nature communications

Article

https://doi.org/10.1038/s41467-025-60044-5

Basal ganglia deep brain stimulation restores cognitive flexibility and exploration-exploitation balance disrupted by NMDA-R antagonism

Received: 29 September 2024

Accepted: 12 May 2025

Published online: 28 May 2025

Check for updates

Nir Asch[1,2] ✉, Noa Rahamim[3], Anna Morozov[3], Uri Werner-Reiss[3], Zvi Israel[4], Rony Paz[2] & Hagai Bergman[1,3,4]

Learning thrives on cognitive flexibility and exploration. Subjects with schizophrenia have impaired cognitive flexibility and maladaptive exploration patterns. The basal ganglia-dorsolateral prefrontal cortex (BG-DLPFC) network plays a significant role in learning processes. However, how this network maintains cognitive flexibility and exploration patterns and what alters these patterns in schizophrenia remains elusive. Using a combination of extracellular recordings, pharmacological manipulations, macro-stimulation techniques, and mathematical modeling, we show that in the nonhuman primate (NHP), the external segment of the globus pallidus (GPe, the central nucleus of the BG network) modulates cognitive flexibility and exploration patterns (experiments were done in females only). We found that chronic, low-dose administration of N-methyl-D-aspartate receptor (NMDA-R) antagonist, phencyclidine (PCP), decreases directed exploration but increases random exploration, as seen in schizophrenia. In line with adaptive working-memory reinforcement-learning models of the BG-DLPFC network, low-frequency GPe macro-stimulation restores the balance of both exploration types. Our findings suggest that exploration-exploitation imbalance reflects abnormal BG-DLPFC activity and that GPe stimulation may restore it.

The ability to learn and adapt in changing environments (cognitive flexibility) provides a significant evolutionary advantage. The exploration-exploitation (E-E) balance is key for adaptive behavior, much like excitation-inhibition balance in neural circuits. Exploration enhances learning, while exploitation optimizes decisions based on known information. Exploration, driven by behavioral policy and environmental state, facilitates learning by exposing individuals to unfamiliar or uncertain elements. As a meta-behavioral policy, cognitive flexibility enables learning through exploration, which occurs in two forms: directed exploration, focused on information-seeking, and random exploration, introduced stochastically into decision-making[1–5]. These strategies complement each other—directed exploration is crucial when knowledge is limited or evolving, while random exploration helps avoid suboptimal solutions when knowledge is abundant. Notably, random exploration is not purely stochastic; it varies with behavioral state, increasing in safe conditions and shifting to exploitation during perilous periods[6].

Research on the neural mechanisms subserving learning and cognitive flexibility has chiefly focused on the DLPFC and the BG[7–11]. These well-integrated regions[12–14] are recognized as the major

