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

Emre Baspinar[1,*], Gloria Cecchini[2,3], Michael DePass[3], Marta Andujar[4], Pierpaolo Pani[4], Stefano Ferraina[4], Rubén Moreno-Bote[3,5], Ignasi Cos[2,5], Alain Destexhe[1]

1 Paris-Saclay University, CNRS, NeuroPSI, Saclay, France
2 Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Catalonia, Spain
3 Center for Brain and Cognition, DTIC, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
4 Department of Physiology and Pharmacology, Sapienza University of Rome, Rome, Italy
5 Serra-Hunter Fellow Programme, Barcelona, Catalonia, Spain

* Corresponding author: emre.baspinar@inria.fr

Abstract

We propose a new AdEx mean-field framework to model two networks of excitatory and inhibitory neurons, representing two cortical columns. The columns are interconnected with excitatory connections contacting both Regularly Spiking (excitatory) and Fast Spiking (inhibitory) cells. The model is biophysically plausible since it is based on intercolumnar excitation modeling the long range connections and intracolumnar excitation-inhibition modeling the short range connections. This configuration introduces a bicolumnar competition, sufficient for choosing between two alternatives. Each column represents a pool of neurons voting for one of the two alternatives indicated by two stimuli presented on a monitor in human and macaque experiments. We endow the model with a reward-driven learning mechanism which allows to capture the optimal strategy maximizing the cumulative reward, as well as to model the exploratory behavior of the participant. We compare the simulation results to the behavioral data obtained from the human and macaque experiments in terms of performance and reaction time. This model provides a biophysical ground for simpler phenomenological models proposed for similar decision-making tasks and can be applied to neurophysiological data. Finally, it can be embedded in whole-brain simulators, such as The Virtual Brain (TVB), to study decision-making in terms of large scale brain dynamics.

1 Introduction

Decision-making refers to selecting between given alternatives by taking into account the consequences of each choice. To achieve it, the brain employs neural mechanisms which encode and interpret sensory stimuli, weight evidence to select between choice alternatives, and finally, generates oriented motor actions. In parallel to the neural mechanisms, cognitive mechanisms related to learning, memory and risk assessment, modulate the dynamics produced by the neural mechanisms to provide the optimal compromise between gains and risks of the choices. Our aim is to model these neural and cognitive mechanisms in a biophysically plausible manner in the context of a consequence-based decision-making task [12].

There have been proposed several experimental procedures by neuroscientists to unveil the neurological processes underlying decision-making. On the side of psychophysics, behavioral data was sampled while the participant was performing decision-making tasks, such as distinguishing between different stimuli and reaching one of them based on a certain decision-making strategy [13, 33]. On the neurophysiology side, electrophysiological measurements provided single neuron or neural population activity in specific brain areas. This could be used to study the link between these areas, functions, and the behavioral variables [18, 37, 41, 43, 47].

These studies indicated that the decision-making is a complex task involving several cortical and subcortical regions [8, 26, 42, 45]. Among these areas, the prefrontal cortex (PFC) has been

traditionally considered as the key area involved in decision-making in the brain [6,16,22,27,44,50]. We focus on a biophysically inspired modeling setting which achieves to produce the decision-making dynamics and its modulation based on the learning via reward. For this reason, our model does not take into account the subcortical regions and it focuses on the cortical regions (in particular PFC) relevant to the decision-making neural dynamics.

Decision-making relevant PFC regions can be classified into two functional categories: benefit and cost. Our everyday routine is based on analyzing, judging and evaluating the advantages and disadvantages of the choices under different circumstances. During all these processes, the brain areas responsible for benefit and cost estimation are activated [38, 51, 52]. Consequently, a decision is made either to perform a task or to avoid it. We will take into account these two functional categories as the main ingredients training our model to identify the most rewarding decisions.

Several neural population models have been introduced so far for similar decision-making mechanisms [1, 28, 39]. In [9], a recurrent network model of object working memory was proposed. Thereafter, another network model with similar architecture was proposed for a visual discrimination task in [55]. This final architecture was adapted to a two-choice perceptual task, where the effect of memory-based experience on the decisions was included [33]. All these models are based on the connectivity architectures which are not in accordance with the anatomy, which imposes the cortico-cortical connections. These long range connections originate from the pyramidal (excitatory) cells [24, 34], and they make the synapses onto the other pyramidal (excitatory) cells as well as onto the interneuron (inhibitory) cells [30, 35]. To tackle this point, we start from the model topologies presented in [33, 55] and modify them such that this long range cortico-cortical connectivity is taken into account in our model.

Reliable neural responses to stimuli are generally observed at the population level in neurophysiological experiments. In other words, the information processing and the produced response do not correspond to a single neuron but rather to a population activity. Neural response is obtained as averaged time-integrated activity of the individual dynamics of the neurons in interaction within the population. A neural population can be modeled by using a network which consists of a number of stochastic or ordinary differential equations. Averaged network behavior provides the population behavior. This requires high computational power and it is challenging to apply analytical approaches on the network since the network is high dimensional [57]. Alternatively, one can consider the averaged network behavior at the coarse-grained continuum limit. This asymptotic limit of the network can be written in terms of the probability distribution of the state variables describing the neural dynamics in the network. This asymptotic limit is the so-called *mean-field limit* [2, 19, 56].

Our model was extended from the Adaptive Exponential (AdEx) mean-field framework [15, 58]. This mean-field framework approximates closely the neural population behavior modeled by the AdEx network [7]. It is low dimensional, simpler and easier to analyze compared to the AdEx network, yet it approximates closely the network dynamics, motivating our choice of model. Moreover, in the case of the cerebral cortex, the AdEx mean-field framework models the population of two neuron types: Regular Spiking (RS) neurons, displaying spike-frequency adaptation as observed in the pyramidal neurons, and Fast Spiking (FS) neurons, with no adaptation, as observed in the interneurons. This constitutes the origin of the biophysical nature of our model.

Our model consists of three structures: basic module, regulatory module, reward module. The basic module models two distant columns of the prefrontal cortex, where each column is represented by the extended AdEx mean-field equations. We restrict ourselves to the scenario in which the decision results from the competition between (at least) two columns, where each one of them votes in favor of either one of the two alternatives. The winning column determines the decision. This justifies the connectivity of our model, which is based on the long range excitatory connections between the columns as illustrated in Figure 3. The basic module produces the neural dynamics, which can be due to a stimulation or in the absence of any stimulation (spontaneuous brain activity). We focus on the stimulated activity. The basic module is not associated with any cognitive context. To introduce the cognitive context, we integrate the basic module with the regulatory and reward modules. These two modules identify the correspondence of reward to the made choices and optimize the decision-making strategy such that the cumulative reward is maximized. This optimization of the strategy relies on a learning based on the rewards received by the model as the choices are made. The regulatory module is based on a function with two

attractors which was adapted from [12]. It introduces the bias to the stimuli provided to the basic module. It can be thought of as a gating mechanism arranging the motor-plan flexibility for the response to the stimuli (or the inputs evoked by the stimuli). This mechanism might provide an explanation to the observations showing that relevant sensory inputs, distractors, and direct perturbations of decision-making circuits affect the behavior more strongly when they are introduced in the early phase of the stimulation [20, 29, 31, 49, 54, 59]. Finally, the reward module is based on a reinforcement learning type equation which motivates the choices providing more reward and demotivates the decisions providing less or no reward. This is associated with the increased dopamine signaling during predicted reward anticipation: Motivation to repeat the same decision, as well as the dopamine signaling, increases if the reward is as predicted or more [36, 48]. This results in reward-based learning.

Two parameters characterize the behavioral performance of our model: learning speed and flexibility parameter. The former determines how quickly the model captures the optimal strategy maximizing the cumulative reward. It can be thought of as the modulation effects of acetylcoline and dopamine on the learning dynamics [10, 11]. The latter determines how much the model is flexible to make mistakes, or to try different decision-making strategies to explore, sometimes even after it learns the optimal strategy.

Challenges are at both mathematical and numerical levels. Firstly, the model dimension is increased compared to the classical AdEx mean-field framework [15, 58] since we consider two cortical columns. Moreover, the intercolumnar connections introduce the derivatives of cross correlations between the population firing rates corresponding to these two different columns. This demands high memory and computational power. Secondly, noise should be incorporated in the model at the level of both neural dynamics and cognitive layer. These two types of noise should be introduced in a balanced manner such that they do not dominate the effects of each other. Finally, the increased number of parameters makes it difficult to fit them to the experimental behavior data.

The model reproduces the behavioral results obtained from the experiments conducted on humans and macaques. The novelties of the model are in terms of both connectivity and structure as explained in Discussion. The implementation of the model is provided in Python as a Jupyter notebook [5], which can be simulated online on EBRAINS as well [4].

2 Experiment setup

Decisions are made by taking into account immediate and longer term consequences. In many cases, making decision with a consideration of stronger benefits in the long term is important for survival. This requires a resistance to immediate reward and a control over future planning to identify the optimal decision making strategy. To study the behavioral background of such mechanisms in human and macaque, an experiment protocol was designed for each species [12, 14, 21]. The protocol differs in terms of the forms of the stimuli and reward between the two species. The goal for both species is to obtain the maximum possible cumulative reward, which is proportional to the sum of all rewards obtained throughout the experiment. This requires: (i) to identify the reward values and the correspondence of these values to the made decisions, (ii) to capture the preset strategy in the task to make the optimal sequence of decisions providing the maximum cumulative reward.

2.1 Human experiment

The human experiments were conducted at Universitat de Barcelona, Facultat de Matemàtiques i Informàtica. The task was run with 28 participants. We focus on one participant as an example to fit our model, however the model can be fitted to every participant. The participant which we considered is a 20-year-old healthy female. The relevant data is accessible on EBRAINS [14]. Here we provide a summary of the task design and refer to [12] for details.

Two types of experiments are considered: Horizon 1 and Horizon 0. Each experiment is composed of K episodes, with K denoting the total number of episodes in the experiment. In Horizon 1, each episode is composed of two successive trials, which are named as Trial 1 and Trial 2. In Horizon 0, each episode is composed of a single trial. Behavioral results are quantified in

terms of two metrics: performance and reaction time. Performance provides a measure of the overlap between the decisions made by the participant and the preset strategy. It is registered per episode. Reaction time shows how long it takes for the participant to make the decision in each trial. It is registered per trial. If there is an overlap between the decision of the participant in an episode and the preset strategy, a reward is provided. The goal of the participant is to obtain at the end of the experiment the highest possible cumulative reward, which is maximum if and only if the participant obtains the maximum performance in every episode of the corresponding experiment.

The participant is instructed to choose one of the two stimuli in each trial. Each stimulus is a partially filled vertical bar; see Figure 1a. The percentage of the filled part is different for each bar. The preset strategy in Horizon 1 experiments is to choose the smaller stimulus in Trial 1 and the larger stimulus in Trial 2. We increase the filled parts of both bars with the same amount at the end of Trial 1 if the choice of the participant is in coherence with the preset strategy. Otherwise, we decrease the filled parts with the same amount. This amount is called gain. It is generated at the beginning of the experiment and kept fixed throughout the whole experiment. In Horizon 0, the preset strategy boils down to choosing the larger stimulus in each episode.

In Figure 1a, we provide an example scenario of one episode of Horizon 1. The amounts of the filled parts of the bars are $M \pm d/2$ at the beginning of Trial 1. Here M is randomly and independently generated from a uniform distribution at the beginning of each episode. Five different values of d are used in the experiments. Each one of them is fixed during $K/5$ episodes of the experiment. Here M denotes the mean value of the stimuli, and d determines the difference between the two stimuli. In Horizon 1, if the participant chooses the small stimulus in Trial 1, the gain G is added to both stimuli and $M + G \pm d/2$ become the filled quantities of the two stimuli in Trial 2. Otherwise, the gain G is subtracted from the stimuli and the filled quantities become $M - G \pm d/2$ in Trial 2. The same procedure is applied in Trial 2 but now the expected choice by the preset strategy is the large stimulus. Moreover, the participant does not see at the end of Trial 2 any added or subtracted gain since there is no Trial 3. This forces the participant to learn the consequence of the episode instead of learning a sequence of choices, because there is no feedback at the end of the episode. See also Figure 2 for some relevant example experiment results.

