scientific reports

Check for updates

OPEN

A generalized reinforcement learning based deep neural network agent model for diverse cognitive constructs

Sandeep Sathyanandan Nair[1], Vignayanandam Ravindernath Muddapu[1,4], C. Vigneswaran[1], Pragathi P. Balasubramani[2,5], Dhakshin S. Ramanathan[2,3], Jyoti Mishra[2] & V. Srinivasa Chakravarthy[1]✉

Human cognition is characterized by a wide range of capabilities including goal-oriented selective attention, distractor suppression, decision making, response inhibition, and working memory. Much research has focused on studying these individual components of cognition in isolation, whereas in several translational applications for cognitive impairment, multiple cognitive functions are altered in a given individual. Hence it is important to study multiple cognitive abilities in the same subject or, in computational terms, model them using a single model. To this end, we propose a unified, reinforcement learning-based agent model comprising of systems for representation, memory, value computation and exploration. We successfully modeled the aforementioned cognitive tasks and show how individual performance can be mapped to model meta-parameters. This model has the potential to serve as a proxy for cognitively impaired conditions, and can be used as a clinical testbench on which therapeutic interventions can be simulated first before delivering to human subjects.

High-level human cognition consists of a variety of functions or capabilities, including selective processing of goal-relevant information, suppression of goal-irrelevant information, action selection, reward processing, working memory, etc. There is a long history of empirical research that studies the various cognitive functions individually while excluding other functions. However, to understand these cognitive functions as functions of an integrative agent, it is essential to study them holistically, revealing the synergies among these functions that come into play as an agent interacts with its environment.

While empirical research on cognitive functions suffers from this fragmented approach due to several challenges including participant burden, limited resources and expertise, theoretical investigation also often reflects this piecemeal approach, offering a wide variety of models that describe individual cognitive functions. There have been efforts to construct integrative computational frameworks that capture a range of cognitive functions. For example, the "ACT-R" system has been proposed as a general framework for modeling a wide variety of cognitive processes[1]. Subsequently, it was extended to include visual attention, and its properties, like speed and selectivity, as they vary from subject to subject[1]. Similarly, the "Soar" architecture can successfully integrate different levels of reasoning, planning, reactive execution, and learning from experience[2]. The importance of more holistic and integrative models has been emphasized by several researchers, resulting in many unified computational models of cognition. As these models evolved, a certain similarity among these modeling architectures began to reveal itself. For example, common features of three such cognitive architectures viz., ACT-R[1], Soar[2,3], and SIGMA[4] have been described[5]. More on fragmented and integrative approach is updated in the supplementary material.

Neuropsychiatric disorders are characterized by a wide range of cognitive dysfunctions, and the degree to which these disorders are mapped to specific neural substrates is still being resolved[6]. Just as theoreticians

[1]Computational Neuroscience Lab, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Room 505, Block 1, Sardar Patel Road, Adyar, Chennai, Tamil Nadu 600036, India. [2]Neural Engineering and Translation Labs, Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [3]Department of Mental Health, VA San Diego Medical Center, San Diego, CA, USA. [4]Present address: Blue Brain Project, École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech, 1202 Geneva, Switzerland. [5]Present address: Department of Cognitive Science, Indian Institute of Technology, Kanpur, Kanpur, India. ✉email: schakra@ee.iitm.ac.in

sought to create unified architectures of cognition, experimental cognitive scientists also made efforts to move away from exclusivist approaches and began to study multiple cognitive functions in human population cohorts simultaneously[7]. One such experimental system is the *BrainE* platform, which includes a range of cognitive assessments like selective attention (SA), response inhibition (RI), working memory (WM), and distractor processing (DP) in both non-emotional and emotional context[8]. The *BrainE* system measures both behavioral parameters and electroencephalography (EEG) signals, thereby creating an opportunity to relate cognitive behavior to neural substrates.

In this paper, we present a unified architecture of cognition that can model a range of cognitive functions, including SA, RI, WM, DP, etc. The proposed model has the following components: 1) sensory representation, 2) memory, 3) value computation, 4) exploration, and 5) action selection. The model is cast broadly within the framework of reinforcement learning (RL)[9–11]. Notably, the action selection strategy, which involves pursuit of explorative and exploitative modes, each of which is regulated based on the underlying value dynamics is novel to our approach. The model has elements common to deep neural networks and two novel neural elements that are not typically found in such networks viz., 1) flip-flop neurons and 2) oscillator neurons. First proposed in[12], the flip-flop neurons are fashioned after flip-flops in digital systems theory and can store memories. In the oscillator network, the lateral interactions are designed such that the oscillators exhibit desynchronized dynamics. Such oscillator networks have been used before to implement exploratory functions essential for achieving randomness in action selection in RL models[13]. We hypothesized that this modeling framework can replicate the subject's performance with respect to diverse cognitive decision-making tasks.

In the following section, we will describe the methods starting with a brief overview of the experimental setup, the cognitive testing paradigms used, and the various tasks conducted to evaluate cognitive abilities. This is followed by a brief overview of the model architecture that is used to simulate the experimental tests, various building blocks of the model architecture and their mathematical formulation, and how they are integrated to mimic an experimental subject. In the "Results" section, we summarize the main results, including the training phase, the performance evaluation and the meta-parameters used for tuning the performance. We compare the model performance with experimental results. The final section concludes the results and discusses the utility, limitations, and future scope.

Materials and methods

The proposed Generalized Reinforcement Learning-based Deep Neural Network (GRLDNN) agent model, as shown in the Fig. 1, can simulate various experimental paradigms that can test different cognitive functions such as SA, RI, WM, and DP.

The experimental tasks. A repertoire of tasks was prescribed to capture a range of cognitive functions that comprise human decision-making. The cognitive functions that we focus on in these tasks are the ability to selectively attend to relevant stimuli, inhibit responses to irrelevant stimuli, avoid distractors during an attention task, and working memory. We primarily focused on modeling the cognitive assessments like SA, RI, WM, DP conducted using *BrainE* platform[8]. In addition to this we have also modeled additional experimental paradigms such as the N-back task to assess working memory load and the 2×5 task, which evaluates the sequence processing capabilities. To test the effectiveness of the Q-learning network we have also modeled the T-maze and Grid-world tasks. The details of the tasks are described in the supplementary material.

GRLDNN (Generalized reinforcement learning-based deep neural network) agent model. We present a unified RL-based deep network architecture that can simulate all the experimental tasks described in Supplementary material (Section S1).

A schematic of the model architecture is given in Fig. 1. The model has five distinctly identifiable components viz.—1) Representational System (RS) consists of a series of layers—convolutional layers followed by fully-connected layers, that generate compact representations of the input images. 2) Memory System (MS) is a layer of flip-flop neurons that receives the inputs from the RS via a fully connected weight stage. This system has the memory property. 3) Value Computation System (VC) combines the neural outputs of the MS and computes the value function. 4) Explorer comprises a nonlinear oscillator network, wherein the oscillators interact via inhibitory connections, generating desynchronized oscillations. The randomness inherent in the chaotic oscillatory dynamics of this system introduces a level of randomness in the action selection at the output layer, driving exploratory behavior. 5) Action Selection System (AS) is the output of the entire architecture that combines the outputs of the MS via a trainable weight stage and the output of the Explorer. The AS is trained using Q-learning described in detail in the supplementary materials[14].

