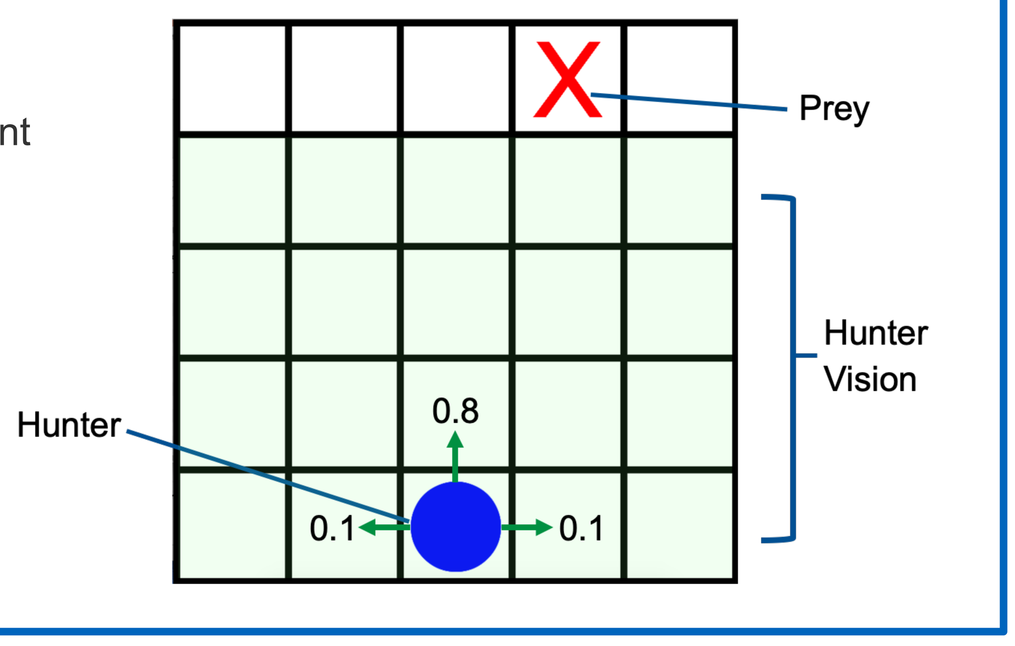
Initialization

- Hunter and prey randomly placed in gridworld environment

Action Selection

$$P(s_t, a_t) = \frac{e^{Q_t(s_t,a_t)/\tau}}{\sum_{i=1}^{n} e^{Q_t(s_t,i)/\tau}}$$

τ = 0.01
s: state
a: action



Training Agents

Q-Learning: Bellman Equation

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q_t(s_{t+1}, a) - Q_t(s_t, a_t))$$