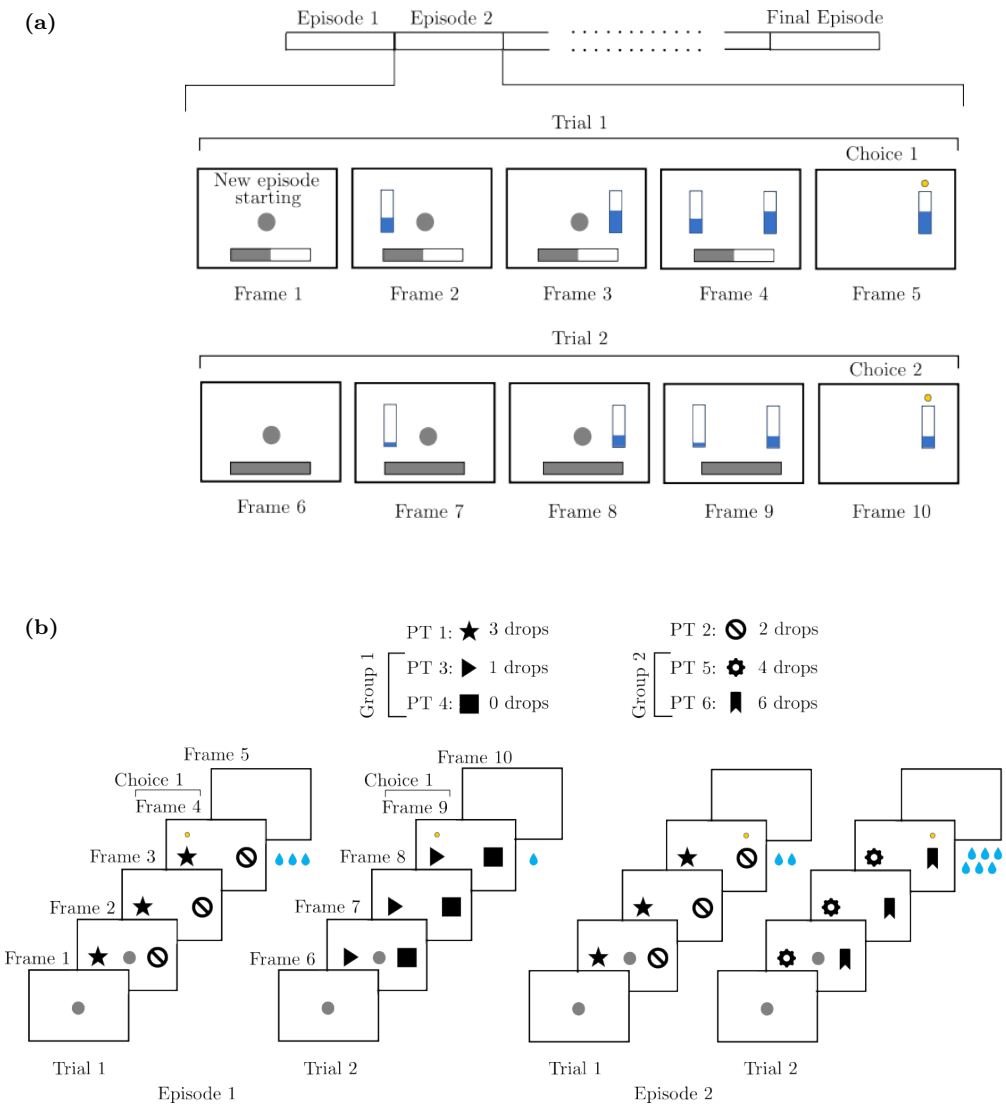

Fig 1. Example scenarios from the human and macaque experiments. **(a)** An episode of the Horizon 1 human experiment: A central target (CT) appears in Frame 1 to indicate the beginning of the episode. A progress bar at the bottom indicates in which trial the participant is. The stimuli are shown separately in Frames 2 and 3. Then the stimuli are shown together in Frame 4. The participant moves the pointer on the larger stimulus 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. The stimuli after the subtracted gain are shown separately in Frames 7 and 8. Finally, the updated stimuli are shown together in Frame 9 and the participant chooses the larger one as highlighted in Frame 10. **(b)** Two episodes of the macaque experiment. A CT appears in Frame 1 to indicate the beginning of the episode. The participant touches the CT and holds its finger on it to initiate the episode. Then the peripheral targets (PTs) are shown together on the both sides of the CT in Frame 2. The CT disappears (Go signal) as the participant starts to move its hand to choose one of those PTs as shown in Frame 3. The chosen PT (PT 1) is highlighted by the yellow dot in Frame 4. The corresponding reward (3 water drops) is provided and in Frame 5 a blank screen is shown to indicate the end of Trial 1. The same procedure is followed in Trial 2 but now with the PTs of Group 1. In the second scenario, the cumulative reward is higher since PT 2 is chosen in Trial 1 and hence the Group 2 PTs, which provide a higher reward, are shown on the monitor in Trial 2. The nomenclature and corresponding rewards to the PTs are given at the top.

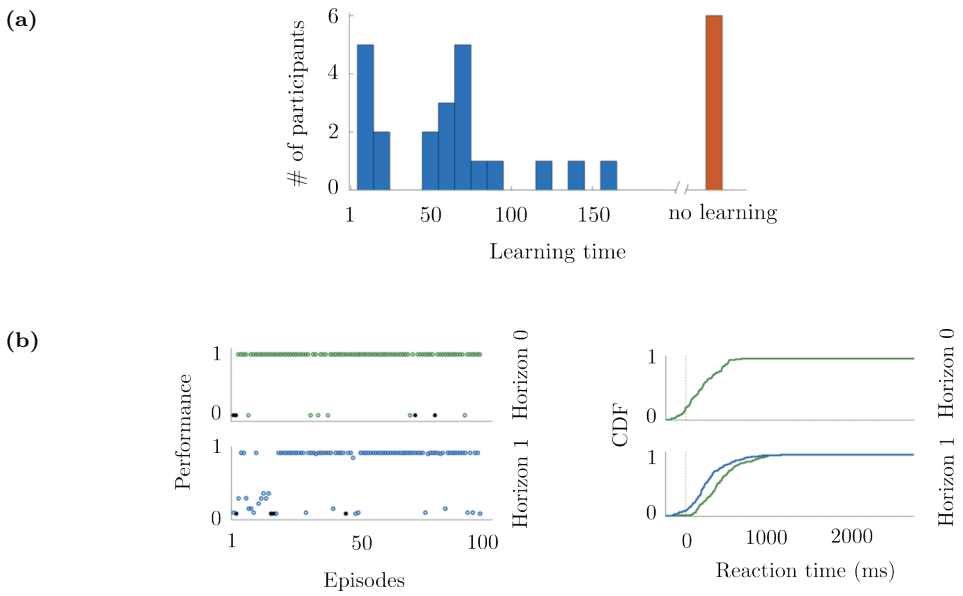

Fig 2. Some experiment results from Cecchini et al. [12]. **(a)** The histogram of the learning times obtained from 28 human subjects. The learning time is in terms of episode numbers. The long learning time is classified as no learning. **(b)** The performance of one of the participants, and the cumulative distribution function (CDF) of the reaction times of the same participant. The results are presented separately for Horizon 0 and Horizon 1, where in the latter, the CDFs of Trial 1 and Trial 2 are provided in green and blue, respectively.

2.2 Macaque experiment

The macaque experiments were conducted at Sapienza Università di Roma, Dipartimento di Fisiologia e Farmacologia. One 16-year-old, healthy, male macaque weighting 9.5–10.5 kg was trained for performing a Horizon 1 task similar to its counterpart in the human task. The relevant data can be found on EBRAINS [21].

The task can be summarized based on the two possible scenarios provided in Figure 1b. A dataset of black and white stimuli were shown to the macaque in each trial. Six stimuli (peripheral targets; PTs) were extracted randomly from *Microsoft PowerPoint* shapes library. Two of these six randomly extracted PTs were shown in Trial 1 and they were attributed to different rewards: PT 1 – 3 drops, PT 2 – 2 drops. The remaining four stimuli were separated into two groups of two PTs, Group 1 with PT 3 and PT 4, and Group 2 with PT 5 and PT 6. Each of those four PTs corresponds to a different reward: PT 3 – 1 drop, PT 4 – 0 drop, PT 5 – 4 drops and PT 6 – 6 drops. The group which is shown in Trial 2 was determined by the choice made in Trial 1. More precisely, if the participant chooses PT 1 in Trial 1, Group 1 will be shown in Trial 2 as illustrated in Episode 1 of Figure 1b. Otherwise, Group 2 will be shown in Trial 2 as illustrated in Episode 2 of Figure 1b. At the beginning of each trial, a central target (CT) is shown at the center of a touchscreen monitor to indicate the beginning of the trial. The macaque is required to touch the CT and hold its finger on it for 550 $\pm$ 50 ms to initialize the trial. Once this requirement is fulfilled, Stimuli 1 and 2 are shown as the PTs appearing on the left and right hand sides of the CT. As the participant starts to move his hand after an additional holding time ($\approx 400 - 600$ ms), the CT disappears. This is considered as the Go signal initiating the reaction time, which is tracked for a maximum duration of 2000 ms. If a choice is made between two PTs within this 2000 ms, the reward is delivered to the participant after an additional holding time of 600 ms on the chosen PT. Once this protocol for Trial 1 is completed successfully by the participant, Trial 2 begins after an inter-trial interval of 1200 ms and it follows the same protocol.

Although both human and macaque tasks are based on perceptual distinguishing of the targets, there are three major differences between them. Firstly, the stimuli are partially filled bars in the human task, whereas they are some figures chosen randomly from the Powerpoint shapes library in the macaque task. Secondly, in the human task, the higher and lower rewards are provided as the addition of the gain to the filled parts and the subtraction of the gain from the filled parts

of the stimuli, respectively. In the macaque task, the reward is provided as a certain number of water drops to the macaque. Finally, reward is provided at the end of Trial 2 as well, in the macaque task; and it depends on the choices made in both Trial 1 and Trial 2.

3 Results

3.1 Model

3.1.1 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 the inhibitory populations by i_A, i_B, where the subindices mark the corresponding pool. There are only excitatory connections across the pools. Each cross-pool excitatory connection targets both the excitatory and the inhibitory populations of the other pool. The excitatory and inhibitory populations within each pool are fully connected with recurrent and cross-population connections; see Figure 3.

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. The ratio N_i/N_e is 0.2. The weights of the intercolumnar connections are denoted by w_c^e and w_c^i, whose supindices denote the target populations. The recurrent connections have unit weights.

Base drive v_{AI} keeps the excitatory populations in the asynchronous irregular (AI) state. We fix v_{AI} to 5 Hz. The terms $\lambda^A(v_A, v_B)$ and $\lambda^B(v_A, v_B)$ represent the external inputs with a bias introduced by a regulatory module modeling plasticity as explained in the following section. The biased inputs are fed to Pools A and B through the function λ simultaneously.

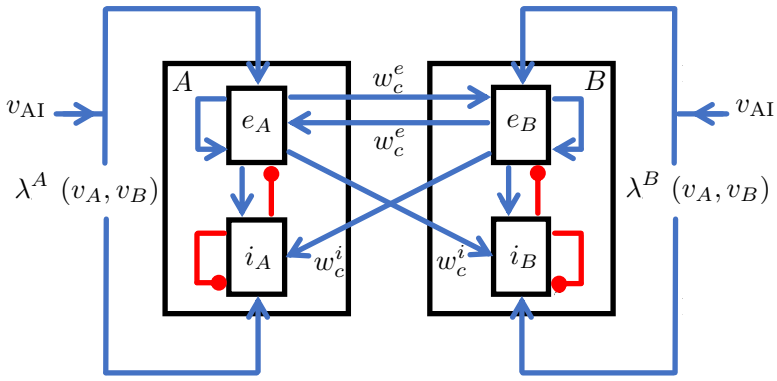

Fig 3. Basic module with two pools of excitatory and inhibitory populations. Pools A and B vote in favor of Stimulus A (v_A) and Stimulus B (v_B), respectively. The excitatory populations are denoted by e_A, e_B and the inhibitory populations are represented with i_A, i_B. The excitatory and inhibitory connections are in blue and red, respectively. The weights of the intercolumnar connections are denoted by w_c^e and w_c^i. Here λ^A and λ^B represent the outputs of the regulatory module introducing a bias to the input signals v_A and v_B. Finally, v_{AI} is the base drive, which ensures that the model performs in the AI state.

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

We assume that the derivatives of the adaptation variables with respect to the firing rates v_α are 0. This is due to the fact that the adaptation variables W_α evolve on a slow time scale and its change with respect to the firing rates v_α evolving on a 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 the noise intensity. The function $A_{\alpha\beta}$ is defined as follows [15]:

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

with N_α denoting the number of neurons in the population α. The 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 [15]. We use the same term μ_V as given in [15, 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 τ_w is a time scale-like parameter for W_α. We express the probability of the intercolumnar connectivity with $p_c \geq 0$. Finally, we write the regulated inputs $\tilde{v}_\alpha$ of the transfer functions given in (2) as follows:

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

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

3.1.2 Regulatory module

Regulatory module was adapted from [12] and it introduces a bias by weighting the stimuli in each pool. The pool voting for the promoted stimulus is more likely to have a higher excitatory firing rate than the other pool. Consequently, the pool with the higher excitatory firing rate wins the bicolumnar competition and makes the decision; see Figures 5 and 12. The instant at which the decision is made is the reaction time. It is the instant at which the difference between the excitatory firing rates exceeds a prefixed threshold (see Supporting information).