Representational system. The input stimulus is presented as an image to the input layer of the RS, which is trained as a convolutional autoencoder[15,16]. The RS module consists of an input layer followed by four convolutional and max-pooling layers. A fully connected layer follows the four layers of the convolutional and max-pooling layers. Another fully connected later is used to reduce the encoder output to 1×64. The convolutional layers use a 3×3 filter window size. Mean squared error is used as the output loss. At the decoder end, the 1×64 feature output is expanded, followed by deconvolutional layers and pooling layers, at the end of which the original image is reconstructed back. Output from the fully connected layer of the encoder part of RS is provided as the input to the MS.

Memory system. The output of the RS module is presented to the MS via a fully connected weight stage. The MS, as mentioned before, is a 1D layer of flip-flop neurons. This layer is divided into two equal sections—MS1

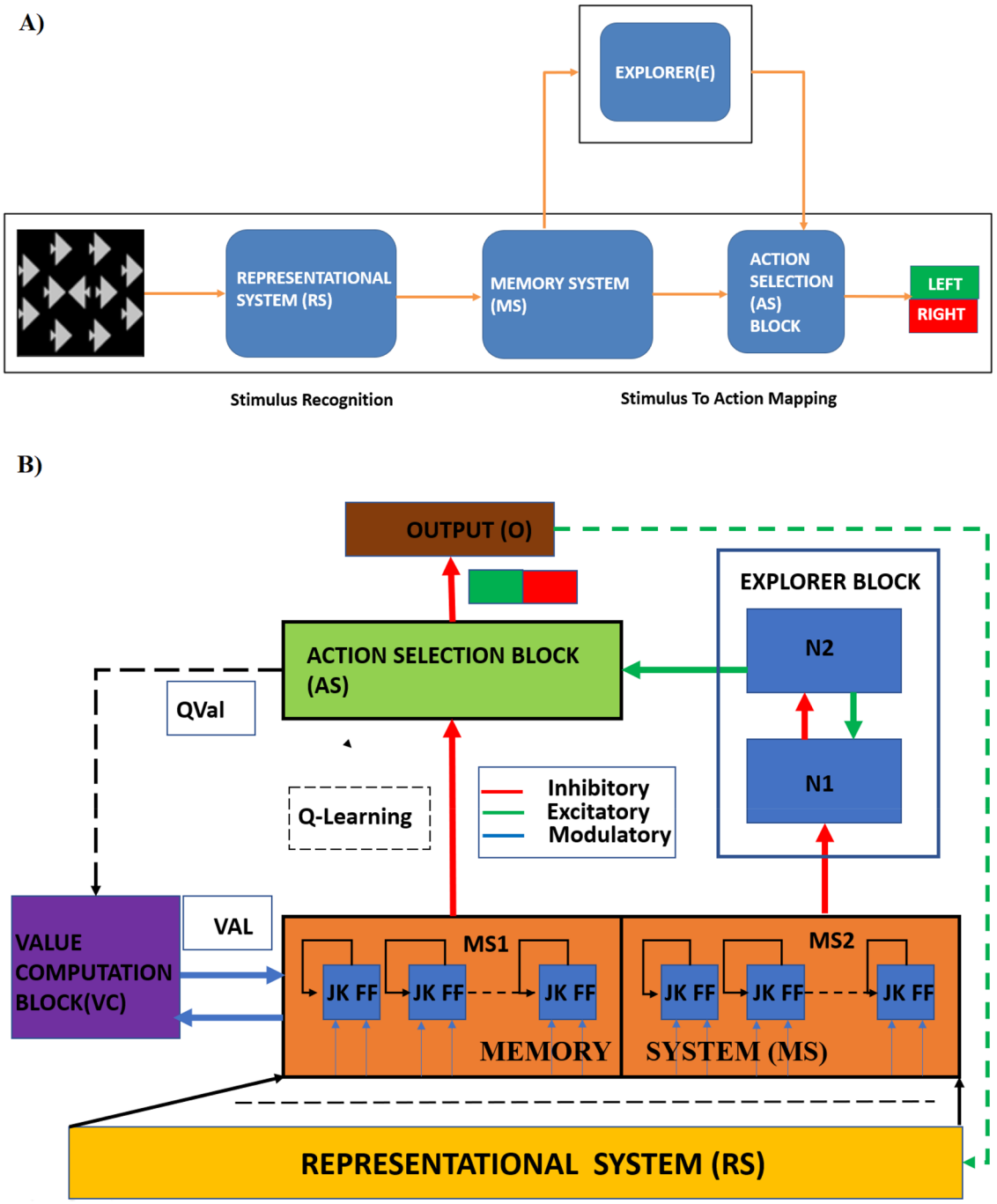

Figure 1. (**A**) Overview of (Generalized Reinforcement Learning-based Deep Neural Network) GRLDNN model architecture. RS, Representational System is used for stimulus recognition; Memory System (MS) and Action Selection (AS) block along with the Explorer (E) is used for stimulus to action mapping. The encoder output from RS is presented to the MS where the stimulus is processed, and the action selection takes place at AS. (**B**) Block diagram of the Agent Model architecture. RS, Representation System Block; Explorer Block consisting of N1/N2 Oscillator pair; AS, Action Selection Block; O, Output block; VAL, Value Computation Block where the Value function is computed; QVal, Q-value function; MS, Memory System block.

and MS2. MS1 has D1-type flip-flop neurons, whereas MS2 has D2-type flip-flop neurons. The flip-flop architecture is provided in the supplementary material.

The feature vector from the RS module (X_{RS}) reaches the flip-flop neurons of MS1 and MS2 over the weight stages ($W_i^{RS\to MS1}$) and $W_i^{RS\to MS2}$, respectively. So, the effective input received at MS1 is $W_i^{RS\to MS1}(t) * X_{RS}$ (Supplementary material).

The output of the encoder part of the RS module is presented as the input, $X_{RS}(t)$, to the J and K ports of the flip-flop neurons present in the MS1/MS2 sub-blocks of the MS, $W_i^{RS\to JMS1}$, $W_i^{RS\to KMS1}$, $W_i^{RS\to JMS2}$ and

$W_i^{RS\rightarrow KMS2}$ denote the weights from the RS layer to the respective J and K inputs of the MS1/MS2 sub-blocks (Supplementary material).