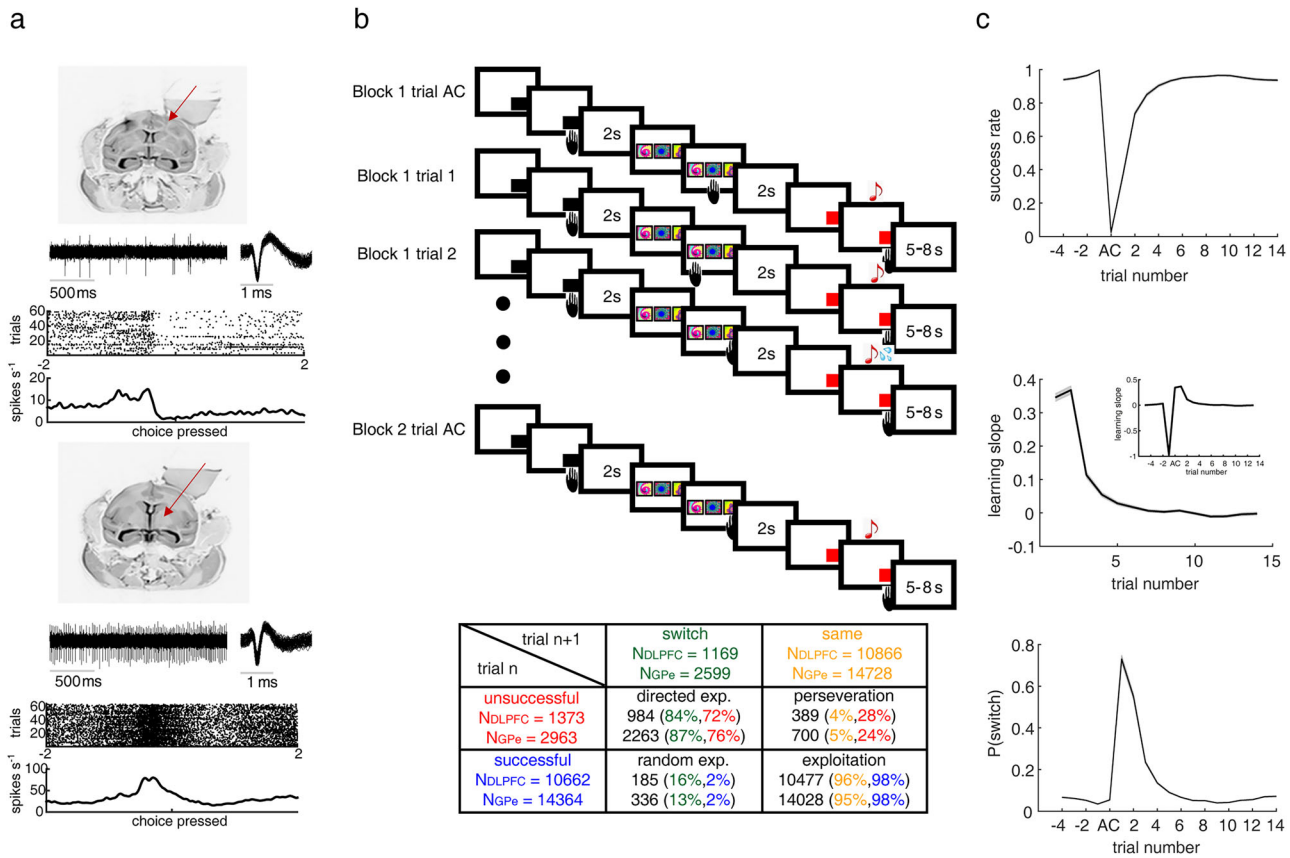
[1]Department of Medical Neurobiology, The Hebrew University of Jerusalem, Jerusalem, Israel. [2]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel. [3]Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem, Israel. [4]Functional Neurosurgery Unit, Department of Neurosurgery, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. ✉e-mail: nir.ash@mail.huji.ac.il

underlying neural circuitry of goal-directed (probably governed by the cortico-cortical network) and implicit/habitual (hypothesized to be controlled by sub-cortical structures, e.g., the basal ganglia network) learning paradigms[7,15–17]. The GPe, the central nucleus of the main axis of the BG network, connected to both input and output BG layers[17–19], and the DLPFC[3,20,21] are, therefore, possible hubs for controlling the E-E balance. The BG has been suggested to function as a generator of random exploration[22,23], as well as an active gating mechanism, dynamically regulating the cortical working memory (WM) system[23–25]. Memory is critical for the exploitation of learned knowledge and the efficient exploration of new options. Therefore, the dynamic gating of incoming stimuli into WM considerably influences the E-E balance.

Subjects with schizophrenia exhibit impaired cognitive flexibility[26] in tasks like reversal learning and multiple-arm bandit, showing reduced directed exploration and increased random exploration[27,28]. These deficits are linked to WM impairment and DLPFC dysfunction[29,30], as well as hyperactivity and enlargement of the GPe, which correlates with symptom severity[31,32]. The increased random exploration may stem from an overactive "exploration generator" (the GPe)[33] or WM deficits.