The regulatory module evolves during the ith trial of the Eth episode via

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

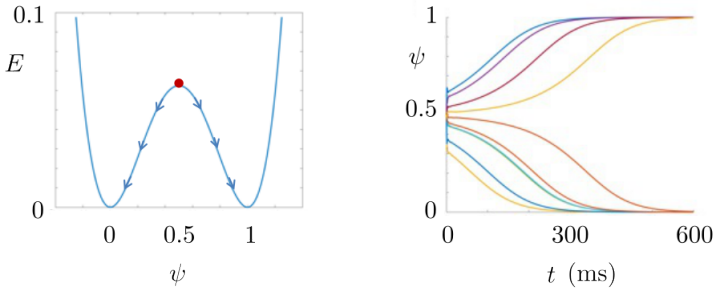

Fig 4. Regulatory module. Left: Energy functional of ψ. The neutral initial condition is $\psi(0) = 0.5$. The high and low rewards introduce a bias making $\psi(0) > 0.5$ or $\psi(0) < 0.5$ in the next episode. This results in $\psi = 1$ or $\psi = 0$ at the end of the corresponding trial of the next episode. The former introduces the bias favoring the large stimulus and the latter introduces the bias favoring the small stimulus. Right: Time evolution of ψ starting from different initial conditions. At the beginning, a small noise is introduced for modeling whatsover which might perturb the regulatory process, such as the exploratory behavior or perceptual difficulties.

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

The time evolution process given by (4) is reinitialized at the beginning of each trial of the Eth episode from the initial condition fixed to the value determined by the reward ϕ_i^E corresponding to the same trial but of the $(E-1)^{\text{th}}$ episode. In this way, the reward module provides for each trial a feedback to the regulatory module at the end of the $(E-1)^{\text{th}}$ episode. This feedback determines towards which decision the regulatory module will introduce the bias to the basic module in the Eth episode.

The dynamics of the regulatory module is transmitted to Pool A and Pool B via

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

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

3.1.3 Reward module

Reward module endows the model with online learning. It allows the model to learn the preset strategy maximizing the cumulative reward in each episode. Once the strategy is learned, the system makes the decisions in almost complete coherence with the strategy.

The reward module is updated through a discrete evolution where the temporal variable is the episode 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 the 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 gain. We write N to express the number of trials in one episode. Then we write the evolution for the reward function ϕ as follows:

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

with C denoting a constant, which is fixed to 0.5 in our framework. This system initiates the reward module 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 the 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 appearing on the monitor to a numerical value, and to a value associated with the corresponding reward. We replace the first line of (6) with

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

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.

3.1.4 External stimuli

A common choice to model the external stimuli is Heaviside function in mean-field models. However, since Heaviside functions have discontinuity due to the discrete jump, our biophysical model undergoes a transient phase which might obscure the instant marking the reaction time. For this reason, we performed simulations by using sufficiently sharp sigmoid function as external stimulus; see Figure 12a and Figure 13 in Supporting information.

3.1.5 Complete model with the reward module

We show in Figure 5 the results of our model simulation with 6 Horizon 1 episodes. The reward module is initiated from 0.5 for both trials in the first episode. We use the same parameters given in (17)-(19). The parameter d is kept constant for all episodes.

The preset strategy requires choosing the smaller stimulus (the bar which is filled less) in the first trial and the larger stimulus (the bar which is filled more) in the second trial at each episode. In Figure 5, we observe that the model explores the strategy in Episode 1. Its decision is completely random since $\psi_i^1(0) = 0.5$ for all $i \in \{1, 2\}$. We observe that in Trial 1, there is no winner of the competition. In Episode 2, the reward module introduces a bias in the regulatory function ψ by feeding the reward to ψ as the initial condition at the beginning of each trial. The same procedure is repeated for every episode. We observe in Figure 5 that after Episode 2, the model learns the strategy and starting from Episode 3, it makes the choice in accordance with the strategy. The speed of learning depends on the parameter k in (6). In Figure 5, we chose $k = 0.1$ and we fixed the reward gain G to 0.75. The gain is added to the stimuli if the correct decision according to the preset strategy is made, otherwise, it is subtracted from the stimuli; see Figure 1a.

3.2 Noise sources

We model the distortion effects as an Ornstein-Uhlenbeck (OU) process which can be sampled directly from a Gaussian distribution once the initial condition of the OU process is chosen properly. This is similar to the previous framework [15], however with one difference: We choose the initial condition of the OU process as a zero mean Gaussian, where its variance is scaled by the convergence rate of the OU process as given below. This allows to avoid explicit simulation of the OU process.

Formally, the noise in mean-field models is a stochastic process evolving in time independently of the rest of the state variables. For a white Gaussian noise, this evolution can be written as an OU process of the following type:

$$\begin{aligned} d\omega_\alpha(t) =& -\theta_\alpha \omega_\alpha(t)dt + \sigma dW_\alpha(t), \\ \omega_\alpha(0) =& \Omega, \end{aligned} \tag{9}$$

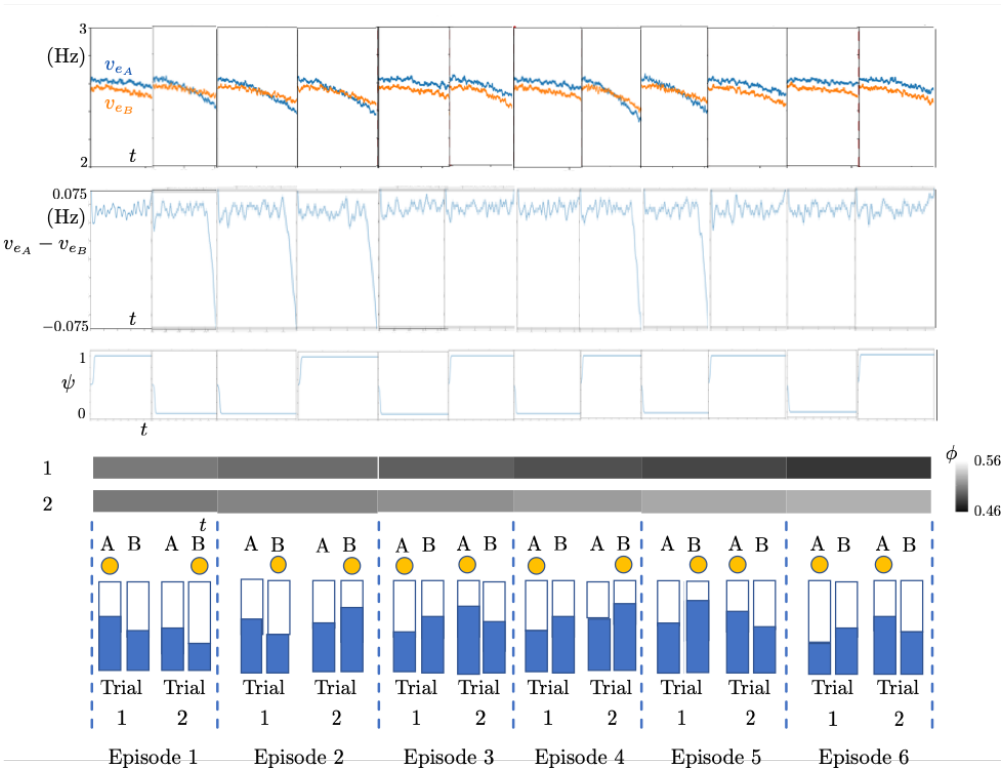

Fig 5. Simulation results of 6 episodes. First row: Time courses of the excitatory population firing rates. Second row: Time courses of the difference of the excitatory population firing rates. Third row: Time course of the regulatory function ψ. Fourth row: Time course of the reward function ϕ for each trial of the corresponding episode shown at the bottom row. Bottom row: External stimuli. Chosen stimuli are highlighted by the yellow dot at the top. Trial and episode numbers are given at the bottom.

where $W_\alpha(t)$ is a standard Brownian motion, θ_α is a positive constant associated to the convergence rate of the process to its long term mean, and Ω is the initial condition, which is a centered Gaussian noise with variance $\sigma_0^2 > 0$. This is a linear stochastic differential equation and its unique solution [3] is

$$\omega_\alpha(t) = \mathrm{e}^{-\theta_\alpha t}\,\Omega + \sigma \int_0^t \mathrm{e}^{-\alpha(t-s)}\, dW_\alpha(s), \tag{10}$$

where the solution $\omega_\alpha(t)$ is a Gaussian process. It can be written for a small time step $0 < h \ll 1$ as

$$\omega_\alpha(t+h) = \mathrm{e}^{-\theta_\alpha h}\,\omega_\alpha(t) + \xi(t), \tag{11}$$

where $\xi(t)$ is a centered Gaussian white noise with variance $\sigma_\xi^2 = \frac{\sigma^2}{2\theta_\alpha}(1 - \mathrm{e}^{-2\theta_\alpha h})$, and which is sampled independently for each t and population α. We observe that $\omega_\alpha(t+h)$ is equivalent to a zero mean Gaussian with variance

$$\sigma_{\omega_\alpha}^2(t+h) = \mathrm{e}^{-2\theta_\alpha h}\,\sigma_{\omega_\alpha}(t)^2 + \sigma_\xi^2. \tag{12}$$

Once we choose $\sigma_{\omega_\alpha}^2(0) = \sigma_0^2 = \frac{\sigma^2}{2\theta_\alpha}$, we observe that $\sigma_{\omega_\alpha}^2(t+h)$ is time independent and it is equal to $\frac{\sigma^2}{2\theta_\alpha}$. This allows us to generate $\omega_\alpha(t)$ from the Gaussian distribution $\mathcal{N}(0, \frac{\sigma^2}{2\theta_\alpha})$ independently at each instant $t > 0$, avoiding explicit forward time simulations of the OU process given in (9). This is the idea behind introducing the noise terms ω_α in the mean-field system (1) and the noise term ζ_i appearing in the regulatory mechanism (4) as Gaussian white noise sampled directly from a Gaussian distribution at each time and for each population, but not as an explicit OU process evolving in time.

3.2.1 Extracellular media distortion effects

In the previous AdEx mean-field model [15], the background noise was introduced as an additive noise to the base drive potential v_{AI}. This noise was evolving as an OU process and it modeled dynamically the distortion effects appearing in the firing rates due to the extracellular medium and synaptic perturbations.

In our model, we introduce the white Gaussian noise ω_α scaled by the noise intensity σ explicitly in the firing rate equations as shown in (6). In this way, we avoid that the noise undergoes the nonlinear effects of the transfer functions since now it is additive in the model equations.

3.2.2 Exploratory behavior

We quantify the coherence between the choices of the participant and the preset strategy in terms of performance values. In each episode, the maximum performance refers to the case in which the participant makes the choice in coherence with the preset strategy in every trial of the episode. The maximum performance value is 1. The minimum performance refers to the case in which the participant makes the incoherent choice in every trial of the episode. The minimum performance value is 0. All other cases with partial coherence are scaled between 0 and 1; see (13).

One of the typical behaviors in the decision-making task is that the participant makes decisions which do not comply with the preset strategy even after the participant learned the strategy. This behavioral pattern is observed as irregular and rather sparse deviations from the maximum performance value. That is, we observe few performance results which are less than 1 within the blocks of maximum performance results; see Figure 11 in Supporting information. This behavior occurs due to two reasons: (i) perceptual difficulty in the human experiment, i.e., the difference between stimuli is small, thus the participant makes the wrong decision as a result of perceptual difficulty; (ii) the participant is willing to explore whatsoever related to the experiment setup, and this might happen in both human and macaque experiments.

The perceptual difficulty is due to the stimuli, and it does not require any additional mechanism to be represented in the model. On the other side, the exploratory will of the participant is a part of the cognitive framework. We model it via the Gaussian white noise ζ_i introduced in the regulatory mechanism (4). This noise is sampled independently at each time instant, and noise level σ is scaled by $(c_0 t^2)$, where $c_0 > 0$. Here t denotes the time instant and it is reinitialized at the beginning of each trial. This scaling term introduces a strong initial perturbation to the bicolumnar competition. This perturbation decays in time. This decay is needed, since, otherwise the noise could dominate the whole bicolumnar competition instead of perturbing only the initial bias. The parameter c_0 determines how strong the initial perturbation is. The regulatory mechanism therefore, is not only for the online reward-driven learning of the preset strategy but also for providing a natural behavior which is open to making wrong decisions. This is similar to what was observed in both human and macaque experiments.

3.3 Quantification of behavior

3.3.1 Global measures

The performance deviations arising from the aforementioned exploratory behavior and perceptual difficulties are momentary, they are locally concentrated around certain episodes appearing after the participant learned the preset strategy. Therefore, the quantification of the behavioral performance requires global measures which are robust to such local deviations and which can quantify these deviations as well. For this, we use three objects, which are highlighted in Figure 11; see Supporting information. They are defined as follows:

Definition (*Performance deviation*): A performance deviation is a point with a performance lower than the maximum performance value ($= 1$) in the set of performance samples, except for the first sample.