Computations in the MS1/MS2 sub-blocks of the memory system. The J and K inputs for the flip-flop neurons of MS1 and MS2 are given by Eqs. (1,2,3,4) below. The output of the flip-flop neuron is expressed by Eqs. (5, 6), which is in line with the circuit diagram and the truth table given in the supplementary material. The output is also influenced by the modulatory input received from the VC[17].

$$J_{MS1}(t) = W_i^{RS\rightarrow JMS1}(t)X_{RS}(t) \tag{1}$$

$$K_{MS1}(t) = W_i^{RS\rightarrow KMS1}X_{RS}(t) \tag{2}$$

$$J_{MS2(t)} = W_i^{RS\rightarrow JMS2}(t)X_{RS}(t) \tag{3}$$

$$K_{MS2}(t) = W_i^{RS\rightarrow KMS2}X_{RS}(t) \tag{4}$$

$$V_i^{MS1}(t) = J_{MS1}\left(1 - V_i^{MS1}(t-1)\lambda_{MS1}\right) + (1 - K_{MS1})V_i^{MS1}(t-1)\lambda_{MS1} \tag{5}$$

$$V_i^{MS2}(t) = J_{MS2}\left(1 - V_i^{MS2}(t-1)\lambda_{MS2}\right) + (1 - K_{D2})V_i^{D2}(t-1)\lambda_{D2} \tag{6}$$

$$V_i^{MS1}(t) = \lambda^{MS1}(t)W_i^{RS\rightarrow MS1}(t)X_{RS}(t) \tag{7}$$

$$V_i^{MS2}(t) = \lambda^{MS2}(t)W_i^{RS\rightarrow MS2}(t)X_{RS}(t) \tag{8}$$

$$\lambda^{MS1}(t) = \left(\frac{1}{1 + e^{K_1*(VAL(t)-\theta_{MS1})}}\right) \tag{9}$$

$$\lambda^{MS2}(t) = \left(\frac{1}{1 + e^{K_2*(VAL(t)-\theta_{MS2})}}\right) \tag{10}$$

where $K_1 < 0$ and $K_2 > 0$ and λ^{MS1} *and* λ^{MS2} are the sigmoid gain parameters.

Value computation system. The value is computed using a weighted sum of the outputs of the flip-flop neurons of MS1. Thus, the value function 'VAL', is computed as per Eq. (11) below,

$$VAL(t) = \sum_{i=1}^{n} W_i^{MS1\rightarrow VC}(t)V_i^{MS1} \tag{11}$$

Explorer module. The explorer module consists of a network of nonlinear oscillators. These are thought to be implemented by two pools of neurons, N1 and N2, connected back-to-back. The MS2-type flip-flop neurons of MS project to N1, whereas the output of the N1 neural layer, in turn, influences the N2 neural layer. N1 and N2 form a loop, with inhibitory projections from N1 to N2 and excitatory projections from N2 to N1. Such excitatory-inhibitory pairs of neurons pools have been shown to exhibit oscillations[10,18]. In the present case, it turns out that the equations that couple a single N1 neuron bidirectionally to a single N2 neuron can be classified as a general oscillator system called Lienard system, which exhibits limit cycle oscillations[19]. The dynamics of N1-N2 neuronal pools is defined as,

$$\tau_{N1}\frac{dV_i^{N1}}{dt} = -V_i^{N1} + \sum_{j=1}^{n} W_{ij}^{N1\rightarrow N1}V_j^{N1} + W_i^{N2\rightarrow N1}V_i^{N2} - V_i^{MS2}(t) \tag{12}$$

$$\tau_{N2}\frac{dU_i^{N2}}{dt} = -U_i^{N2} + \sum_{j=1}^{n} W_{ij}^{N2\rightarrow N2}V_j^{N2}(t) - W_i^{N1\rightarrow N2}V_i^{N1} \tag{13}$$

$$V_i^{N2}(t) = \tanh\left(\lambda^{MS2}(t)U_i^{N2}\right) \tag{14}$$

$$V_j^{A}(t) = \sum_{i=1}^{n} W_{i,j}^{MS1\rightarrow AS}(t)V_i^{MS1} \tag{15}$$

$$V_j^B(t) = \sum_{i=1}^{n} W_{i,j}^{N2\to AS}(t) V_i^{N2}(t) \tag{16}$$

where V_j^{N1} and U_j^{N2} are the internal states of N1 and N2 neurons, respectively, V_i^{N2} is the output of the N2 neuron, $W^{N1\to N1}$ and $W^{N2\to N2}$ are weight kernels representing lateral connectivity in N1 and N2 modules, respectively, τ_{N1} and τ_{N2} are the time constants of N1 and N2, respectively, $W^{N1\to N2}$ is the connection strength from N1 to N2, $W^{N2\to N1}$ is the connection strength from N2 to N1, and λ_{N2} is the parameter which controls the slope of the sigmoid in N2. $V_j^A(t)$ *and* $V_j^B(t)$ are the inputs arriving at AS1 block from MS1 and MS2 (via N1 &N2) respectively.

Lateral connections among coupled oscillators of the explorer module. The lateral connectivity in the N1 or N2 network is modeled using constant weights where weights between two neurons is given by $W^{N1\to N1}$ and $W^{N2\to N2}$,

$$\begin{aligned} &- W_{ij}^{N1\to N1} = W_{ij}^{N2\to N2} = 1 \; for \; i = j; \\ &- W_{ij}^{N1\to N1} = W_{ij}^{N2\to N2} = \epsilon, \; otherwise; \end{aligned} \tag{17}$$

where ϵ is the magnitude of the connection strength of lateral connections among the neighbouring neurons in both N1 and N2 modules.

Action selection module: Q-learning. The network's final output is a linear sum of the output of the MS1 block and the output of the explorer module.

$$V_i^A(t) = \sum_{i=1}^{n} W_{i,j}^{MS1\to AS}(t) V_i^{MS1} \tag{18}$$

The output V_i^A is combined with the output V_i^B (Eq. 16) at the AS block as shown in (Eq. 19) below.

$$\tau_{AS} \frac{dV_i^{AS}}{dt} = -V_i^{AS} - V_i^A(t) + V_i^B(t) \tag{19}$$

$$\tau_{AS} \frac{dV_i^{AS}}{dt} = -V_i^{AS} - \lambda^{MS1}(t) V_i^{MS1}(t) + \lambda^{MS2}(t) W_i^{N2\to AS} V_i^{N2}(t) \tag{20}$$

The action selection mechanism at AS is facilitated using a race model[20–23]. The output of the AS block ($-V_i^{AS}$) is compared against a threshold *Vth*. The neuron (ith) whose output crosses the threshold first, is considered a winner, and the *ith* action is selected.

$$If, V_i^{AS}(t) > V_{th}; \; then \; AS = i \tag{21}$$

Reward and learning. The weights between the neurons in the MS1 block and the AS module are updated using Q-Learning[24] as shown in (Eqs. 22,23,24).

$$Q_t(s,a) = V_j^{N1}(t) \tag{22}$$

$$\delta(t) = r(t) + \gamma \max_a Q_{t+1}(s,a) - Q_t(s,a) \tag{23}$$

where $\gamma = 0$ (discount factor).

$$Q_{t+1}(s,a) = Q_t(s,a) + \eta\Big(r(t) + \gamma \max_a Q_{t+1}(s,a) - Q_t(s,a)\Big) \tag{24}$$

Backward propagation. In this model, trainable weights are located in three areas: i) MS1-block to AS weight stage, ii) MS1 block to VC block weight stage, and iii) RS to MS1/MS2 block weight stage (Supplementary material). The weights between the various modules are updated using the backpropagation algorithm. The weight update between the MS1 and the AS blocks is governed by Q-learning, while the weight update between MS1 to VC is done using temporal difference (TD)-learning. The MS1/MS2 subblocks used flip-flop neurons, and the weight update between RS and MS is done using TD-learning. The weight training equations are described in detail in the supplementary material.

