Model summary: A biologically plausible decision-making model based on interacting cortical columns

E. Baspinar[1], G. Cecchini[2], R. Moreno-Bote[4], I. Cos[2], A. Destexhe[1]

[1]CNRS, NeuroPSI, Saclay
[2]University of Barcelona, Faculty of Mathematics and Computer Science, Barcelona
[3]University of Rome, Department of Physiology and Pharmacology, Rome
[4]Pompeu Fabra University, Center for Brain and Cognition, Center for Brain and Cognition, Barcelona

Abstract

We extend the classical AdEx mean-field framework [1] to model two networks of excitatory-inhibitory neurons, representing two cortical columns, and interconnected with excitatory connections contacting both Regularly Spiking (excitatory) and Fast Spiking (inhibitory) cells. This connection scheme introduces bicolumnar competition. Each column represents a pool of neurons making the decision in favor of one of the two choices represented by two partially filled bars appearing on a monitor in human experiments. Task is based on maximizing total reward provided at the end of each episode consisting of a number of trials. The total reward depends on the coherency between choices of the subject/model and implemented strategy in the experiment. We endow the model with reward-driven plasticity mechanism allowing to capture the implemented strategy, as well as to model subject exploratory behavior. The model provides a biophysical ground for simpler phenomenological models proposed for similar decision-making tasks and can be applied to neurophysiological data obtained from the macaque brain. Finally, it can be embedded into whole-brain scale platforms such as The Virtual Brain (TVB) to study decision making neural dynamics in terms of large scale brain dynamics. This Jupyter notebook produces performance and reaction time measurements obtained from model simulations. It allows to simulate the model with different parameters settings and contains three case studies to create acquaintance with the model the user. It provides the model implementation which was used in [2].

1 Experiment background

Decision making is crucial for survival of many species. It refers to selecting a perceptual or motor action between alternatives by taking into account the consequences of each alternative. To produce a cognitive perceptual and/or motor behavior, the brain employs several decision-making mechanisms which encode and interpret sensory stimuli, weight evidence to select between choice alternatives, and finally, which generates a perceptual decision and/or motor action selection.

Decisions are made in general by taking into account immediate and longer term consequences. In many cases, following an optimal decision making in consideration of stronger benefits in long term is important for survival of many species. Therefore, control over future planning is required to identify optimal decision making strategy. To study behavioral background of such mechanisms in human, we designed a task with two versions: Horizon 0 and Horizon 1. Behavioral results are recorded in terms of performance measurements and reaction times (RTs).

Each horizon is composed of 100 episodes. In Horizon 0, one episode has one single trial. In Horizon 1, one episode is composed of two successive trials, which are named as Trial 1 and Trial 2. In each trial, the subject is asked to choose one of the two stimuli, which each one is a partially filled identical vertical bar. Percentage of the filled part is different for each bar. Subject receives a reward at the end of each trial if the decision is in coherence with the preset strategy in the task, otherwise the subject receives a loss. Reward (loss) is a randomly generated quantity which is added to (subtracted from) the filled parts of the two stimuli at the end of the trial. Goal of the subject is to maximize (minimize) the total reward (total loss) at the end of the episode. This