Definition (*Deviation cluster*): A deviation cluster is a set of at least 3 successive performance deviations with respect to the episode numbers.

Definition (*Performance cluster*): Performance cluster is the set of performance samples in which there is no deviation cluster and whose last performance sample is also the last sample of the whole experiment.

Definition (*In/out-cluster deviation*): A performance deviation is called *in-cluster* if it occurs in a performance cluster, and *out-cluster* otherwise.

A cluster starts at the episode where the participant begins to make decisions coherently with the preset strategy. Within the cluster, the participant makes coherent decisions almost in each episode until the end of the cluster. Almost; because the participant might make decisions in the aforementioned exploratory manner, producing in-cluster deviations. This does not break the performance cluster as long as in-cluster deviations do not constitute a deviation cluster.

3.3.2 Characterization of the model behavioral performance

Learning speed determines how quickly the model learns the preset strategy. The higher it is, the faster the model identifies the strategy. Flexibility parameter determines how much the model deviates from the preset strategy after it learns the strategy. It models the exploratory behavior and the deviations caused by perceptual difficulties. Finally, the gain rescales the learning speed. Here we provide the effects of these three parameters on the model behavior performance.

Initial stimuli in the model are $M_0^E \pm \frac{d}{2}$, where M_0^E is generated from a uniform distribution independently for each episode and d is a constant. Here d determines the difficulty of the trial. If d is small, then it is more difficult to make a distinction between two stimuli. The value of d remains constant throughout one episode, however, this value need not necessarily be the same for different episodes.

We show in Figure 6a the performance plots of three cases in which the learning speed k appearing in (6) varies from 0.1 to 0.3. We denote by V_i^E the value of the chosen stimulus at the end of the ith trial of the Eth episode. We express as $V_{\min}^E$ and $V_{\max}^E$ the maximum and minimum values, respectively, which $\sum_{i=1}^{N} V_i^E$ can attain among all possible scenarios of the Eth episode, with N denoting the number of trials in the episode. We measure the performance of the model in the Eth episode by using

$$\text{Performance}(E) := \frac{\sum_{i=1}^{N} V_i^E - V_{\min}^E}{V_{\max}^E - V_{\min}^E}. \tag{13}$$

We observe in Figure 6a that as we increase k, the system captures the strategy earlier, and makes constantly right decisions thereafter, with very few deviations.

The gain G is another factor affecting the learning speed. The learning speed decreases as the gain decreases since the gain G is equal to the multiplying factor $(M_{i+1}^E - M_i^E)$ of the learning speed k in (6). As shown in Figure 6b, the model learns the preset strategy faster as we increase the gain.

Finally, the flexibility parameter c_0 determines the number of deviations in the model performance results. The higher it is, the closer to the deterministic case the model is. This is due to the fact that the noise given in (4), and which produces the performance deviations, vanishes very quickly after that the regulatory variable ψ starts to evolve at the beginning of each trial; see Figure 6c. Moreover, c_0 determines the beginning of the performance cluster together with the learning speed. As c_0 increases, the performance cluster is more likely to begin from smaller episode numbers.

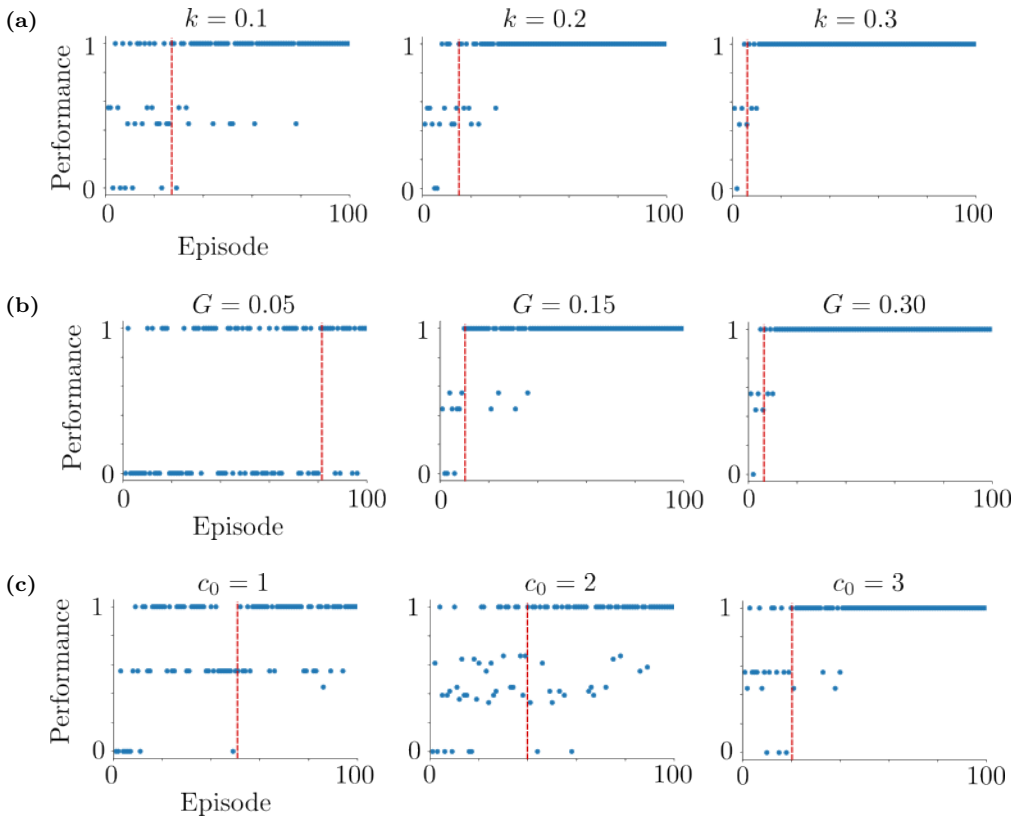

Fig 6. Simulation results of Horizon 1 performances in the human task. Initial episodes of the performance clusters are highlighted by the red vertical lines. The difficulty d is 0.2. See Supporting information for the rest of the parameters. **(a)** The learning speed k is 0.1, 0.2 and 0.3 from left to right. The flexibility parameter c_0 is 2 and the gain G is 0.3. The initial episodes of the performance clusters are episodes 24, 15 and 5 from left to right. **(b)** $G = 0.05$, $G = 0.15$ and $G = 0.3$ from left to right. Here $k = 0.3$ and $c_0 = 2$. The initial episodes of the performance clusters are episodes 81, 10 and 5 from left to right. **(c)** $c_0 = 1$, $c_0 = 2$ and $c_0 = 3$ from left to right. Here $k = 0.05$ and the gain $G = 0.3$. The initial episodes of the performance clusters are episodes 52, 40 and 20; the number of in-cluster deviations are 12, 5 and 4 from left to right.

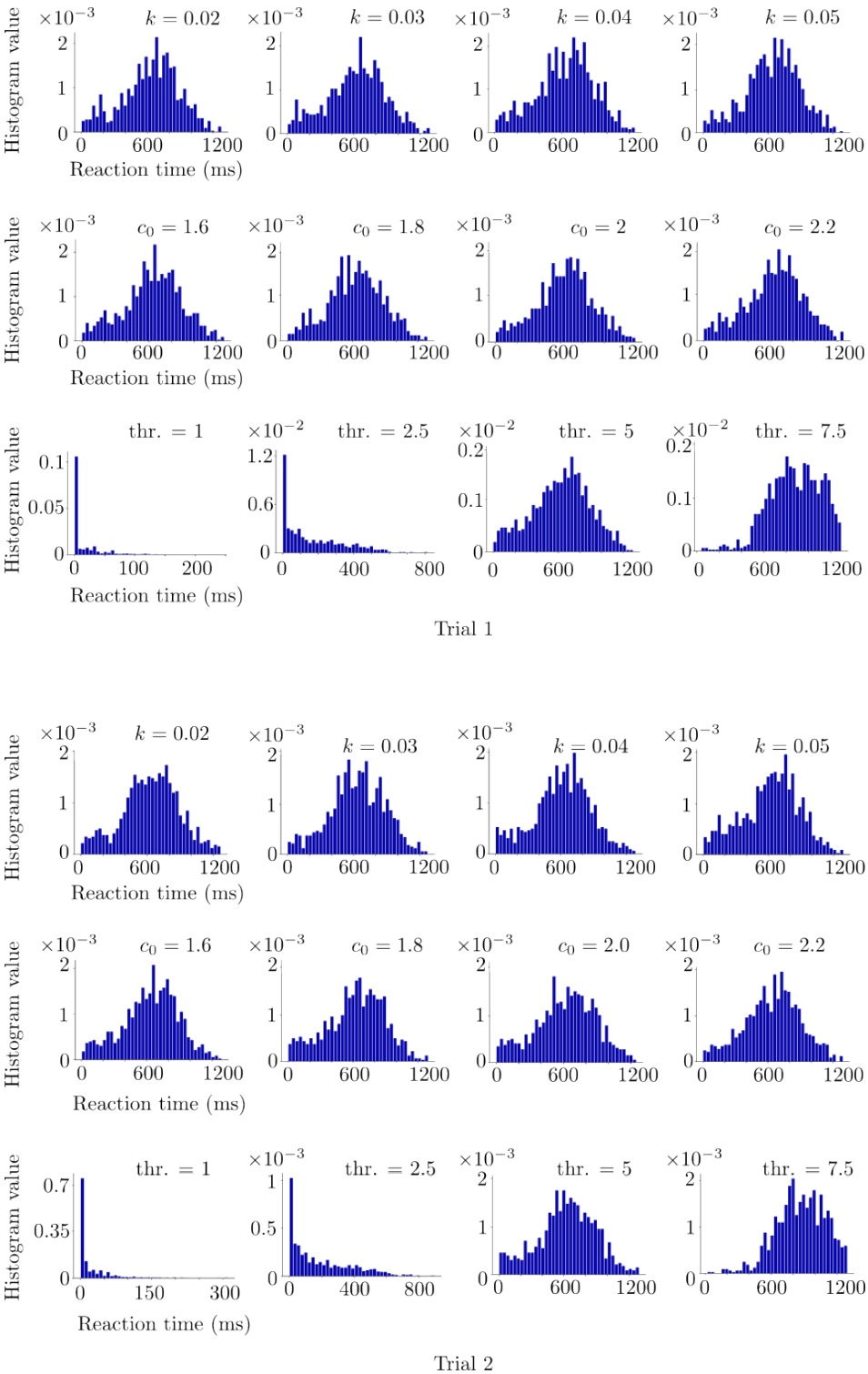

Fig 7. Model simulation results regarding Horizon 1 reaction time histograms of Trials 1 and 2. Varying the learning speed k and the flexibility parameter c_0 does not affect noticeably the distribution of the reaction times in terms of mean and standard deviation. Increasing the decision threshold increases the mean and the standard deviation of the distribution. The histograms are obtained from 10 realizations of the same simulation setup for each parameter. The parameters are the same as those of Figure 8.

3.3.3 Characterization of the model reaction time

In Figure 7, we provide the histograms of the reaction time measurements obtained from the Horizon 1 simulations. In these simulations, we varied the learning speed k and the flexibility parameter c_0, separately in Trial 1 and Trial 2. We observe that there is no distinguishing change in the histograms of the simulation results, suggesting that these two parameters do not have noticeable effects on the reaction times. This is due to the fact that the reaction times are calculated by checking the threshold crossing of the difference between the firing rates v_{e_A} and v_{e_B}. Here k and c_0 do not enter in the equations that describe the firing rates, so they do not influence the reaction times. As the decision threshold increases, the competition between v_{e_A} and v_{e_B} lasts longer, thus the reaction times increase as shown in the bottom row of Figure 7. In Supporting information, similar results for Horizon 0 simulations can be found in Figure 15.

3.4 Model simulations compared to experiments

We provide a comparison of the simulation results to the experiment results for Horizon 1. The comparisons for Horizon 0 can be found in Supporting information; see Figures 14 and 16. In Figure 8, we provide the performance results of the model and their comparison to the performance results of the human participant. Then, in Figure 9, we provide the histograms of the reaction times corresponding to the simulations and we compare them to the reaction time histograms of the human participant. Then we continue by providing the simulation performance results compared to the performance results obtained from the macaque experiment in Figure 10. The simulation results of the reaction time histograms and their comparison to the macaque histograms are given in Supporting information; see Figure 17.

3.4.1 Comparison to the Horizon 1 human performance

We provide the Horizon 1 simulation and human experiment results as quantified based on the mean and standard deviation of the initial episode numbers of performance clusters, as well as the mean and standard deviation of the in-cluster performance deviations. This quantification fully describes the learned decision-making phase of the performance results. We obtain those simulation statistics from 10 realizations of each parameter set presented in Figure 8. Each realization is composed of 100 episodes. Red horizontal lines indicate the initial cluster episode number and the number of in-cluster deviations obtained from the human experiment on the top and bottom rows, respectively. We provide the Horizon 0 results in the same fashion in Figure 14.