More details about the Markov decision model and the state and action spaces are given in the supplementary material.

Performance assessment. The model performance is assessed in terms of the metrics of accuracy, reaction time, speed, consistency, and efficiency, as defined in Supplementary material. We have also.

Results

In this section, we describe the performance of the GRLDNN model that is used to simulate the four cognitive functions of SA, RI, WM, and DP experimentally assessed using the *BrainE* platform. In the subsequent subsections, we will first show the progress of state value functions and the Q-value functions during the learning process. Then we will show the impact of parameter tuning with respect to lateral connections strengths and threshold (ϵ and V_{TH}). Then we implement an inverse model using a neural network that can calibrate the meta parameters of the GRLDNN model so that the model can simulate the experimental performance. We then present a comparison of the average performance between both the model and the experimental results. In addition to the results of the cognitive assessments conducted using *BrainE* platform[8], we also show the performance results of the N-back task, which tests the working memory load and present a comparison between the model and the experimental results. The results of the other tasks such as the 2×5(sequence processing) and T-maze are updated in the supplementary material.

Training phase. As the learning progresses, the magnitude of the value functions at the end of each trial keeps increasing and approach the maximum value of 1, as the model learns the task accurately. The Q-values for the respective state and action pairs corresponding to the correct action are higher, whereas the values corresponding to the wrong action are lower. Figure 2A represents the state value function for the *Go Green* (SA) task, and Fig. 2B represents the Q-value at the end of training epochs, at the last instance of the *Go Green* trials. The state value and Q-value functions for other tasks are described in the supplementary material.

The test dataset contains 33% of the green-colored rockets and 67% of other colored rockets in the SA task. The blue bar represents the 'Go' action and the yellow bar represents the 'No Go' action.

Effect of parameter tuning. The performance of cognitive tasks depends on various factors. We observed variations in the model performance by tuning specific meta parameters, using performance metrics of accuracy, reaction time, speed, consistency, and efficiency. The meta-parameters that are varied are ϵ and V_{TH}.

- ϵ influences the *lateral connectivity strength* of the N1 and N2 coupled oscillator system, which controls the level of exploration in the model.
- V_{TH} is the *threshold* that appears in the race model, used in the AS module where the action selection occurs, which controls how fast the decision can be made (reaction time).

Effect of the threshold ($\boldsymbol{V_{TH}}$) and lateral connectivity strength (ϵ) on the performance. Figure 3 shows the variation in performance with respect to speed and consistency for various values of lateral connection strength (ϵ) and threshold (V_{TH}) illustrated for Go Green (SA) task. V_{TH} is varied in the range of 0.3 to 0.5 in steps of 0.1 and ϵ takes values of 0.01, 0.03, 0.05 and 0.1. Figure 3A1 and Fig. 3A2 show the variation in decision-making *speed* with respect to changes in ϵ and V_{TH}. Figure 3A3 and Fig. 3A4 shows a similar variation in *consistency*. From the results, it can be seen that speed is inversely proportional to both the threshold and lateral connection strength.

Our results show that for lower values of the action selection threshold, the model performance exhibit higher speeds and vice versa. This is expected since the lower the threshold, the faster the action selected. The AS block being implemented as a race model decides on the action to be selected based on the threshold.

Hence by searching for an optimum in the meta-parameter space consisting of lateral connection strength and action selection threshold (ϵ, V_{TH}), it is possible to match the model's performance with experimental data. We modeled the mapping between the experimental parameters (speed, consistency) and the model meta-parameters (ϵ, V_{TH}), tuned using a simple multilayer perceptron model (MLP). Given the input values of speed and consistency, we can predict the corresponding values of ϵ and V_{TH}. Hence by varying the values of the two meta-parameters, we were able to calibrate the GRLDNN model so that the model can approximately simulate the experimental performance. The model fit between the predicted and desired values of ϵ and V_{TH} is as shown in Fig. 4 below for the *Go Green* (SA) task. Twelve data points were used for training from the combination of

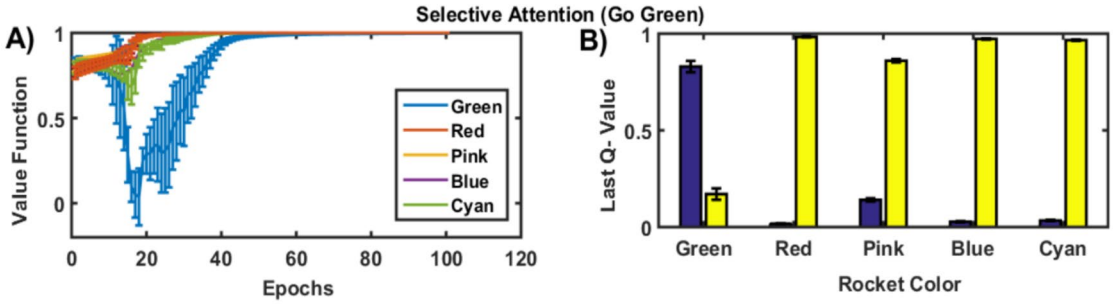

Figure 2. (**A**) The state value functions over the epochs during the training phase of the Go Green (SA) task. (**B**) The Q-value at the end of training for the Go Green task that requires selective attention to the green colored rockets while ignoring other isoluminant colors—red, pink, blue, cyan. The blue bar represents the 'Go' action and the yellow bar represents the 'No-Go' action.