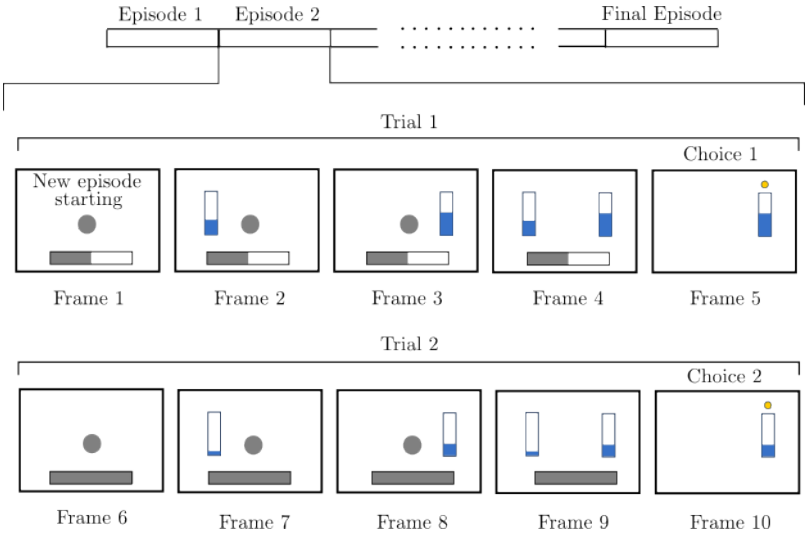

Figure 1: One example episode of Horizon 1 experiment. CT appears in Frame 1 to indicate the beginning of the episode. PTs are shown separately in Frames 2 and 3. Then PTs are shown together in Frame 4. Subject clicks on the large PT as highlighted by the yellow dot in Frame 5. The end of the episode is indicated in Frame 6 with a dot at the center and a progress bar. PTs after the subtracted loss ($= G$) are shown separately in Figures 7 and 8. Finally, updated PTs are shown together in Frame 9 and the subject chooses the large one as highlighted in Frame 10. Performance measure of the episode is the normalized sum of the chosen stimuli by the corresponding sum of the chosen stimuli to the best case scenario, which is in complete coherence with the preset strategy. The performance is expressed in terms of a loading bar. A new episode starts without feeding back the performance to the subject.

requires: (i) to identify the correspondence of reward and loss to the made choices, (ii) to capture the preset strategy to optimize the made choices in accordance with the strategy.

Both horizons were applied to a healthy human subject. Two stimuli are shown to the subject in each trial. Amount of fillings of the bars are $M \pm d/2$ where both M and d are randomly generated numbers from a uniform distribution. Here M denotes the mean value of the stimuli, and d represents the difficulty since it determines the perceptual difficulty to distinguish between two stimuli. Preset strategy is to choose the small stimulus in Trial 1 and the large stimulus in Trial 2. This boils down to choosing the small stimulus in each episode in Horizon 0. In Horizon 1, if the subject chooses the small stimulus in Trial 1, reward G is added to both stimuli and $M + G \pm d/2$ become the filling quantities of the two stimuli in Trial 2. Otherwise, the subject receives loss and the filling quantity becomes $M - G \pm d/2$ in Trial 2. The same procedure is applied in Trial 2 but now the expected choice by the strategy is the large stimulus. At the end of the episode, the total reward the subject receives is equal to the sum of the chosen stimuli. In this way, difficulty of the episode can be taken into account when performance of the subject is measured; see Figure 1 for an example scenario of Horizon 1.

In Horizon 1, protocol is the same for each trial. Trial begins with central target (CT), which appears at the center of the monitor and remains 500 ms to indicate that the episode starts. In addition to the CT, a progress bar appears and remains at the bottom of the screen during the whole episode to indicate in which trial the subject is. Once the CT disappears, Stimuli 1 and 2 are shown as peripheral targets (PT) appearing on the left and right hand sides of the center location where the CT was shown previously, at 500 ms and 1050 ms, respectively. Each one of them remains on the screen for 500 ms and then disappears. At 1600 ms, both of them are shown at the same time and Trial 1 begins. As the subject moves the pointer, the trajectory of the pointer is recorded and the time duration between the instant the subject starts to move the pointer and the instant where the subject clicks on one of the stimuli is saved as the RT. The decision is made for the clicked stimulus within 2000 ms after Trial 1 begins. After clicking, both stimuli disappear at 3600, and after 750 ms a new CT and the updated progress bar appear at 4350 ms to indicate