In Figure 8, we observe that as k and c_0 increase, the initial episode number of performance clusters decreases as shown on the left column. This is the case also in Horizon 0 as seen in Figure 14. However, differently from Horizon 0, the decreasing curves in Horizon 1 are concave, meaning that the decrease rate of the mean initial episode number of performance clusters increases. We observe that this cannot be compensated by the number of in-cluster deviations, which increases as k and c_0 increase. This is different from the Horizon 0 simulations. We observe that the model can produce close results to the experimental values once k and c_0 are chosen properly, for example $k = 0.05$ and $c_0 = 2$.

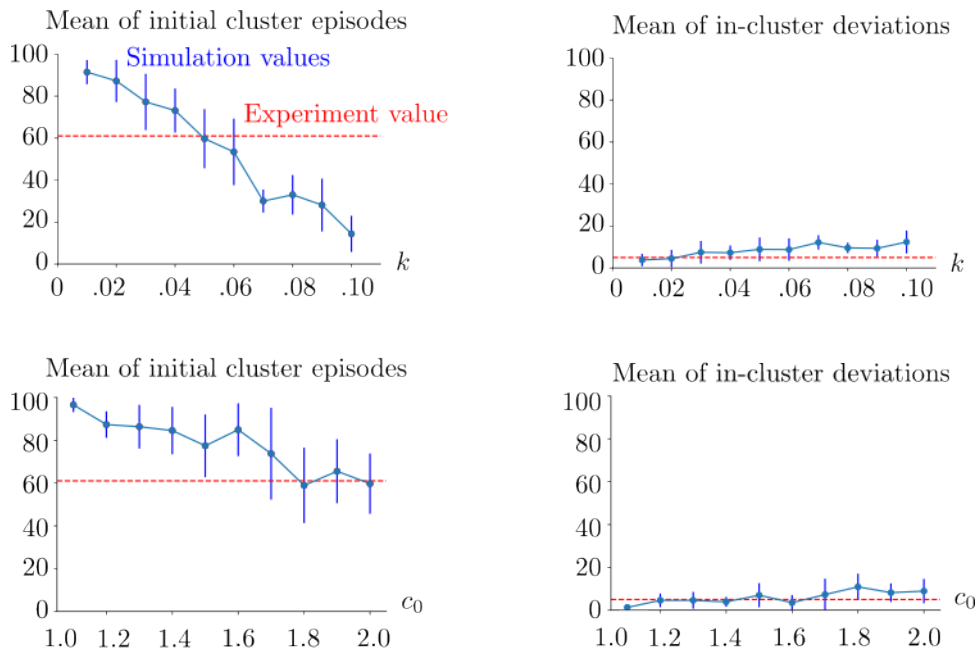

Fig 8. Horizon 1 human case. Simulation statistics with respect to the varied learning speed k in the top row, and with respect to the varied flexibility parameter c_0 in the bottom row. In the top row, $c_0 = 2$. In the bottom row, $k = 0.05$. The vertical blue lines show the standard deviations of the simulation results. The red horizontal lines indicate the initial cluster episode number and the number of in-cluster deviations obtained from the human experiment. The mean and standard deviation are obtained from 10 realizations of the same simulation setup for each (k, c_0) pair. The parameters k and c_0 can be seen as rescaling constants, therefore they are unitless. See Supporting information for the rest of the parameters.

3.4.2 Comparison to the Horizon 1 human reaction times

Reaction time in one trial is measured as the time duration between the instant when the stimuli are shown simultaneously on the monitor and the moment when the participant starts to move the pointer. In the simulations, we measure the reaction time as the time duration between the instant when the two stimuli are provided to the model and the time instant at which the decision is made, i.e., when the difference between the excitatory population firing rates v_{e_A} and v_{e_B} exceeds the decision threshold, which is fixed to 5 Hz.

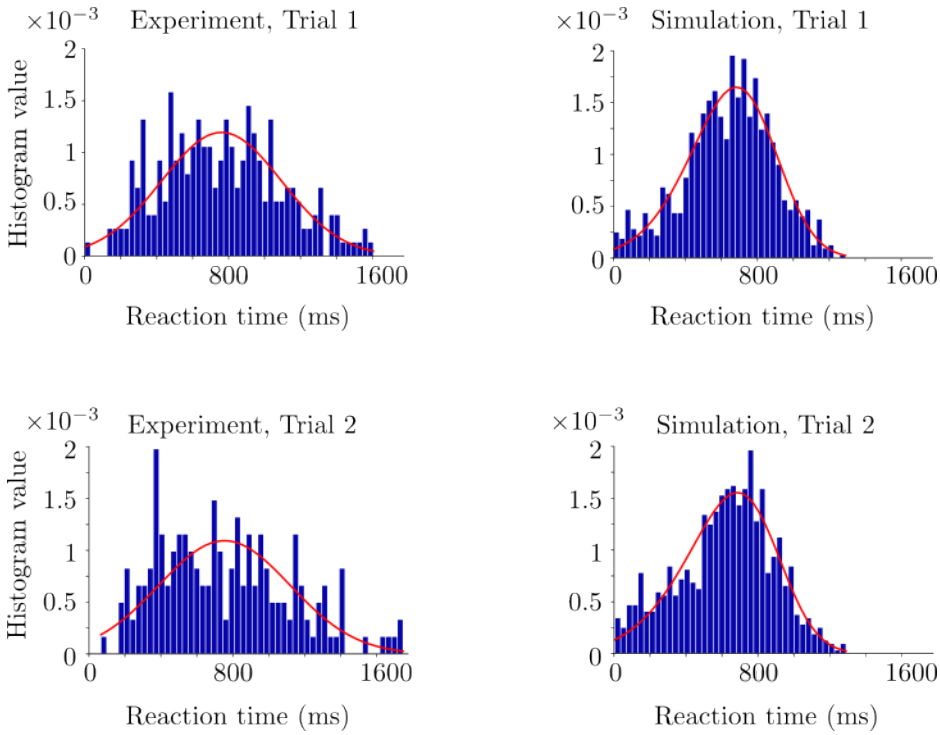

Fig 9. Horizon 1 human case. Histograms obtained from the human experiment and simulations for comparison. The red curves denote the fitted distributions via Python. Trial 1 and 2 histograms are in the top and bottom rows, respectively. The simulations are performed with $k = 0.05$, $c_0 = 2$ and decision threshold $= 0.5$.

In Figure 9, we show the simulation and experiment histograms of the reaction times corresponding to Horizon 1, together with the fitted distributions for the best performance-providing k and c_0 tested values. In Figure 9, we fit the distributions of different types to both simulation and experiment histograms. We use the *Fitter* function from the Python *Fitter* package. We consider the histograms of the reaction times obtained from the first and second trials separately. In Trial 1, skewrow distribution and hypergeometric distribution are fitted to the experiment and simulation histograms, respectively. Corresponding squared sum error values are 1.3e-5 and 9e-6 for the experiment and simulation histograms, respectively. Pearson correlation coefficient of these fitted distributions is 0.92. In Trial 2 histograms, skewrow distribution and hypergeometric distribution are fitted to the experiment and simulation histograms, respectively. Corresponding squared sum error values are 1.7e-5 and 1.1e-5 for the experiment and simulation histograms, respectively. Pearson correlation coefficient of these fitted distributions is 0.63. Trial 2 has a weaker overlap between the experiment and simulation histograms compared to Trial 1.

3.4.3 Comparison to the macaque performance

We observe that the trends in the macaque performance results are similar to the ones found in the Horizon 1 human task simulations. We see that the simulation results overlap with the macaque experiments for properly chosen parameters, for example $k = 0.12$, $c_0 = 0.106$. For a comparison based on the reaction time histograms, we refer to Figure 9 given in Supporting

information.

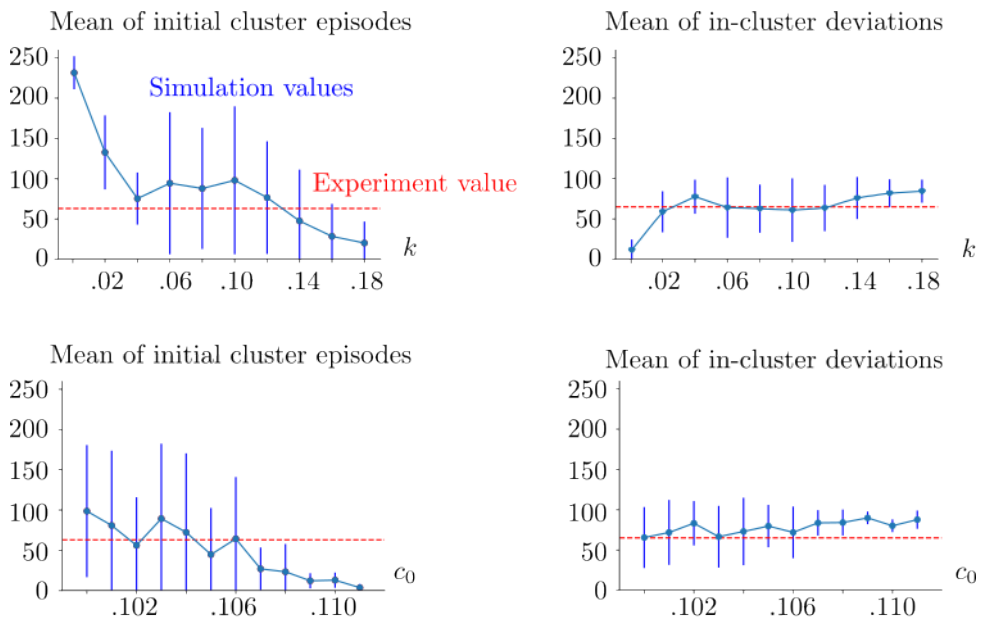

Fig 10. Macaque case. Comparison of the Horizon 1 cluster statistics to the macaque experiment with respect to the varied learning speed k in the top row, and with respect to the varied flexibility parameter c_0 in the bottom row. In the top row, $c_0 = 0.106$ and k is varied. In the bottom row, $k = 0.12$ and c_0 is varied. The vertical blue lines show the standard deviation of the corresponding statistical sample. The red dashed lines show the value obtained from the macaque experiment. The mean and standard deviations are obtained from 10 realizations of the same simulation setup for each (k, c_0) pair. The decision threshold is 5 Hz. The rest of the parameters are given in Supporting information.

4 Discussion and conclusion

We presented a biologically plausible AdEx mean-field model based on interacting cortical columns. We extended the classical AdEx mean-field model [15] to two columns with bicolumnar competition. We fitted this model to a reward-driven consequential decision-making task at behavioral level. We compared the simulation results of the model to the human and macaque experiment results in terms of behavioral performance and reaction time. Novelties of the model are at both connectivity and structure levels.

Regarding connectivity, the intercolumnar connections between the two columns are only excitatory. Two cross connections originate from the excitatory population of one column. One of these connections is onto the excitatory population and the other one is onto the inhibitory population of the other column. This connectivity overlaps with the long range cortico-cortical connections found in the prefrontal cortex and allows us to model the cortical columns as pools of the excitatory and inhibitory populations. Inhibitory control over the excitatory population is triggered by the increasing activity of the excitatory population of the other column. This is not the case in the previous models [33, 55]. In [55], there is only one inhibitory population which is connected to all excitatory populations via the long range connections. In [33], the excitatory and inhibitory populations are not explicit. They are intrinsically represented as one neural population. Two such populations are connected to each other via inhibitory connections, and recurrently connected to themselves via excitatory connections.

One of the goals of our study was to take into account the long range cortico-cortical connections in the decision-making process. These connections are necessary to provoke the bicolumnar competition in a biophysically realistic way. To achieve it, we extended rigorously the AdEx mean-field model which was proposed in [15] for AdEx network of an isolated single column with adaptation. This extension provides the bicolumnar architecture in which the columns are

connected to each other via the long range cortico-cortical connections. This allows the model to provoke the bicolumnar competition. The model chooses one of the two alternatives in accordance with the choice represented by the column which wins the competition.

On the structure side, we model each column as the mean-field limit of a pair of excitatory and inhibitory neural populations. The parameters are directly linked to the biophysical mechanisms of RS and FS cells. This allows our model to be applied to not only behavioral data but also to relevant neurophysiological data, in particular to multi-site array recordings of the macaque cortex. In this way, the model can link the phenomenological decision-making models [12, 33, 55] to the neural dynamics observed at mesoscopic level.

Moreover, we adapted the regulatory module given in [12] to the AdEx system to introduce the plasticity which learns the preset strategy in the task. This machinery is essential for the model to find the optimal decision-making strategy. The regulatory module sustains the optimal strategy once this strategy is learned. This is due to the fact that once the introduced bias provides high rewards successively over episodes, the regulatory module reinforces and retains the bias. This relates the regulatory module to working memory at phenomenological level.

