**Diseñar una propuesta experimental, usando ERPs. Luis Luarte. 07/06/2025.**

*Preamble.* My fundamental objective is to constrain a model of the brain’s decision-making algorithm Fbrain, that maps a history of experiences onto a behavioral and observable policy (the strategy that someone uses to determine the next action based on its current state) and a set of corresponding neurophysiological dynamics. I propose that Fbrain can be approximated by a formal reinforcement learning model MRL, such as Q-learning (Sutton & Barto, 2020). This model proposes that behavior is guided by a set of unobservable latent variables (λ), which are updated by specific computations, most notably the reward prediction error (Schultz et al., 1997). While the model’s utility is partially confirmed if its policy can fit observed behavior, a more rigurous validation requires demonstrating a direct correpondence between its latent computations and neural activity. This experimental proposal therefore tests the central hypothesis that there exists a significant mapping function, G, between the model’s latent variables λ and quantifiable features Φ extracted from the ERP waveform Φ(E(t)). Specifically, we seek to establish that Φ(E(t)) ≈ G(λt), where Feedback-Related-Negativityamplitude ≈ G1(δt) and PCA-Entropy ≈ G2(τt). The existance of a robust and specific mapping G would support the abductive inference that MRL is not merely descriptive of behavior, but is a computationally plausible representation of the neurophysiological processes instantiated by Fbrain.

*Research question.* Do specific ERP features systematically and dynamically track the key latent parameters of a standard Q-learning model. Does the Feedback related negativity reliably correlates with the model-derived reward prediction error (RPE)?. Does the PCA-derived entropy of the pre-decision ERP waveform correlate with the model-derived decision uncertainty temperature parameter. Then, can a computational model that incorporates these ERP features as trial-by-trial inputs achieve superior prediction of choice behavior compared to a standard model that relies on behavioral data alone?

*Hypothesis.* (I) The amplitude of the feedback-related negativity will significantly correlate with the model-derived reward prediction error, confirming the FRN as a neural idnex of RPE. (II) Decision uncertainty τt and pre-decision ERP entropy will show a positive correlation, such that PCA-Entropy ~ β1τt + β2RT will yield β1 > 0 after accounting for reaction time. (III) A neurally-informed model MRL where parameters αt, τt are modulated by ERP features will better predict choice behavior than a standard Mstandard, confirmed via formal model comparison.

*Experimental pardigm*. Participant will run a two-armed bandit task, in this task the main objective is to maximize the cumulative obtained rewards, by repeatedly choosing between two options, each delivering rewards of a variable magnitude drawn from a normal distribution (with a random walk with normal distribution steps over trials). On each trial, participants are first presented with a fixation cross for 3±0.5 seconds, then two visually distinct symbols appear on the screen for 3±0.5 seconds (pre-decision phase), afterwards a instruction is printed on the screen to make a selection via a key press, finally the reward obtained is presented for 3±0.5 seconds (post-decision phase).

*Analysis*.Using data from the pre-decision phase, I will compute frontal and central activity with a band-pass range of 0.1 – 30 Hz, from 300-600 ms after phase start as this correspond to P3 ERP which has been related with choice-evaluation modulation (Cui et al., 2013). However, as the computaitonal model uses a stochasticity-related parameter for the choice-evaluation stages we will perform a principal component analysis in a trial by trial basis, and compute eigen values shannon entropy, to determine how well-defined or stochastic is P3 related activity. For the post-decision phase I will use feedback-related negativity (FRN) (Miltner et al., 1997) potential as it directly assess reward estimation errors. As the model related parameter is directional I will use ERP amplitude as the main variable, and to obtain trial by trial estimates, I will use spatiotemporal filtering and maximum correntropy criterion to use information from multiple channels and generate noise-resistant estimates (Li et al., 2007). Afterwards, linear mixed models will be used to account for trials within subjects hierarchical structure, with MRL parameters as dependents variables and ERP derived features as independents variables, τ ~ P3 ERP-Entropy and RPE ~ FRN amplitude, respectively. For model comparison, two Q-learning model will be fit, the null model will use only behavioral data (choices and rewards), whereas the neurally informed model will use P3-ERP-Entropy and FRN amplitude within the model fitting procedure.

**Supplementary.**

*Standard Q-learning model fitting procedure (Mstandard).* The standard Q-learning model will be fitted to each participant’s behavioral data to find the parameter set that maximizes the likelihood of their observed choices. The model assumes a static learning rate α and decision temperature τ. On each trial t, after choosing action ct and receiving reward Rt, the action-value function Qt(ct) was updated according to the reward prediction error, δt:

The probability of selecting an action was determined by the softmax function:

The free parameters (α, τ) will be estimated for each subject by minimizing the negative log-likelihood of the choice data using non-linear bounded optimization.

*Neurally informed Q-learning model fitting procedure (MRL).* For this model the learning rate is going to be modeled as a logistic-sigmoid function of the standarized FRN amplitude from the previous trial (FRNt-1):

The decision temperature is going to be modeled as an exponential function of the standarized pre-decision P3-ERP-Entropy:

The free parameters for this model are going to be the weight of the mapping functions, (β0, β1, γ0, γ1). These are going to be estimated using the same maximum likelihood procedure as Mstandard.

Model comparison between Mstandard and MRL will be performed using the Akaike Information Criterion to account for the difference in the number of free paramaters.

**References.**

Cui, J., Chen, Y., Wang, Y., Shum, D. H. K., & Chan, R. C. K. (2013). Neural correlates of uncertain decision making: ERP evidence from the Iowa Gambling Task. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00776

Li, R., Principe, J. C., Bradley, M., & Ferrari, V. (2007). Robust single-trial ERP estimation based on spatiotemporal filtering. *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5206-5209. https://doi.org/10.1109/IEMBS.2007.4353515

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a “Generic” Neural System for Error Detection. *Journal of Cognitive Neuroscience*, *9*(6), 788-798. https://doi.org/10.1162/jocn.1997.9.6.788

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, *275*(5306), 1593-1599. https://doi.org/10.1126/science.275.5306.1593

Sutton, R. S., & Barto, A. (2020). *Reinforcement learning: An introduction* (Second edition). The MIT Press.