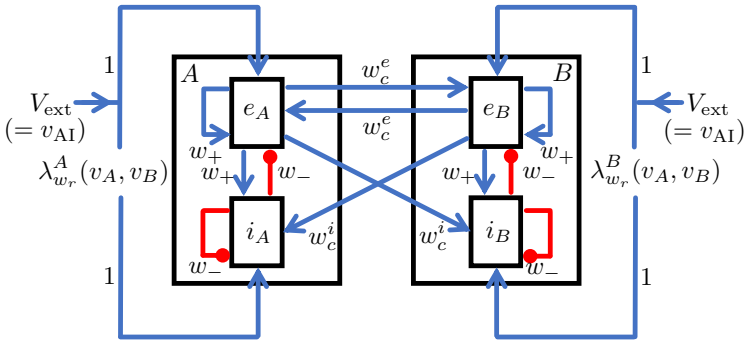

Figure 2: Basic module with two pools excitatory and inhibitory populations

that Trial 2 begins. If there is no clicking within 2000 ms between 1600 and 3600 ms, the stimuli disappear at 3600 ms and a new CT appears at 4350 ms with the progress bar updated. Then the CT disappears after 500 ms and the updated stimuli with reward (or loss) appear for Trial 2 by following the same protocol as in Trial 1. In case of an error, the protocol does not stop until the end but the anomaly is indicated in the behavioral recordings. A new episode is initialized after an inter-episode interval of 2000 ms. The protocol is the same for Horizon 0, with a single difference: there is no progress bar.

2 Model

2.1 Basic module

Basic module produces the dynamics of two columns competing with each other. Each column is modeled as a pair, called pool, of excitatory and inhibitory populations. Pools A and B vote for Stimulus A represented with v_A and Stimulus B denoted by v_B, respectively. We will denote the excitatory populations by e_A, e_B and inhibitory populations by i_A, i_B, where subindices mark the corresponding pool. There are only excitatory connections across the pools. Each cross-pool excitatory connection targets both excitatory and inhibitory populations of the other pool. Excitatory and inhibitory populations within each pool are fully connected with recurrent and cross-population connections; see Figure 2.

In each pool, there exist N_e excitatory and N_i inhibitory neurons accounting all together to $N_{\text{tot}} = N_e + N_i$ neurons. We denote the internal excitatory and inhibitory connection weights by w_+ and w_-, respectively. We employ w^e_c and w^i_c for the intercolumnar connections and the supindices denote the target populations.

External input for the excitatory populations consists of two components. The first one, v_{AI}, represents the base drive keeping excitatory populations in asynchronous irregular (AI) state. We fix v_{AI} to 5 Hz. The second component, $\lambda^A_{w_r}(v_A, v_B)$ or $\lambda^B_{w_r}(v_A, v_B)$, represents the external input with bias which is introduced by regulatory module modeling plasticity as explained in the following section. Biased input is fed to Pools A and B through λ function simultaneously.

Our model consists of 18 state variables: v_α, $C_{\alpha\beta}$ and W_α where $\alpha, \beta \in \{e_A, i_A, e_B, i_B\}$ and $C_{\alpha\beta} = C_{\beta\alpha}$ for all α, β. Here v_α is the firing rate of population α and $C_{\alpha\beta}$ denotes the covariance between v_α and v_β. Variable W_α denotes slow adaptation for population α. Inhibitory neurons are known to have no adaptation, therefore $W_{i_A} = W_{i_B} = 0$ for all $t \in [0, \infty)$ with t denoting the time variable.

We assume that the derivatives of adaptation variables with respect to firing rates v_α are 0. This is due to that the adaptation variables W_α evolve on slow time scale and its change with respect to firing rates v_α evolving on fast time scale is negligible.