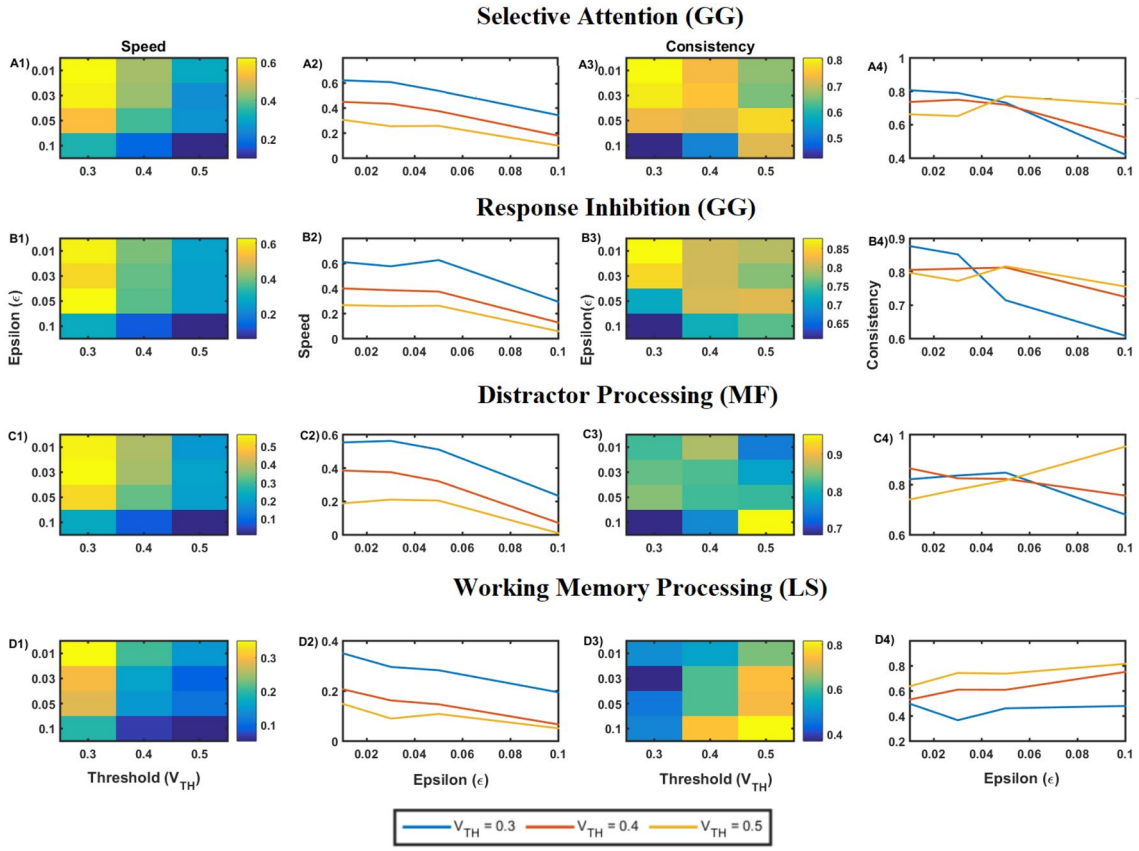

Figure 3. Plot of performance parameters for various tasks, with threshold (varying between 0.3 to 0.5 and lateral connection strength ϵ ranging between 0.01 to 0.1. (**A1**,**A2**) The speed with which the decision is made for Go-Green (SA) task, (**A3**,**A4**) Consistency for Go-Green (SA) task, which indicates how consistent the performance was with respect to speed across trials. It also indirectly indicates the standard deviation of the speed of performance. Similarly, (**B1**–**B4**) represents the Go-Green (Response Inhibition) task, (**C1**–**C4**) indicates the Middle Fish (Distractor Processing) task, (**D1**–**D4**) indicate performance on the Lost Star task.

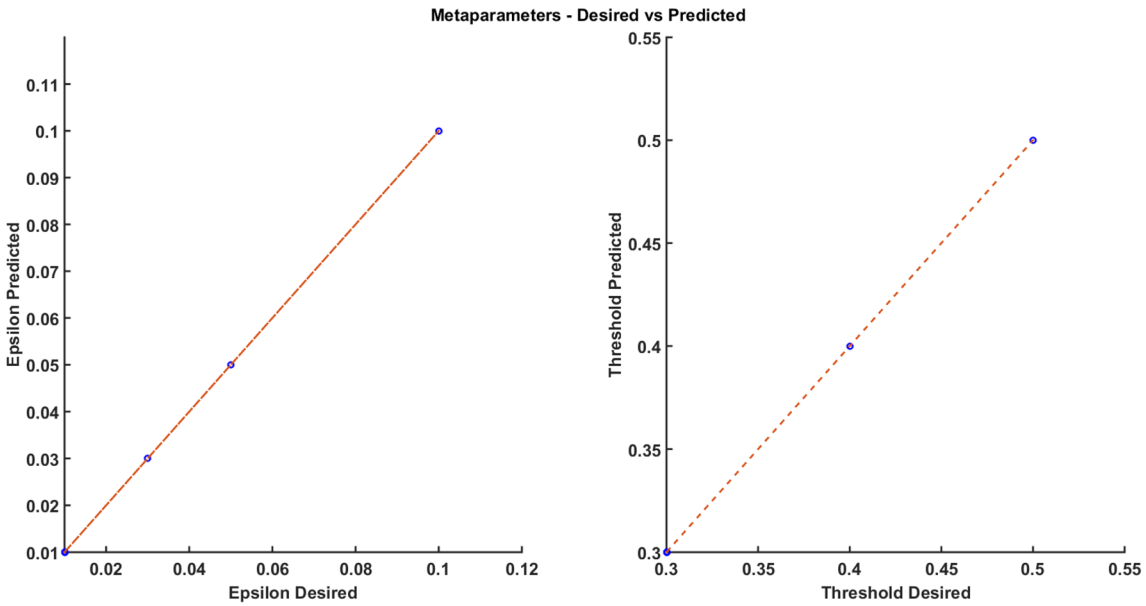

Figure 4. (**A**) The parameter fit is checked for the predicted vs. desired meta parameter (ϵ). (**B**) The parameter fit is checked for the predicted vs. desired meta parameter (V_{TH}).

meta-parameters (ϵ and V_{TH}), and the training error was in the order of ~ 10^{-5} after 50,000 epochs. The predicted and desired values of both the threshold and epsilon were matched.

The performance results of the GRLDNN agent model were compared with the experimental results of healthy subjects[8]. In the GRLDNN agent model, decision-making scenarios were simulated for the different cognitive functions of SA, RI, WM, DP by modeling the test paradigms for *Go Green*, *Middle Fish*, and *Lost Star* tasks. The model's performance was tuned using the meta-parameters ($V_{TH(pred)}$ *and* ϵ) to match the experimental results. By navigating through the parameter space as shown in Fig. 3, we selected the values of $V_{TH} = 0.4$ *and* $\epsilon = 0.05$, which was found to be closely matching with the average performance results of the experimental subjects. The RMSE (root mean squared error) was found to be lowest at 0.007 for $V_{TH} = 0.4$ *and* $\epsilon = 0.05$ when compared with the average performance of experimental results.

Figure 5 shows the performance comparison of the model and experimental data, where the blue bar represents the model performance, and the yellow bar represents the experimental performance results for all modeled tasks. As seen in the case of the *Go Green* task, the GRLDNN agent model recorded an average speed of 0.3510 ± 0.065 and 0.3756 ± 0.0279 for SA and RI, respectively (Fig. 5A, dark blue bar) compared to the experimental results, which recorded an average speed of 0.3580 ± 0.0534 and 0.3976 ± 0.0612, for SA and RI, respectively (Fig. 5A, yellow bar). Similar comparisons can be made for performance metrics for all modeled tasks.

Hence by mapping the performance characteristics across different cognitive tasks onto the meta-parameter space and navigating through the same, we are able to replicate empirical performance on different cognitive abilities using the GRLDNN model.

The performance characteristics of the N-back task is as shown in Fig. 6. The model performance results of the N-back task are comparable and relatable to the experimental results[25]. The response time and the accuracy were evaluated for both target and non-target stimuli. We have simulated up to N = 4. The blue bar indicates the model performance and the orange bar indicates the experimental results.