Discount factor: γ = 0.99
Learning rate: α = 0.10

Convergence Criteria

Rolling mean of maximum policy change (Window size = 100) < 1e-6

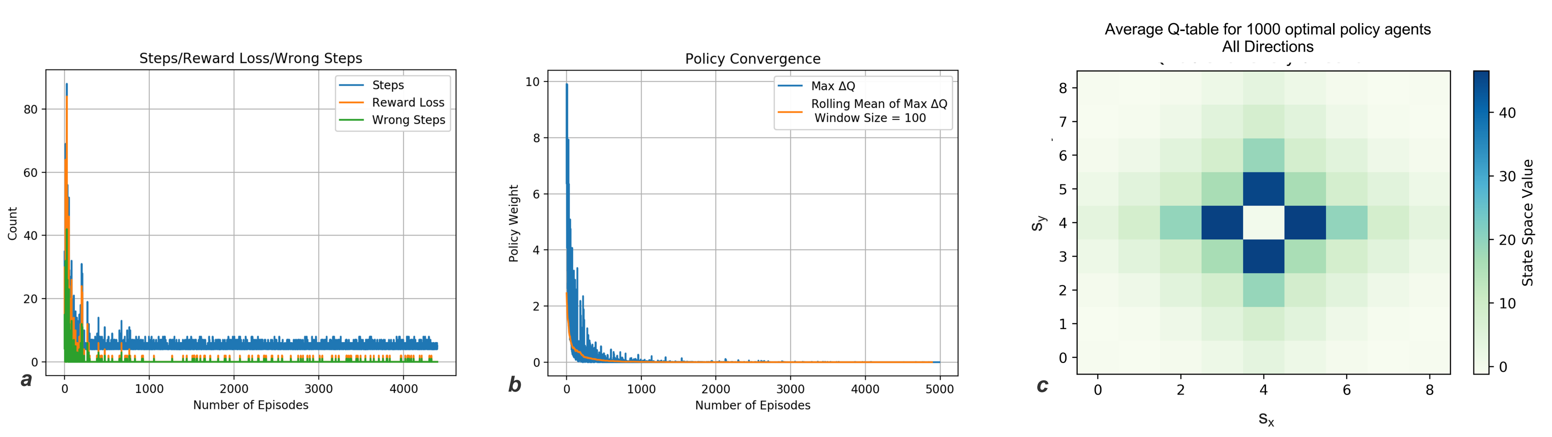


Figure 1. (a),(b) Plot of a single optimal policy agent training. (a) Evaluation of agent accuracy throughout training based on total steps taken per episode, reward loss (total possible reward - reward received), and number of wrong steps. (b) Policy convergence during training based on the maximum change in Q-values and rolling mean (window size=100) of the maximum policy change. Convergence was determined when the rolling mean of the maximum policy change was less than $1x10^{-6}$. (c) Average optimal policy Q-value heat map for 1000 trained agents based on state space in 5x5 gridworld environment.

Experiment Setup

Optimal Policy Agent

- Average of 1000 optimal policy agents

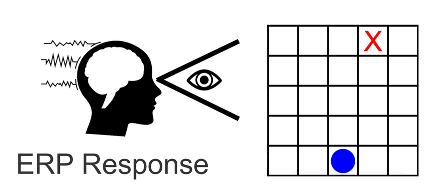
Error Generation

- Generate random number [0 1]
  if number < error rate:
  Random non-optimal action chosen
  else:
  Action chosen based on probabilities

Block Number	Error Rate	Timestep (ms)
1	0.00	1200
2	0.20	1200
3	0.40	1200
4	0.00	800
5	0.40	800
6	0.20	800
7	0.20	[800-1200]
8	0.40	[800-1200]
9	0.00	[800-1200]

Experiment Details

- One test subject for pilot experiment
- Subject asked to guess agent error rate after each block to help maintain focus
- All hunter and prey starting locations and actions recorded



Simulation Setup

Simulation Prediction Error based on Observation

- How certain the observing agent views the action taken by the hunter in the experiment based on its own policy

1000 optimal policy agents observed same experiment

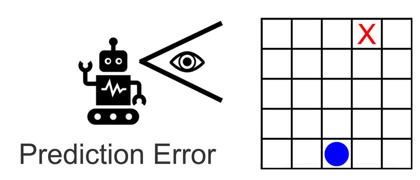
$$Q_{norm}(s_t, a_t) = \frac{Q_t(s_t, a_t)}{\sum_{i=1}^{n} Q_t(s_t, i)}$$

Normalize Q-Values for each action given the current state

$$P(s_t, a_t) = \frac{e^{Q_{norm}(s_t,a_t)/\tau}}{\sum_{i=1}^{n} e^{Q_{norm}(i)/\tau}}$$

Determine softmax probabilities for actions given state and normalized Q-values for each action (τ = 0.50)

$$Error = \begin{cases} \left|1 - \max_a P(s_t, a_t)\right|, & a = \arg\max_a P(s_t, a_t) \\ \left|0 - \max_a P(s_t, a_t)\right|, & a \neq \arg\max_a P(s_t, a_t) \end{cases}$$