Finally, the AdEx mean-field framework was never used for a cognitive task and at a behavioral level. We show here that it can reproduce the output of the whole-brain scale activity represented on a behavioral scale, and in the context of the considered decision-making tasks. This supports the idea of that the AdEx mean-field framework can be embedded in a whole-brain simulator such as The Virtual Brain (TVB) [25, 46, 53] and can reproduce whole-brain dynamics induced by cognitive tasks, in particular by decision-making tasks.

The model was tested at the behavioral level by comparing its predictions to the experimental data obtained from the human and macaque participants. The comparison was made based on three statistical metrics: mean value of initial episode numbers of the performance clusters, mean value of in-cluster performance deviations and reaction time distributions. It was found that the model reproduces several characteristics of the experiment results. To begin with, the model predicts correctly the performance measures in the cases of both human and macaque for properly chosen parameter sets. It reproduces closely the reaction time distributions of the human data. The correlation values between the simulation and experiment reaction time histograms are lower in Trial 2 of Horizon 1 in both human and macaque compared to Trial 1 decision times. A possible reason for this is that the decision threshold is fixed. It is not dynamically adapted over the episodes as the strategy is learned.

Optimal decision-making refers to making choices which provide the maximum possible reward [17]. Albeit the absence of a precise definition, sub-optimal decision-making refers to the decision-making deviating from the optimal decision-making in terms of performance. Our model is designed for the optimal decision-making, and it has the potential to be extended to the sub-optimal decision-making. The model already reproduces partially sub-optimal behavior in the performance clusters thanks to its exploratory behavior as shown, for example, in Figures 6a–6c. This property can be improved: In the regulatory module, we can introduce a term which weights the flexibility parameter based on the loss or the augmentation of motivation of the participant to conduct the task. This can be combined with the decision thresholds which are dynamically adapted to the performance. This could project the changes in the motivation of the participant on the performance results. This modification could provide a better understanding of the neural dynamics which modulate the behavioral strategies of the participant in accordance with the increased or decreased motivation. This has been an active research area in the experimental side [23, 32, 40], and it can benefit enormously from the computational side.

Finally, a further extension of the model can be towards the large-scale brain dynamics of decision-making. For this, our bicolumnar AdEx mean-field framework can be integrated in TVB to model the decision-making brain dynamics. In this way, the perceptual inputs can be provided directly from the visual (or other perceptual) areas. The output of the bicolumnar framework can be provided to motor areas in a biologically realistic way. Nevertheless, it is not straightforward how to and for which regions such embedding should be done, evoking interesting questions regarding the information routing in the brain.

Supporting information

Simulation parameters

The parameters which are used in Figures 6-10 and Figures 14-17 are as follows:

$$\begin{gathered} T = 5\,\mathrm{ms}^{-1}, \quad \sigma = 0.01, \quad \tau_w = 5000\,\mathrm{ms}^{-1}\ (\text{for RS}), \quad 1^{-9}\,\mathrm{ms}^{-1}\ (\text{for FS}), \\ a = 4\ (\text{for RS}), \quad 0(\text{for FS}), \quad b = 40\ (\text{RS}),\ 0\ (\text{FS}), \quad E_L = -65\,\mathrm{mV}, \\ N_{e_A} = N_{e_B} = 8000, \quad N_{i_A} = N_{i_B} = 2000, \end{gathered} \tag{14}$$

and it is given for (3) as

$$v_{\mathrm{AI}} = 5\,\mathrm{Hz}, \quad w_c^e = w_c^i = 2.5 \times 10^{-4}, \quad p_c = 0.8. \tag{15}$$

Finally, the parameters appearing in (4) and (5) are

$$\tau_\psi = T = 5, \qquad \sigma = \sigma_r = 0.01, \qquad w_r = 1. \tag{16}$$

Each trial lasts 15 seconds in both Horizon 0 and 1 simulations. We set the decision threshold to 5 Hz in all the simulations.

Global measures

In Figure 11, we provide a visual highlight of the definitions of the global measures given in Section 3.3.1.

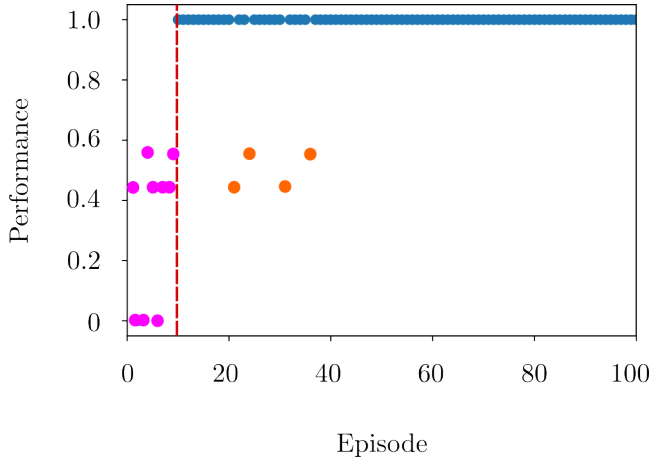

Fig 11. Highlighted global objects in Horizon 1 human experiment results. The pink samples correspond to a deviation block, and they are out-cluster deviations. The orange samples are in-cluster deviations, they do not form a deviation block. Both the pink and the orange samples are performance deviations. The blue samples are maximum performance samples and they form a performance cluster. The vertical red line marks the beginning of the performance cluster.

4.1 Effect of varying the bias

In this section, the results focus on the effects of varying the initial condition $\psi(0)$, therefore the effects of varying the bias. The threshold to measure the reaction time $\bar{t}$ is $|v_{e_A}(\bar{t}) - v_{e_B}(\bar{t})| = 7.5\ Hz$. We use Euler-Maruyama scheme with time step $\Delta t = 0.5$ and $t_f = 4$ for the results presented in Figure 12. Stimuli are applied at $t_0 = 2$. The parameters in this framework are as follows:

$$\begin{gathered} T = 0.003, \quad \sigma = 0.01, \quad \tau_w = 500\ (\text{for RS}),\ 10^{-9}\ (\text{for FS}), \\ a = 1\ (\text{RS}),\ 0(\text{FS}), \quad b = 100\ (\text{RS}),\ 0\ (\text{FS}), \quad E_L = -65, \\ N_{e_A} = N_{e_B} = 8000, \quad N_{i_A} = N_{i_B} = 2000. \end{gathered} \tag{17}$$

Moreover, in (3) we fix

$$v_{\mathrm{AI}} = 5\,\mathrm{Hz}, \quad w_c^e = w_c^i = 0.01, \quad p_c = 0.025. \tag{18}$$

Finally, the parameters appearing in (4) and (5) are

$$\tau_\psi = 10\,T = 0.03, \qquad \sigma = 0.01, \qquad c_0 = 1000, \qquad w_r = 1. \tag{19}$$

In Figure 12a, we show an example of the applied stimuli and the results of the case with $\psi(0) = 0$. In Figure 12b, we provide the results of the cases with $\psi(0)$ varied from 0 to 1, where the same stimuli shown in Figure 12a were applied. In the cases where ψ converges to 1, i.e., where Stimulus A is chosen, the reaction time is lower (≈ 0.1 s) than the cases where ψ converges to 0, i.e., where Stimulus B is chosen (≈ 0.18 s). This is due to that we chose $v_A(0) > v_B(0)$ in the trials given in Figure 12a, and therefore $v_{e_A}(\bar{t}) - v_{e_B}(\bar{t}) = 7.5$ is reached in a shorter time duration compared to $v_{e_A}(\bar{t}) - v_{e_B}(\bar{t}) = -7.5$.

In Figure 12b, there is no clear difference in terms of reaction time between the cases with different initial conditions $\psi(0)$ as long as the stimuli remain the same and ψ converges to the same value. It is due to the fact that, the convergence of the regulatory mechanism is rapid, therefore the competition is promoted towards the same pool and as we do not change the stimuli, the evolution of v_{e_A}, v_{e_B} becomes different realizations of almost the same random processes. Consequently, the reaction times of those realizations fluctuate around the same value.

Finally, we refer to Figure 13 for the effects of varying the stimuli difference on the reaction time.

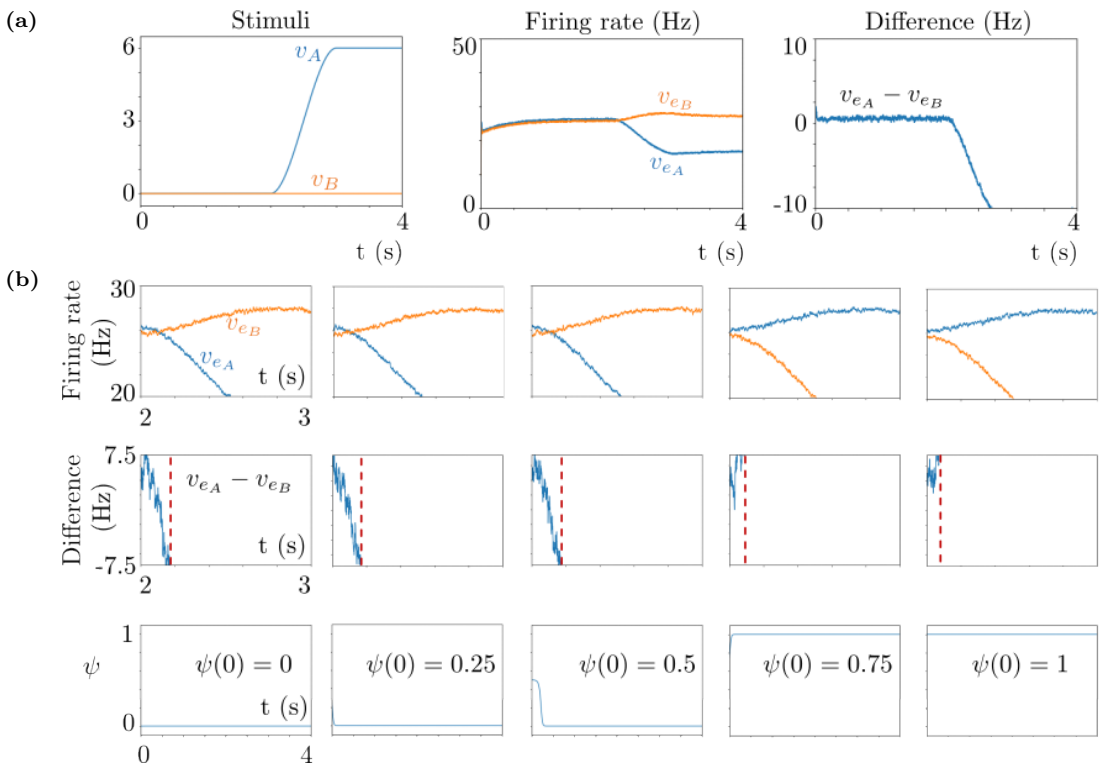

Fig 12. Simulation results of isolated trials, with different regulatory module initial conditions $\psi(0)$. **(a)** The result for $\psi(0) = 0$. Left: Applied stimuli v_A and v_B. Middle: Time course of the excitatory population membrane voltages of both Pools A and B. Right: Difference of the evolving membrane voltages in time. **(b)** The results for varied $\psi(0)$ values. Top: Time courses of v_{e_A} and v_{e_B}. Middle: Difference of the firing rates where the red vertical line denotes the decision instant. Bottom: Time evolution of the regulatory pool $\psi(t)$. The initial values $\psi(0)$ are 0, 0.25, 0.5, 0.75, 1 from left to right.

Effect of varying the difference between stimuli

In Figure 13, we observe the effects of varying the difference between the stimuli on the model reaction time. The reaction time is measured in the same way as in Figure 12b. We set the amplitude of Stimuli A to 6 and vary the amplitude of Stimuli B between 0 and 6. We keep constant ($= 0.75$) the initial value $\psi(0)$ of the regulatory function ψ such that the system privileges always Stimulus A. We observe that the reaction time increases as the difference between the stimuli decreases. It is expected since it becomes more difficult to make a distinction between the stimuli as the difference between the filled quantities of the bars decreases.