The performance characteristics of the sequence processing (2×5) and other tasks are updated in the supplementary material. We have considered two additional RL tasks that involves action based state transitions to show the effectiveness of our model. We have also discussed the stability aspects and convergence of the Q networks in the supplementary material.

Discussion

Using the proposed GRLDNN model, we successfully modeled tasks to assess a variety of cognitive abilities, including SA, RI, WM, and DP. By varying the meta-parameters, V_{TH} *and* ϵ, we were able to tune the performance outputs of the model (Fig. 3). Notably, our model results were comparable to that of experimental results for healthy subjects, as shown in Fig. 5.

The current GRLDNN model is an agent model built using a reinforcement learning framework and implemented partly using a deep neural network. Although the model is proposed as a generic agent model that can simulate a variety of cognitive and decision-making tasks, the model's architecture was originally inspired by an earlier model of the basal ganglia[10] as shown in Fig. 7A. The representational system is analogous to the cortico-striatal projections that are thought to be capable of compressing the cortical state and generate abstract

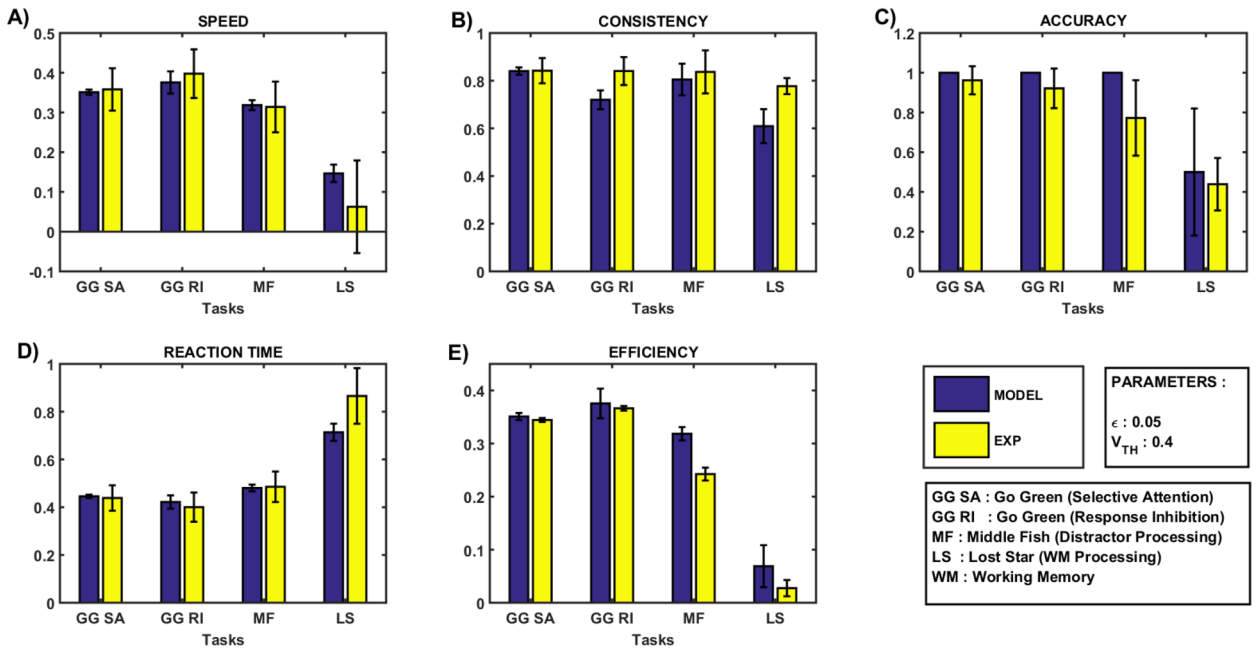

Figure 5. Comparison of performance of the GRLDNN model with the experimental results and data adapted from[8] (**A**) Speed (**B**) Consistency, (**C**) Accuracy, (**D**) Reaction Time, (**E**) Efficiency. EXP, experimental results; MODEL, Model performance Results.

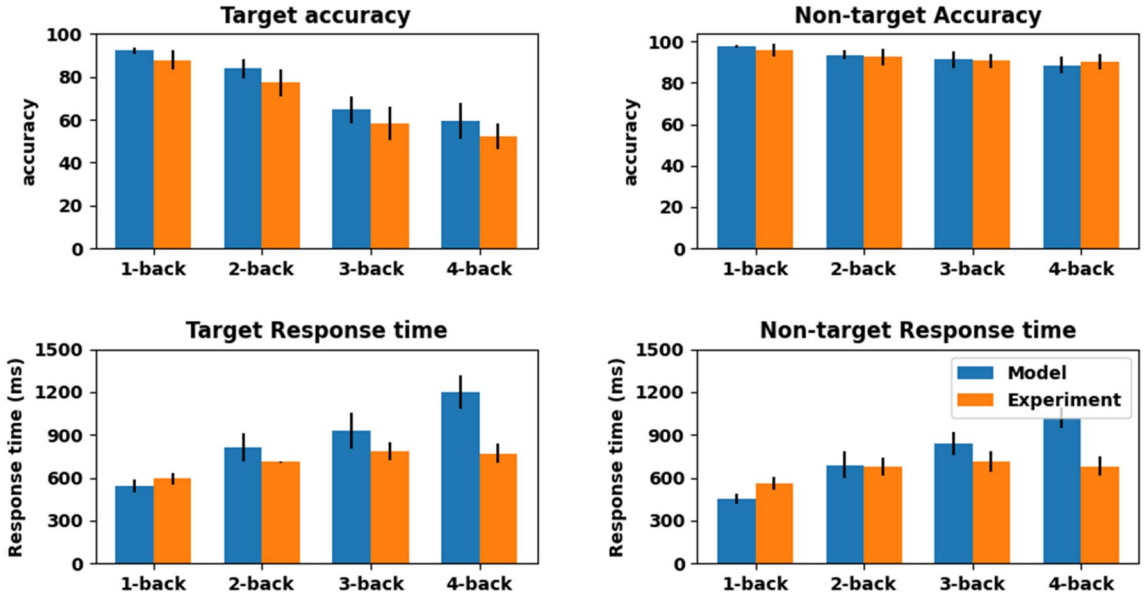

Figure 6. Comparison of performance of the GRLDNN model with the experimental results and data adapted from[25] (**A**) Accuracy when the current stimulus is the target (**B**) Accuracy for non-target stimulus , (**C**) Response time for Target Stimulus, (**D**) Response Time for non-Target Stimulus.