The model equations read as:

$$\begin{aligned} T\,\partial_t v_\alpha =&(F_\alpha - v_\alpha) + \frac{1}{2}C_{\xi\eta}\partial_{\xi\eta}F_\alpha + \sigma\,\omega_\alpha \\ T\,\partial_t C_{\alpha\beta} =&\delta_{\alpha\beta}A^{-1}_{\alpha\alpha} + (F_\alpha - v_\alpha)(F_\beta - v_\beta) + C_{\beta\xi}\partial_\xi F_\alpha + C_{\alpha\xi}\partial_\xi F_\eta - 2C_{\alpha\beta} \\ \partial_t W_\alpha =&-\frac{W_\alpha}{\tau_w} + \Big(\delta_{\alpha e_A} + \delta_{\alpha e_B}\Big)\Big(bv_\alpha + a\Big(\mu_V(v_{e_A}, v_{e_B}, W_\alpha) - E_L\Big)\Big), \end{aligned} \tag{1}$$

where

$$\begin{aligned} F_{e_A} &= F_{e_A}(\tilde{v}_{e_A}, \tilde{v}_{i_A}, W_{e_A}), \qquad F_{i_A} = F_{e_A}(\tilde{v}_{e_A}, \tilde{v}_{i_A}, W_{i_A}), \\ F_{e_B} &= F_{e_B}(\tilde{v}_{e_B}, \tilde{v}_{i_B}, W_{e_B}), \qquad F_{i_B} = F_{e_A}(\tilde{v}_{e_B}, \tilde{v}_{i_B}, W_{i_B}), \end{aligned} \tag{2}$$

are the population transfer functions with subindices indicating the corresponding population. Here $\omega_\alpha = w_\alpha(t)$ is a white Gaussian noise:

$$\mathbb{E}[\omega_\alpha(t)] = 0, \qquad \mathbb{E}[\omega_\alpha(t)\omega_\beta(t')] = \delta_{\alpha\beta}\delta_{tt'}, \quad \text{for all } t, t' \geq 0.$$

In (1), $\sigma > 0$ denotes noise intensity. Function $A_{\alpha\beta}$ is defined as follows [1]:

$$A_{\alpha\beta} = \delta_{\alpha\beta}\frac{N_\alpha}{F_\alpha(1/T - F_\beta)},$$

with N_α denoting the number of neurons in population α. Term T is the time scale parameter both for the firing rate and for the cross correlation variables appearing in (1) and it should be chosen properly not to violate Markovian assumption [1]. We use the same term μ_V as given in [1, Section 2.3.1]. We use the notation given by $\partial_\alpha = \frac{\partial}{\partial v_\alpha}$ and $\partial_{\alpha\beta} = \frac{\partial^2}{\partial v_\alpha \partial v_\beta}$ for the partial derivatives. Here δ is the Dirac delta function, E_L is a constant representing the reverse leakage potential, and finally τ_w is a time scale-like parameter for W_α. We express the probability of intercolumnar connectivity with $p_c \geq 0$. Finally we write regulated inputs $\tilde{v}_\alpha$ of the transfer functions given in (2) as follows:

$$\begin{aligned} \tilde{v}_{e_A}(t) = v_{e_A}(t) &+ v_{\mathrm{AI}} + \lambda^A_{w_r}(v_A(t), v_B(t)) \\ &+ w^e_c\Big(v_{e_B}(t) + v_{\mathrm{AI}} + \lambda^B_{w_r}(v_A(t), v_B(t), t)\Big)p_c N_{e_B}, \\ \tilde{v}_{i_A}(t) = v_{i_A}(t) &+ \lambda^A_{w_r}(v_A(t), v_B(t)) \\ &+ w^i_c\Big(v_{e_B}(t) + v_{\mathrm{AI}} + \lambda^B_{w_r}(v_A(t), v_B(t), t)\Big)p_c N_{e_B}, \\ \tilde{v}_{e_B}(t) = v_{e_B}(t) &+ v_{\mathrm{AI}} + \lambda^B_{w_r}(v_A(t), v_B(t)) \\ &+ w^e_c\Big(v_{e_A}(t) + v_{\mathrm{AI}} + \lambda^A_{w_r}(v_A(t), v_B(t))\Big)p_c N_{e_A}, \\ \tilde{v}_{i_B}(t) = v_{i_B}(t) &+ \lambda^B_{w_r}(v_A(t), v_B(t)) \\ &+ w^i_c\Big(v_{e_A}(t) + v_{\mathrm{AI}} + \lambda^A_{w_r}(v_A(t), v_B(t), t)\Big)p_c N_{e_B}. \end{aligned} \tag{3}$$