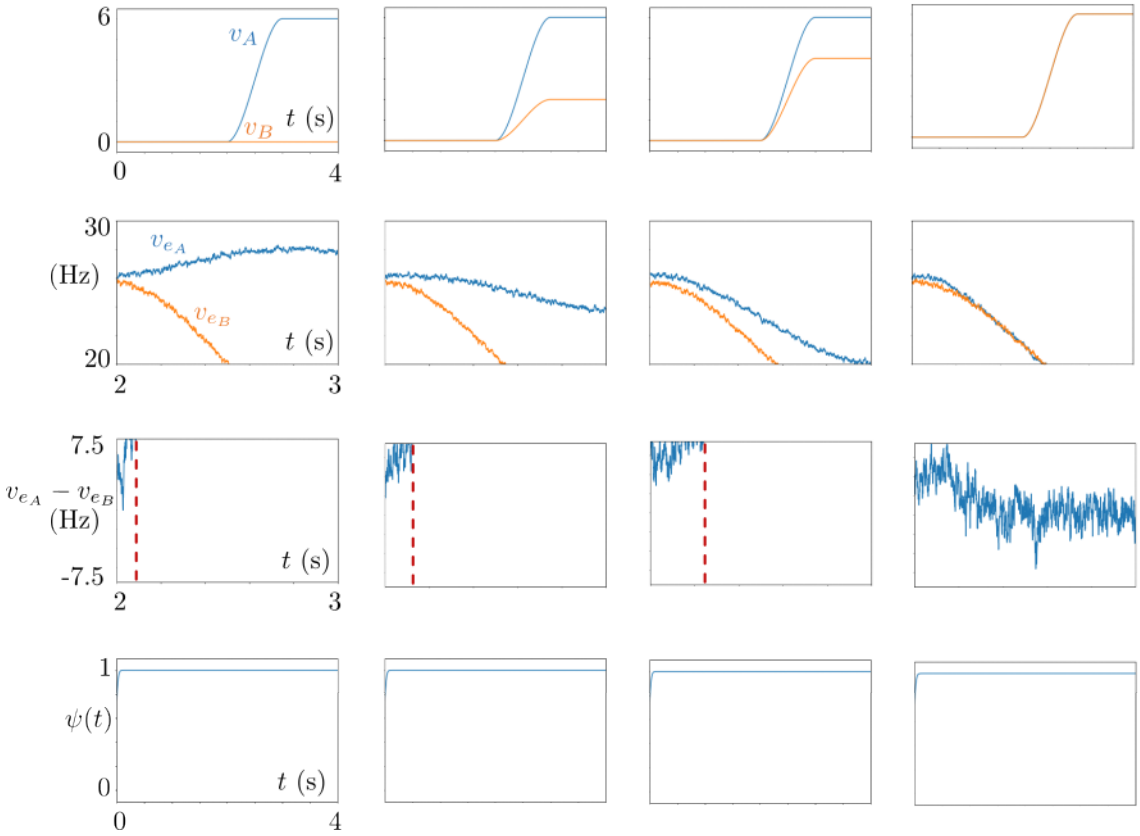

Fig 13. Simulation results with changing v_B. Top row: Time courses of the external stimuli v_A and v_B. Second row: Time courses of v_{e_A} and v_{e_B}. Third row: Time course of the difference between the membrane potentials where the red vertical line denotes the decision instant. Bottom row: Time course of the regulatory function ψ. The initial value $\psi(0)$ is 0.75 in all plots.

Comparison to the Horizon 0 human performance

In Figure 14, we provide the results of the Horizon 0 simulations similarly to the Horizon 1 results presented in Figure 8. The red horizontal lines in Figure 14 show the experimental values. We observe in the left column that the learning speed decreases both with increasing k and with increasing c_0, resulting in smaller values of the initial episode number of the performance clusters. The parameter k directly controls the learning speed, and c_0 contributes to it by avoiding the deviation blocks which might break the performance clusters. In both cases, the decrease in the mean value profiles is convex. We see that in both cases, the model can reproduce close statistics once k and c_0 are chosen properly, for example $k = 0.3$ and $c_0 = 1.1$.

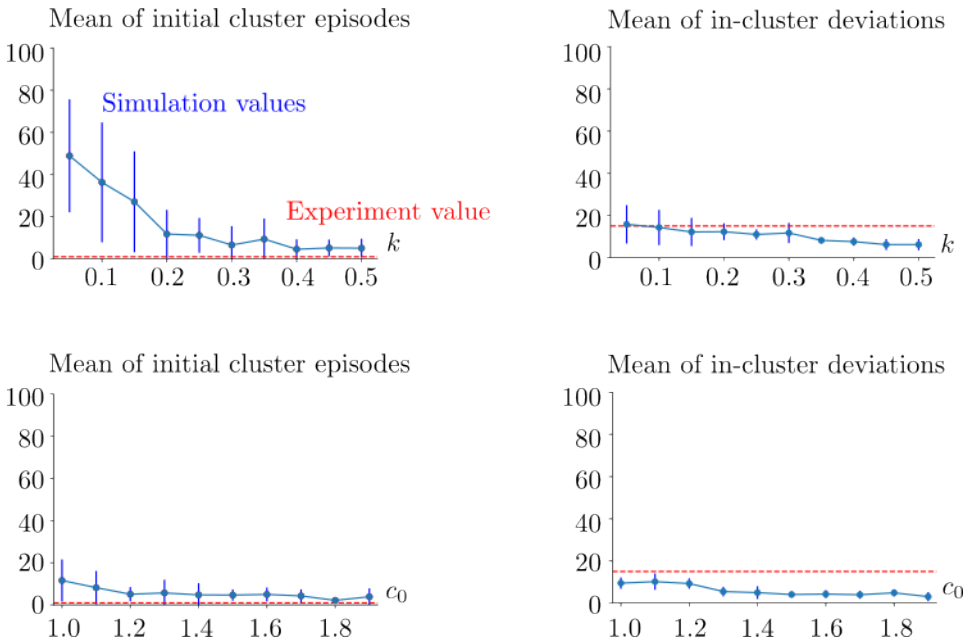

Fig 14. Horizon 0 human case. Simulation statistics with respect to the varied learning speed k in the top row, and with respect to the varied flexibility parameter c_0 in the bottom row. In the top row, $c_0 = 1.1$. In the bottom row, $k = 0.3$. The blue vertical lines show the standard deviations of the corresponding statistical sample. The red dashed lines are the experimental values. The mean and standard deviations are obtained via repeating the simulation 10 times for each (k, c_0) pair. See (14)-(16) for the rest of the parameters.

Comparison to the Horizon 0 human reaction times

In Figure 15, we provide the histograms of the reaction times obtained from Horizont 0 simulations with the varied learning speed k and flexibility parameter c_0. We observe that there is no distinguishing change in the histograms as we vary k and c_0, suggesting that these two parameters do not have noticeable effects on the reaction times.

In Figure 16, we show the simulation and experiment histograms of the reaction times corresponding to Horizon 0, with the fitted distributions for the best performance k and c_0 values as in the Horizon 1 case. The best fit is achieved with skewrow distribution and hypergeometric distribution for the experiment and simulation histograms, respectively. Squared sum errors are 13e-6 and 12e-6 for the experiment and simulation histograms, respectively. Pearson correlation coefficient between the fitted distributions is 0.91.

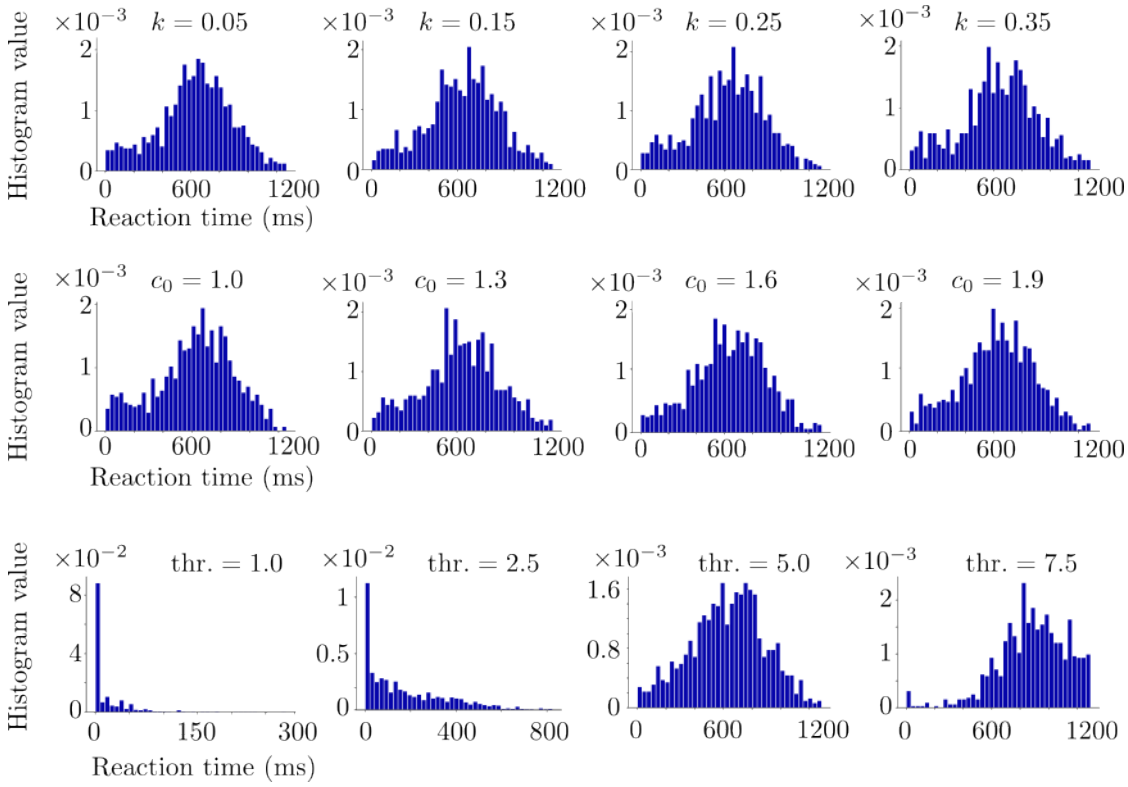

Fig 15. Horizon 0 reaction time histograms. Top: The histograms of the varied k, with $c_0 = 1.1$, decision threshold = 5 Hz. Middle: The histograms of the varied c_0, with $k = 0.3$, decision threshold = 5 Hz. Bottom: The histograms of the varied decision threshold, where $k = 0.3$, $c_0 = 1.1$. The histograms are obtained for each parameter from 10 realizations of the same simulation setup. The parameters are the same as those of Figure 14.

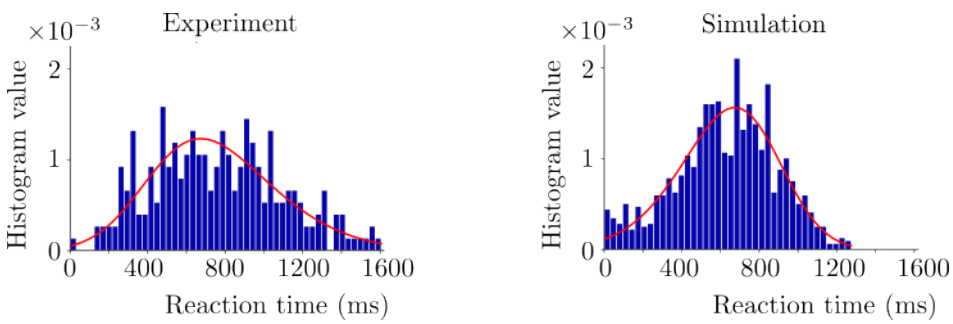

Fig 16. Horizon 0 human case. Histograms obtained from the human experiment and simulations for comparison. The red curves denote the fitted distributions via Python. The simulations are performed with $k = 0.3$, $c_0 = 1.1$ and decision threshold = 0.5.

Comparison to the macaque reaction times

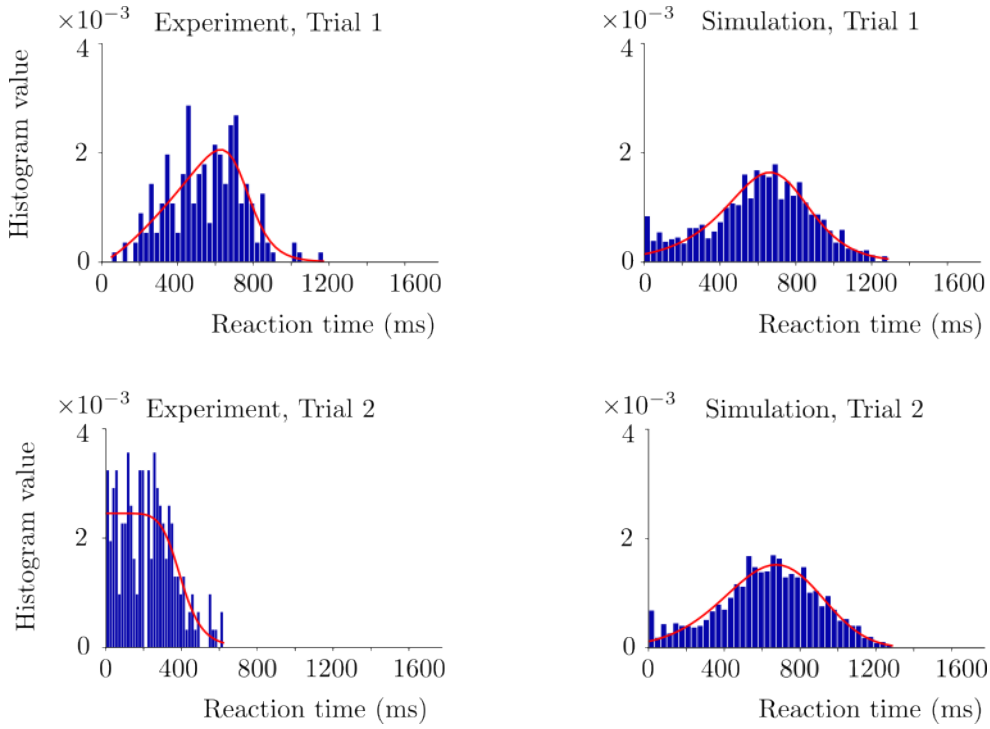

Fig 17. Horizon 1 macaque case. Histograms obtained from the experiment and simulations for comparison. The red curves denote the fitted distributions via Python. Trial 1 and 2 histograms are in the top and bottom rows, respectively. The simulations are performed with $k = 0.12$, $c_0 = 0.106$ and decision threshold $= 5$. The rest of the parameters is found in (14)-(16).