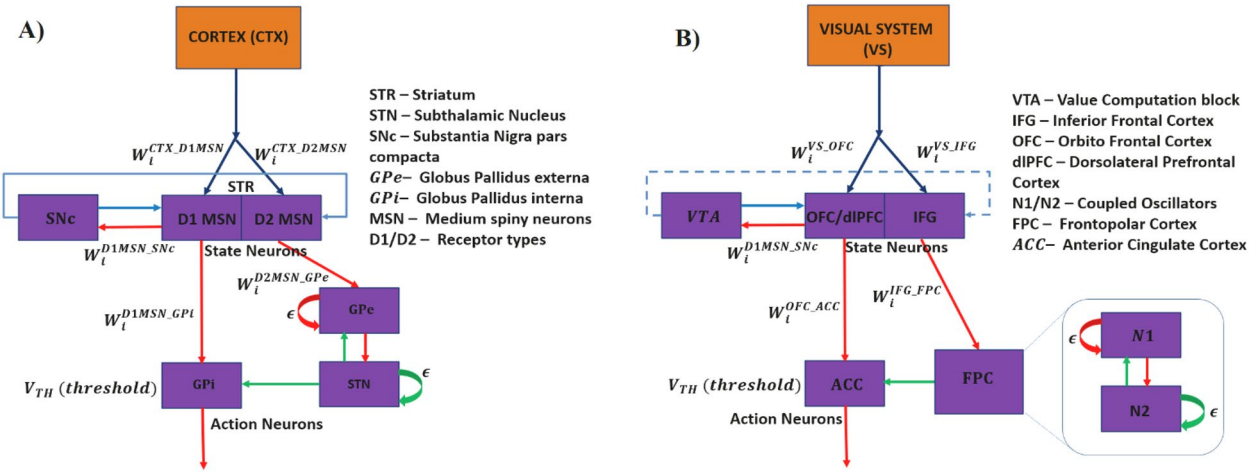

Figure 7. Biological Representation of GRLDNN Agent Model inspired by (**A**) Basal Ganglia (BG) architecture and (**B**) the equivalent representation of prefrontal cortex (PFC) Architecture. SNc, Substantia Nigra pars compacta; STR, Striatum; GPi, Globus Pallidus interna; GPe, Globus Pallidus externa; STN, Subthalamic Nucleus; MSN, Medium Spiny neurons; STN-GPe forms a coupled oscillator system with interconnectivity weights W between STN and GPe and ϵ is the lateral connection strengths among the neurons of STN and GPe. Action is selected based on the winner neuron crossing the threshold (V_{TH}) first. VTA, ventral tegmental area; OFC, Orbitofrontal cortex; DLPFC, dorsolateral prefrontal cortex; IFG, inferior frontal gyrus; FPC, frontopolar cortex; ACC, anterior cingulate cortex; FPC is modeled using a coupled oscillator system with interconnectivity weights W between oscillators N1 and N2 and ϵ is the lateral connection strengths among the neurons of N2. OFC/DLPFC to ACC weights are updated using Q-learning.

representations of the same[26]. The memory system is analogous to the striatum proper, and the flip-flops are comparable to the medium spiny neurons (MSNs) of the striatum. The MSNs are known to exhibit UP/DOWN states, a property that is thought to subserve working memory functions[27,28]. In digital systems, flip-flops are used as memory elements that serve as building blocks to implement sequential logic. Thus, in the proposed agent model, the presence of flip-flop neurons in the memory system affords the model the ability to process sequences and perform decision-making functions thereon. The value computation block is analogous to substantia nigra pars compacta (SNc)—it integrates the outputs of the flip-flop neurons of the memory system and computes the value function. The connections from the memory system to the action selection block is analogous to the direct pathway of the basal ganglia. The longer route from the memory system to the explorer block and onward to the action selection block is analogous to the indirect pathway. In modeling literature that describes the decision-making functions of the basal ganglia using reinforcement learning, there is a subclass of models

that attribute the role of exploratory drive to the indirect pathway, which is essential to sample the action space randomly[10]. Finally, the action selection block itself is analogous to globus pallidus interna (GPi), the output port of the basal ganglia.

The proposed agent model can also be compared to another brain region known for its decision-making functions—the prefrontal cortex (PFC). Figure 7B shows the analogy between the components of the proposed GRLDNN agent model and areas of PFC whose contributions to decision making have been described extensively[29–32].

The dorsolateral prefrontal cortex (DLPFC) receives inputs from the primary and secondary sensory association cortices of the posterior brain (Klaus and Pennington 2019). The DLPFC is also considered to be the terminus of the dorsal visual pathway, also called the "where" or "how" pathway that determines how to use visual information by supplying such information to the decision-making mechanisms of the PFC[33]. The DLPFC is also known for its working memory functions, subserved by dopamine-receptor expressing neurons and gated by dopaminergic projections from mesencephalic regions[30]. Thus, the memory system in the proposed agent model is suitably comparable to DLPFC.

Single unit electrophysiological studies have shown the involvement of the orbitofrontal cortex (OFC) in value computation[31]. The role of OFC in value computation was also confirmed by functional imaging studies[29]. Since dopaminergic activity is strongly linked to reward signalling, projections from the ventral tegmental area (VTA) to PFC were implicated in the value computations of OFC[32]. Electroencephalographic studies[34], on subjects engaged in decision-making activities, have implicated the frontopolar cortex (FPC) in exploratory behavior. Thus, the exploratory block in the proposed model is comparable to FPC. The inferior frontal gyrus[35,36] is suggested to encode information about NoGo processes and has strong implications for action selection mechanisms, especially action stopping. On the other hand, Anterior Cingulate Cortex (ACC)[37,38] is suggested to encode information about the uncertainty in choices, hence important for estimating utility values of action choices. Currently only one level of working memory processing is tested in the model. Going forward this aspect will be incorporated where the impact of memory load (by increasing the number of stars and the perceptual levels) can be analysed both experimentally as well as in the model.

The current model also can be made more robust and has the scope of conducting patient profiling. By tuning the appropriate model parameters, we are able to match the experimental results, thereby demonstrating its potential use for profiling real patients. There is a scope to explore further and incorporate aspects of various disease conditions and cognitive disabilities into the model. With respect to the modeling of working memory, we have considered tests at only one difficulty level in our model. There is a scope to scale the model to adopt multiple levels of cognitive loads. The future work also includes integrating the cortical and the subcortical modules into a single framework. In the future, these modeling efforts could also be expanded to include emotion processing and modeling of electrophysiological signals. Altogether, this study opens doors to modeling various cognitive dimensions of the same individual through a unified agent-based modeling framework.

Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Further, inquires can be directed to the corresponding author. The MATLAB code of the proposed GRDLNN model (http://modeldb.yale.edu/267532) is available on the ModelDB server[39] and an access code will be provided on request.