Here the time dependency is explicitly denoted and the terms with no explicit time variable are constants.

2.2 Regulatory module

Regulatory module weighs the stimuli thus promotes one of them by making it more likely that the pool voting for the promoted stimulus has higher excitatory firing rate compared to the excitatory population of the other pool. Consequently, the pool with higher firing rate wins the bicolumnar competition and makes the decision. We name the time instant at which the decision is made as reaction time. It is defined as the instant at which the difference between the firing rates of the excitatory populations of two pools exceeds a prefixed threshold (see Appendix).

The regulatory module evolves during the ith trial of the Eth episode by the following:

$$\begin{cases} \tau_\psi \frac{d\psi^E_i(t)}{dt} = -4\psi^E_i(t)\Big(\psi^E_i(t) - 1\Big)\Big(\psi^E_i(t) - 1/2\Big) + \frac{\sigma}{(c_0 t)^2}\zeta_i, \qquad t \in (0, t_F], \\ \psi^E_i(0) = \phi^{E-1}_i, \end{cases} \tag{4}$$

where τ_ψ is the time scale parameter, ψ is the output of the regulatory pool, and $t_F > 0$ is the final time. This system evolves independently for each trial. Here $\zeta_i = \zeta_i(t)$ is white Gaussian noise whose intensity level is scaled version of $\sigma > 0$, and $c_0 > 0$ is a constant. Noise term ζ_i introduces a strong stochastic behavior initially, then it decays in time. This noise models the exploratory behavior of the subject.

Evolution process given by (4) is restarted at the beginning of each trial of the Eth episode from the initial condition fixed to the value determined by the reward (or loss) ϕ_i^E corresponding to the same trial but of the $(E-1)$th episode. In this way, the reward mechanism provides for each trial a feedback to the regulatory module at the end of the $(E-1)$th episode. This feedback determines towards which decision the regulatory module will introduce bias to the basic module in the Eth episode. Final time t_F appearing in (4) is the final time of the trial and it is the same for every trial.

Regulatory module dynamics is transmitted to Pool A and Pool B via

$$\begin{aligned} \lambda_{w_r}^A(v_A(t), v_B(t), t) =& \psi\, v_A(t) + w_r(1-\psi(t))\, v_B(t), \\ \lambda_{w_r}^B(v_A(t), v_B(t), t) =& \psi\, v_B(t) + w_r(1-\psi(t))\, v_A(t), \end{aligned} \tag{5}$$

where w_r is a constant, which is fixed to 1 in our setting.

At each trial, both stimuli are shown to the subject and they have an excitatory effect on all the populations. The decision depends on the evolution of ψ function, which converges to either 1 promoting the decision for the large stimulus, or to 0 promoting the decision for the small stimulus.

2.3 Reward mechanism

Reward mechanism introduces online learning to the model. It allows the model to capture the strategy maximizing the gain in each episode. Once the strategy is learned, the system makes the decisions in the way which maximizes the final gain, which is the total reward of one episode.