Results

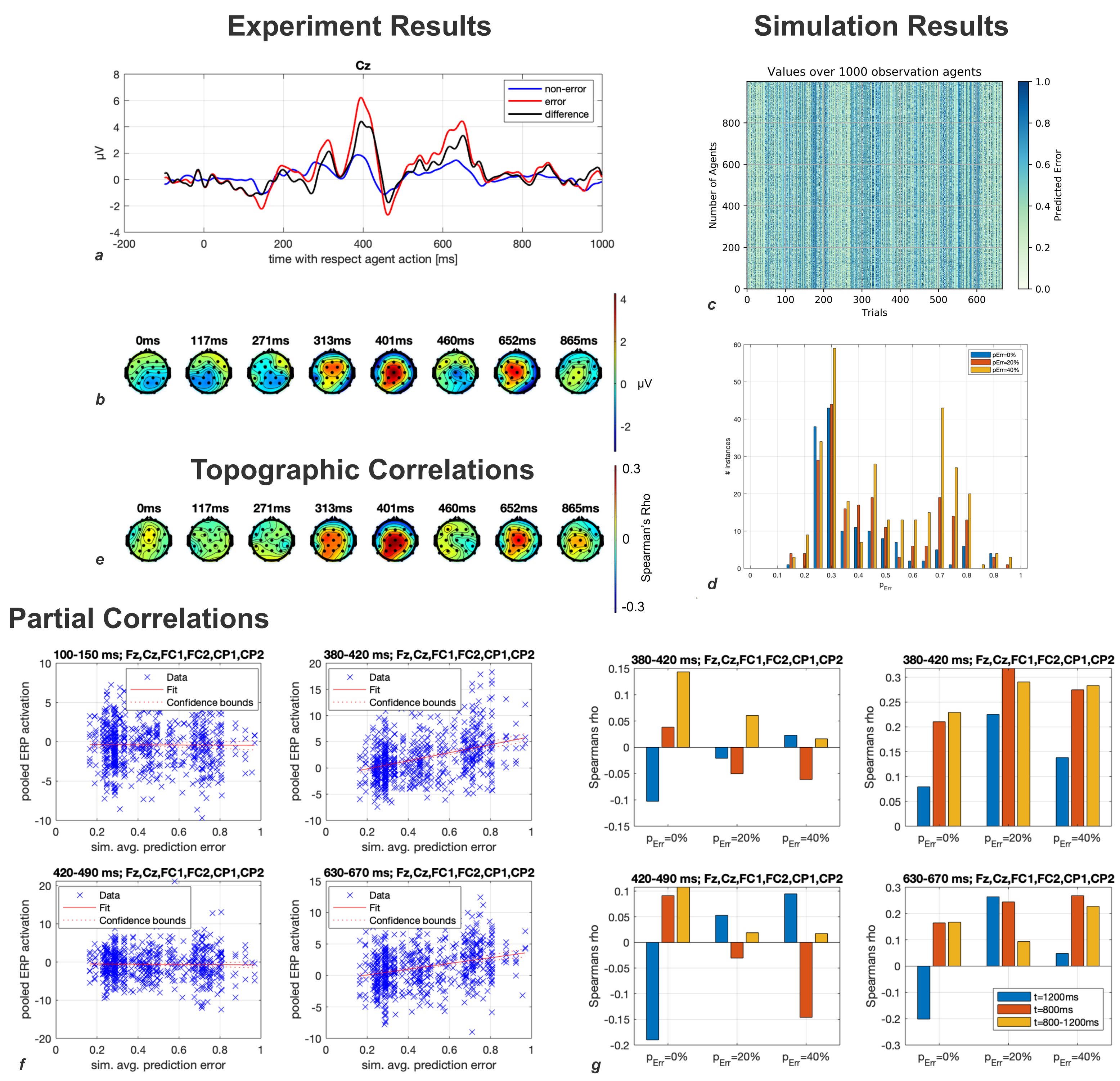


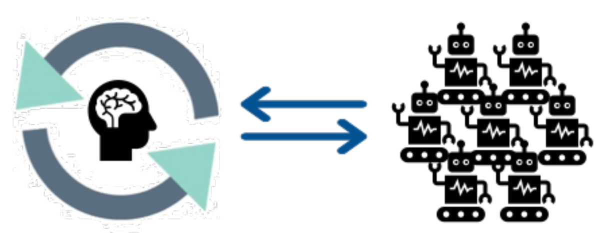
Figure 2. (a) Average ERP time courses for non-error and error events in channel Cz. (b) Topographic representation of spatial patterns post event (time locked to hunter step) for difference between error and non-error trials. (c) Average prediction error for 1000 observing agents across all trials (hunter steps). A lower value indicates the hunter step was more in line with the observing agent's policy, whereas a higher prediction error corresponds to a wrong step according to the observing agent's policy. (d) Distribution of prediction error generated from optimal policy agent observation. (e) Topographic representation of Spearman correlations between ERP response and simulation prediction error. A positive correlation exists between ERP magnitude and simulation prediction error. (f) Plot of pooled ERP activation within selected channels versus average simulation prediction error at different time regions to show partial correlations. Correlation significance depends on ERP time region. (g) Spearman correlation coefficient magnitude between average pooled ERP activation within selected channels and simulation prediction error.

Conclusions

- Correlation strength between ERP magnitude and simulation prediction error varies depending on ERP time region, hunter stepping error rate, and time-step duration.

Outlook

- Potential for the human policy to be a combination of multiple optimal policy agents (adaptive human policy).



- Further studies required to determine relationship between an adaptive human policy and observing agents' policies in order to predict ERP magnitude.

References:

Chavarriaga, R., Sobolewski, A., & Millán, J. D. R. (2014). Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in neuroscience*, *8*, 208.
Ehrlich, S. K., & Cheng, G. (2018). Human-agent co-adaptation using error-related potentials. *Journal of neural engineering*, *15*(6), 066014.
Ehrlich, S. K., & Cheng, G. (2019, October). A computational model of human decision making and learning for assessment of co-adaptation in neuro-adaptive human-robot interaction. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 264-271). IEEE.
Ehrlich, S. K., & Cheng, G. (2019). A feasibility study for validating robot actions using eeg-based error-related potentials. *International Journal of Social Robotics*, *11*(2), 271-283.
Iturrate, I., Chavarriaga, R., Montesano, L., Minguez, J., & Millán, J. D. R. (2015). Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control. Scientific reports, 5, 13893.
Kim, S. K., Kirchner, E. A., Stefes, A., & Kirchner, F. (2017). Intrinsic interactive reinforcement learning–Using error-related potentials for real world human-robot interaction. *Scientific reports*, *7*(1), 17562.