Received: 18 August 2022; Accepted: 24 March 2023
Published online: 12 April 2023

References

1. Anderson, J. R. ACT-R: A theory of higher level cognition and its relation to visual attention. *Hum. Comput. Interact.* **12**, 439–462 (1997).
2. Laird, J. E. *The Soar Cognitive Architecture* (MIT Press, Cambridge, 2018). https://doi.org/10.7551/mitpress/7688.001.0001.
3. Young, R. M. & Lewis, R. L. The soar cognitive architecture and human working memory. *Models Work. Mem.* https://doi.org/10.1017/cbo9781139174909.010 (2012).
4. Rosenbloom, P. S., Demski, A. & Ustun, V. The sigma cognitive architecture and system: Towards functionally elegant grand unification. *J. Artif. Gen. Intell.* **7**, 1 (2016).
5. Laird, J. E., Lebiere, C. & Rosenbloom, P. S. A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Mag.* **38**, 13–26 (2017).
6. Millan, M. J. *et al.* Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. *Nat. Rev. Drug Discov.* https://doi.org/10.1038/nrd3628 (2012).
7. Weintraub, S. *et al.* Cognition assessment using the NIH Toolbox. *Neurology* **80**, S54–S64 (2013).
8. Balasubramani, P. P. *et al.* Mapping cognitive brain functions at scale. *Neuroimage* **231**, 117641 (2021).
9. Chakravarthy, V. S., Joseph, D. & Bapi, R. S. What do the basal ganglia do? A modeling perspective. *Biol. Cybern.* **103**, 237–253 (2010).
10. Chakravarthy, V. S. & Moustafa, A. A. *Computational Neuroscience Models of the Basal Ganglia*. *Movement disorders* vol. 15 (Springer Singapore, 2018).
11. Sridharan, D., Prashanth, P. S. & Chakravarthy, V. S. The role of the basal ganglia in exploration in A neural model based on reinforcement learning. *Int. J. Neural Syst.* **16**, 111–124 (2006).
12. Holla, P. & Chakravarthy, S. Decision making with long delays using networks of flip-flop neurons. in *Proceedings of the International Joint Conference on Neural Networks* vols 2016-October (2016).
13. Balasubramani, P. P., Chakravarthy, V. S., Ravindran, B. & Moustafa, A. A. An extended Reinforcement Learning model of basal ganglia to understand the contributions of serotonin and dopamine in risk-based decision making, reward prediction, and punishment learning. *Front. Comput. Neurosci.* **8**, 47 (2014).

14. De Oliveira, T. B. F., Bazzan, A. L. C., Da Silva, B. C. & Grunitzki, R. Comparing Multi-Armed Bandit Algorithms and Q-learning for Multiagent Action Selection: A Case Study in Route Choice. in *Proceedings of the International Joint Conference on Neural Networks* vols 2018-July (2018).
15. Nerurkar, P. A., Chandane, M. & Bhirud, S. Exploring convolutional auto-encoders for representation learning on networks. *Comput. Sci.* **20**, 273–288 (2019).
16. Lindsay, G. W. Convolutional neural networks as a model of the visual system: Past, present, and future. *J. Cogn. Neurosci.* **33**, 2017–2031 (2021).
17. Sutton, R. S. & Barto, A. G. *Reinforcement Learning, Second Edition: An Introduction—Complete Draft* (The MIT Press, Cambridge, 2018).
18. Gillies, A., Willshaw, D. & Li, Z. Subthalamic-pallidal interactions are critical in determining normal and abnormal functioning of the basal ganglia. *Proc. R. Soc. B Biol. Sci.* **269**, 545–551 (2002).
19. Kawahara, T. Coupled Van der Pol oscillators ? A model of excitatory and inhibitory neural interactions. *Biol. Cybern.* **39**, 37–43 (1980).
20. Packard, M. G. & Knowlton, B. J. Learning and memory functions of the basal ganglia. *Ann. Rev. Neurosci.* https://doi.org/10.1146/annurev.neuro.25.112701.142937 (2002).
21. Smith, Y., Bevan, M. D., Shink, E. & Bolam, J. P. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* https://doi.org/10.1016/S0306-4522(98)00004-9 (1998).
22. Vickers, D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* **13**, 37–58 (1970).
23. Mandali, A., Rengaswamy, M., Chakravarthy, V. S. & Moustafa, A. A. A spiking Basal Ganglia model of synchrony, exploration and decision making. *Front. Neurosci.* **9**, 191 (2015).
24. Rice, P. J. & Stocco, A. Basal ganglia-inspired functional constraints improve the robustness of q-value estimates in model-free reinforcement learning. in *Proceedings of ICCM 2017—15th International Conference on Cognitive Modeling* (2017).
25. Lamichhane, B., Westbrook, A., Cole, M. W. & Braver, T. S. Exploring brain-behavior relationships in the N-back task. *Neuroimage* **212**, 116683 (2020).
26. Bar-Gad, I., Goldberg, J. A., Bergman, H., Havazelet-Heimer, G. & Ruppin, E. Reinforcement-driven dimensionality reduction—A model for information processing in the Basal Ganglia. *J. Basic Clin. Physiol. Pharmacol.* **11**, 305–320 (2000).
27. Wilson, C. J. & Kawaguchi, Y. The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *J. Neurosci.* **16**, 2397–2410 (1996).
28. Ferbinteanu, J. Contributions of hippocampus and striatum to memory-guided behavior depend on past experience. *J. Neurosci.* **36**, 6459–6470 (2016).
29. Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W. & Rangel, A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* **28**, 5623–5630 (2008).
30. Klaus, K. & Pennington, K. Dopamine and Working Memory: Genetic Variation, Stress and Implications for Mental Health. in *Current Topics in Behavioral Neurosciences* vol. 41 (2019).
31. Setogawa, T. *et al.* Neurons in the monkey orbitofrontal cortex mediate reward value computation and decision-making. *Commun. Biol.* **2**, 126 (2019).
32. Takahashi, Y. K. *et al.* The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron* **62**, 269–280 (2009).
33. Takahashi, E., Ohki, K. & Kim, D. S. Dissociation and convergence of the dorsal and ventral visual working memory streams in the human prefrontal cortex. *Neuroimage* **65**, 488–498 (2013).
34. Bourdaud, N., Chavarriaga, R., Gálan, F. & Millán, J. D. R. Characterizing the EEG correlates of exploratory behavior. *IEEE Trans. Neural Syst. Rehabil. Eng.* **16**, 549–556 (2008).
35. Aron, A. R., Robbins, T. W. & Poldrack, R. A. Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* https://doi.org/10.1016/j.tics.2004.02.010 (2004).
36. Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J. & Robbins, T. W. Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* **6**, 115–116 (2003).
37. Heilbronner, S. R. & Hayden, B. Y. Dorsal anterior cingulate cortex: A bottom-up view. *Annu. Rev. Neurosci.* **39**, 149–170 (2016).
38. Monosov, I. E. Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nat. Commun.* **8**, 134 (2017).
39. McDougal, R. A. *et al.* Twenty years of ModelDB and beyond: Building essential modeling tools for the future of neuroscience. *J. Comput. Neurosci.* **42**, 1–10 (2017).

Author contributions

S.S.N.—Conceptualization; Model development; Implementation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing—original draft; V.R.M.—Conceptualization; Model development; Implementation; Data curation; Formal analysis; Investigation; Methodology; Validation; Writing—review & editing; V.C.—Model development; Implementation; Investigation; P.B.—Conceptualization; Data curation; Formal analysis; Writing—review & editing; Validation; J.M.—Formal analysis; Writing—review & editing; D.R.—Formal analysis; Writing—review & editing; V.S.C.—Conceptualization; Model development; Data curation; Formal analysis; Investigation; Methodology; Validation, Writing—review & editing; Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-32234-y.

Correspondence and requests for materials should be addressed to V.S.C.