The reward mechanism is updated through a discrete evolution where the temporal variable is the trial number. In other words, the reward function value remains constant during a trial and it is updated at the end of the trial. The updated value is fed as the initial condition to regulatory function ψ in the trial corresponding to the episode coming after; see (4). We use the notation M_i^E to denote the mean value of the stimuli corresponding to the ith trial of the Eth episode. At the end of the episode, e.g., with $i = 2$ in Horizon 1, M_3^E refers to the mean value updated by the final reward (or loss). We write N to express the number of trials in one episode. Then we write the evolution for the reward function ϕ as follows:

$$\begin{cases} \phi_i^{E+1} = \phi_i^E + k\,(M_{i+1}^E - M_i^E)(2\psi_i^E(T) - 1)(\phi_i^E - 1)^2(\phi_i^E)^2, \\ \phi_i^0 = C, \quad \text{for all} \quad i \in \{0, 1, \ldots, N-1\}, \end{cases} \tag{6}$$

with C denoting a constant, which is fixed to 0.5 in our framework. This system initiates the reward mechanism for each trial separately, yet the trials are not independent due to the coupling effect arising from $(M_{i+1}^E - M_i^E)$ factor in (6).

Finally, the only difference between the frameworks which we use in human and macaque simulations appears in (6). In the case of macaque, we assign to each stimulus, the number of water drops corresponding to the reward of the stimulus. This allows us to translate each symbol to a numerical value, and a value associated to the corresponding reward. We introduce a constant: maximum reward α. We fix α to 8, which is the maximum possible number of water drops that the macaque can receive in one episode. Maximum reward and its half value allow to change the value towards which ψ converges in Trials 1 and 2, respectively, each time the model follows a scenario different from the maximum performance scenario. More precisely, we replace the first line of (6) with

$$\phi_i^{E+1} = \phi_i^E + k\,F_i\,(2\psi_i^E(T) - 1)(\phi_i^E - 1)^2(\phi_i^E)^2, \tag{7}$$

where

$$F_i^E := (\delta_{i-1}(\text{other}_1^E - \text{choice}_1^E) + \delta_{i-2}(\text{choice}_2^E - \text{other}_2^E)), \tag{8}$$

with i and δ denoting the trial number and Dirac delta, respectively. Here choice_i^E and other_i^E denote the numbers of drops given as rewards for the chosen and the other stimuli in the ith trial of the Eth episode, respectively.

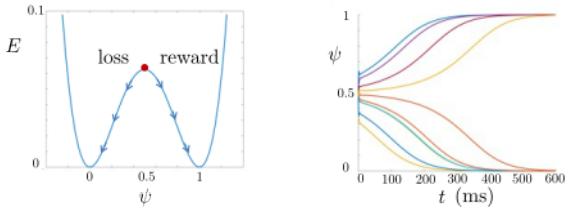

Figure 3: Left: Energy functional of ψ. Neutral initial condition is $\psi(0) = 0.5$. Reward and loss introduce a bias making $\psi(0) > 0.5$ and $\psi(0) < 0.5$ in the next episode, respectively. This results in $\psi = 1$ and $\psi = 0$ at the end of the corresponding trial of the next epsiode, resulting in keeping the same decision as in the previous episode or changing the decision, respectively. Right: Time evolution of ψ starting from different initial conditions. At the beginning, small noise is introduced to model the exploratory behavior and the perceptual difficulties.

3 Usage notes

The Jupyter notebook provided here is self-explained. Please read the explanations in the notebook to use the model, change the parameters and to run the simulations. The notebook requires the built-in Matplotlib, Nump, Random, Os and Scipy.io libraries. It is generated by Python 3.8.5. Contact: *emre.baspinar@cnrs.fr.*

References

[1] M. Di Volo, A. Romagnoni, C. Capone, and A. Destexhe, "Biologically realistic mean-field models of conductance-based networks of spiking neurons with adaptation," *Neural Computation*, vol. 31, no. 4, pp. 653–680, 2019.

[2] E. Baspinar, G. Cecchini, M. DePass, M. Andujar, P. Pani, S. Ferraina, R. Moreno-Bote, I. Cos, and A. Destexhe, "A biologically plausible decision-making model based on interacting cortical columns," *bioRxiv preprint 2023.*