In Figure 17, we show the simulation and experiment histograms of the reaction times corresponding to the macaque experiment, together with the fitted distributions for the best performance k and c_0 values. We fit distributions to both simulation and experiment histograms by using the *Fitter* function from the Python *Fitter* package as in the case of the human experiment. In Trial 1, genlogistic distribution and powernorm distribution are fitted to the experiment and simulation histograms, respectively. Corresponding squared sum error values are 1e-5 and 9e-6 for the experiment and simulation histograms, respectively. Pearson correlation coefficient of these fitted distributions is 0.97. In Trial 2 histograms, burr distribution and kappa3 distribution are fitted to the experiment and simulation histograms, respectively. Corresponding squared sum error values are 9.9e-5 and 5.73e-4 for the experiment and simulation histograms, respectively. Pearson correlation coefficient of these fitted distributions is 0.31. Trial 2 has a weaker overlap between the experiment and simulation histograms compared to Trial 1 as in the case of the human task.

Acknowledgments

We hereby acknowledge that this research was supported by the Human Brain Project (European Union grant H2020-945539).

References

1. L. Albantakis and G. Deco, *Changes of mind in an attractor network of decision-making*, PLoS Comput Biol, 7 (2011), p. e1002086.

2. S.-i. Amari, *Dynamics of pattern formation in lateral-inhibition type neural fields*, Biological Cybernetics, 27 (1977), pp. 77–87.

3. L. Arnold, *Stochastic differential equations*, New York, (1974).

4. E. Baspinar, G. Cecchini, R. Moreno-Bote, I. Cos, and A. Destexhe, *Double columnar Adaptive Exponential mean-field model for decision-making* , *EBRAINS*, 2023.

5. ——, *Jupyter notebook of a biophysically plausible decision-making model based on interacting cortical columns*, *Zenodo*, 2023.

6. A. Bechara, D. Tranel, and H. Damasio, *Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions*, Brain, 123 (2000), pp. 2189–2202.

7. R. Brette and W. Gerstner, *Adaptive exponential integrate-and-fire model as an effective description of neuronal activity*, Journal of Neurophysiology, 94 (2005), pp. 3637–3642.

8. Y. Broche-Pérez, L. H. Jiménez, and E. Omar-Martínez, *Neural substrates of decision-making*, Neurología (English Edition), 31 (2016), pp. 319–325.

9. N. Brunel and X.-J. Wang, *Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition*, Journal of Computational Neuroscience, 11 (2001), pp. 63–85.

10. Z. Brzosko, W. Schultz, and O. Paulsen, *Retroactive modulation of spike timing-dependent plasticity by dopamine*, eLife, 4 (2015), p. e09685.

11. Z. Brzosko, S. Zannone, W. Schultz, C. Clopath, and O. Paulsen, *Sequential neuromodulation of hebbian plasticity offers mechanism for effective reward-based navigation*, eLife, 6 (2017), p. e27756.

12. G. Cecchini, M. DePass, E. Baspinar, M. Andujar, S. Ramawat, P. Pani, S. Ferraina, A. Destexhe, R. Moreno-Bote, and I. Cos, *Consequence assessment and behavioral patterns of inhibition in decision-making: modelling its underlying mechanisms*, bioRxiv, (2023).

13. A. K. Churchland, R. Kiani, and M. N. Shadlen, *Decision-making with multiple alternatives*, Nature Neuroscience, 11 (2008), pp. 693–702.

14. M. DePass, G. Cecchini, and I. Cos, *Consequence assessment and behavioral patterns of inhibition in decision-making (v1)*, EBRAINS dataset, 2023.

15. 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, 31 (2019), pp. 653–680.

16. P. Domenech and E. Koechlin, *Executive control and decision-making in the prefrontal cortex*, Current Opinion in Behavioral Sciences, 1 (2015), pp. 101–106.

17. K. Doya, S. Ishii, A. Pouget, and R. P. Rao, *Bayesian brain: Probabilistic approaches to neural coding*, MIT press, 2007.

18. J. Drugowitsch, R. Moreno-Bote, A. K. Churchland, M. N. Shadlen, and A. Pouget, *The cost of accumulating evidence in perceptual decision making*, Journal of Neuroscience, 32 (2012), pp. 3612–3628.

19. S. El Boustani and A. Destexhe, *A master equation formalism for macroscopic modeling of asynchronous irregular activity states*, Neural Computation, 21.

20. A. Finkelstein, L. Fontolan, M. N. Economo, N. Li, S. Romani, and K. Svoboda, *Attractor dynamics gate cortical information flow during decision-making*, Nature Neuroscience, (2021), pp. 1–8.

21. R. Fontana, M. Andujar, I. B. Marc, V. Giuffrida, S. Ramawat, G. Bardella, E. Brunamonti, P. Pani, and S. Ferraina, *Evaluating consequences in decision-making: behavioral data and neural recordings from monkey premotor and prefrontal cortex (v1)*, EBRAINS dataset, 2022.

22. S. Funahashi, *Prefrontal contribution to decision-making under free-choice conditions*, Frontiers in neuroscience, 11 (2017), p. 431.

23. M. Giamundo, F. Giarrocco, E. Brunamonti, F. Fabbrini, P. Pani, and S. Ferraina, *Neuronal activity in the premotor cortex of monkeys reflects both cue salience and motivation for action generation and inhibition*, Journal of Neuroscience, 41 (2021), pp. 7591–7606.

24. C. D. Gilbert and T. N. Wiesel, *Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex*, Nature, 280 (1979), pp. 120–125.

25. J. S. Goldman, L. Kusch, B. H. Yalcinkaya, D. Depannemaecker, T.-A. E. Nghiem, V. Jirsa, and A. Destexhe, *Brain-scale emergence of slow-wave synchrony and highly responsive asynchronous states based on biologically realistic population models simulated in the virtual brain*, bioRxiv, (2020).

26. Z. Guo, J. Chen, S. Liu, Y. Li, B. Sun, and Z. Gao, *Brain areas activated by uncertain reward-based decision-making in healthy volunteers*, Neural Regeneration Research, 8 (2013), p. 3344.

27. A. N. Hampton, P. Bossaerts, and J. P. O'doherty, *The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans*, Journal of Neuroscience, 26 (2006), pp. 8360–8367.

28. B. Y. Hayden and R. Moreno-Bote, *A neuronal theory of sequential economic choice*, Brain and Neuroscience Advances, 2 (2018), p. 2398212818766675.

29. R. Kiani, T. D. Hanks, and M. N. Shadlen, *Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment*, Journal of Neuroscience, 28 (2008), pp. 3017–3029.

30. Z. F. Kisvárday, K. A. Martin, T. Freund, Z. Magloczky, D. Whitteridge, and P. Somogyi, *Synaptic targets of hrp-filled layer iii pyramidal cells in the cat striate cortex*, Experimental Brain Research, 64 (1986), pp. 541–552.

31. C. D. Kopec, J. C. Erlich, B. W. Brunton, K. Deisseroth, and C. D. Brody, *Cortical and subcortical contributions to short-term memory for orienting movements*, Neuron, 88 (2015), pp. 367–377.

32. L. A. Leotti and T. D. Wager, *Motivational influences on response inhibition measures.*, Journal of Experimental Psychology: Human Perception and Performance, 36 (2010), p. 430.

33. E. Marcos, P. Pani, E. Brunamonti, G. Deco, S. Ferraina, and P. Verschure, *Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance*, Neuron, 78 (2013), pp. 249–255.

34. K. Martin and D. Whitteridge, *Form, function and intracortical projections of spiny neurones in the striate visual cortex of the cat.*, The Journal of Physiology, 353 (1984), pp. 463–504.

35. B. A. McGuire, C. D. Gilbert, P. K. Rivlin, and T. N. Wiesel, *Targets of horizontal connections in macaque primary visual cortex*, Journal of Comparative Neurology, 305 (1991), pp. 370–392.

36. J. Mirenowicz and W. Schultz, *Importance of unpredictability for reward responses in primate dopamine neurons*, Journal of neurophysiology, 72 (1994), pp. 1024–1027.

37. G. Mochol, R. Kiani, and R. Moreno-Bote, *Prefrontal cortex represents heuristics that shape choice bias and its integration into future behavior*, Current Biology, 31 (2021), pp. 1234–1244.

38. I. E. Monosov and O. Hikosaka, *Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex*, Journal of Neuroscience, 32 (2012), pp. 10318–10330.

39. R. Moreno-Bote, J. Rinzel, and N. Rubin, *Noise-induced alternations in an attractor network model of perceptual bistability*, Journal of Neurophysiology, 98 (2007), pp. 1125–1139.

40. S. Padmala and L. Pessoa, *Interactions between cognition and motivation during response inhibition*, Neuropsychologia, 48 (2010), pp. 558–565.

41. A. J. Parker and W. T. Newsome, *Sense and the single neuron: probing the physiology of perception*, Annual Review of Neuroscience, 21 (1998), pp. 227–277.

42. M. P. Paulus, C. Rogalsky, A. Simmons, J. S. Feinstein, and M. B. Stein, *Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism*, Neuroimage, 19 (2003), pp. 1439–1448.

43. R. Romo and E. Salinas, *Touch and go: decision-making mechanisms in somatosensation*, Annual Review of Neuroscience, 24 (2001), pp. 107–137.

44. M. F. Rushworth, M. P. Noonan, E. D. Boorman, M. E. Walton, and T. E. Behrens, *Frontal cortex and reward-guided learning and decision-making*, Neuron, 70 (2011), pp. 1054–1069.

45. A. G. Sanfey, G. Loewenstein, S. M. McClure, and J. D. Cohen, *Neuroeconomics: cross-currents in research on decision-making*, Trends in Cognitive Sciences, 10 (2006), pp. 108–116.

46. P. Sanz Leon, S. A. Knock, M. M. Woodman, L. Domide, J. Mersmann, A. R. McIntosh, and V. Jirsa, *The virtual brain: a simulator of primate brain network dynamics*, Frontiers in neuroinformatics, 7 (2013), p. 10.

47. J. D. Schall, *Neural basis of deciding, choosing and acting*, Nature Reviews Neuroscience, 2 (2001), pp. 33–42.

48. W. Schultz, *Dopamine reward prediction error coding*, Dialogues in clinical neuroscience, (2022).

49. E. Seidemann, E. Zohary, and W. T. Newsome, *Temporal gating of neural signals during performance of a visual discrimination task*, Nature, 394 (1998), pp. 72–75.

50. M. N. Shadlen and R. Kiani, *Decision making as a window on cognition*, Neuron, 80 (2013), pp. 791–806.

51. A. Sirigu and J.-R. Duhamel, *Reward and decision processes in the brains of humans and nonhuman primates*, Dialogues in Clinical Neuroscience, (2022).

52. S. Suzuki, V. M. Lawlor, J. A. Cooper, A. R. Arulpragasam, and M. T. Treadway, *Distinct regions of the striatum underlying effort, movement initiation and effort discounting*, Nature human behaviour, 5 (2021), pp. 378–388.

53. A. Turan, E. Baspinar, and A. Destexhe, *A whole-brain model of auditory discrimination*, bioRxiv, (2023), pp. 2023–09.

54. F. Van Ede, S. R. Chekroud, M. G. Stokes, and A. C. Nobre, *Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors*, Nature Communications, 9 (2018), pp. 1–12.

55. X.-J. Wang, *Probabilistic decision making by slow reverberation in cortical circuits*, Neuron, 36 (2002), pp. 955–968.

56. H. R. Wilson and J. D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophysical Journal, 12 (1972), pp. 1–24.

57. Z.-C. Xiao, K. K. Lin, and L.-S. Young, *A data-informed mean-field approach to mapping of cortical parameter landscapes*, PLoS Computational Biology, 17 (2021), p. e1009718.

58. Y. Zerlaut, S. Chemla, F. Chavane, and A. Destexhe, *Modeling mesoscopic cortical dynamics using a mean-field model of conductance-based networks of adaptive exponential integrate-and-fire neurons*, Journal of Computational Neuroscience, 44 (2018), pp. 45–61.

59. Y. Zuo and M. E. Diamond, *Rats generate vibrissal sensory evidence until boundary crossing triggers a decision*, Current Biology, 29 (2019), pp. 1415–1424.