To better understand the neural mechanisms underlying cognitive flexibility and the E-E balance in both healthy and schizophrenia-relevant states, we recorded neurobehavioral activity in two female African green monkeys (*Chlorocebus aethiops sabaeus*, ~4 kg) across three distinct stages: (1) the naïve state, before any drug administration; (2) under the effect of PCP, an NMDA-R antagonist, during 28 days of chronic administration (using a subcutaneous osmotic pump) to induce learning and memory deficits[34] associated with the negative and cognitive symptoms of schizophrenia[35–39]; and (3) post-PCP, after a washout period of at least two weeks following PCP discontinuation.

These NHPs were extensively trained on a deterministic three-armed bandit (reversal-learning) task while neural activity was recorded from the GPe and DLPFC (Fig. 1a, b), two cytoarchitectonically preserved regions that are highly similar across primates[40]. Unlike probabilistic tasks[3], this design allowed us to better distinguish directed from random exploration. We defined exploration following reward omissions as directed (seeking a new cue-reward association) and exploration after rewarded choices as random (despite no prediction-outcome mismatch). By examining DLPFC-GPe dynamics in the naïve state and under PCP, we investigated how NMDA-R



trial n \ trial n+1	switch	same
	N_{DLPFC} = 1169 N_{GPe} = 2599	N_{DLPFC} = 10866 N_{GPe} = 14728
unsuccessful N_{DLPFC} = 1373 N_{GPe} = 2963	directed exp. 984 (84%,72%) 2263 (87%,76%)	perseveration 389 (4%,28%) 700 (5%,24%)
successful N_{DLPFC} = 10662 N_{GPe} = 14364	random exp. 185 (16%,2%) 336 (13%,2%)	exploitation 10477 (96%,98%) 14028 (95%,98%)

Fig. 1 | Experimental setup and healthy behavioral performance. a Top - MRI of the non-human primates' (NHPs) brain and recording chamber. The red arrow points to the dorsolateral prefrontal cortex (DLPFC). Middle, extracellular recordings of DLPFC exemplary neuron (left) and 100 randomly chosen superimposed spike waveforms of the recorded cell (right). Bottom, Raster display and a post-stimulus time histogram (PSTH) of the neuron's firing during 60 trials around choice selection (two seconds before and two seconds after). Bottom—The same as the top image, but for the external segment of the globus pallidus (GPe). **b** Top—Task design. NHPs had to identify the hidden association change and learn the new association. The first line represents an association change trial (AC) in which the NHP is unaware of the change of the new association and, therefore, chooses the wrong cue. The second line represents the next trial in the block in which the NHP changes its choice to another incorrect choice. In the third line, the NHP changes again, this time receiving a reward. The fourth line represents the following block, where the association changes without the NHPs' knowledge. The NHP then chooses the same stimulus as before, but this time receives no reward. Bottom—A table showing the amount (and proportion) of recorded trial types for both brain regions. Rows indicate what happened in the previous trial (trial *n*) and columns indicate the current trial (*n* + 1). Text color coding indicates the fraction of recorded cells within each category of the total sum. For example, our neural data set of perseveration trials consists of 389 DLPFC and 700 GPe recordings, constituting 28% and 24% of all neural-recorded trials following an unsuccessful trial, respectively, and 4% and 5% of all neural-recorded trials following same-choice trials. **c** Top—The learning curve, i.e., the NHPs' probability of choosing successfully. 'AC' indicates the association change trial; trials prior/post to the AC trial are of negative/positive sign, respectively. Middle—The learning slope (i.e., the derivative of the learning curve). Inset—the learning slope across the AC trial. Bottom—Switch probability.

antagonism disrupts cognitive flexibility and E-E control, paralleling schizophrenia-related deficits and informing potential treatments.

Results

Naïve behavior and associated neural activity patterns

Two NHPs were trained to acquire and reacquire their response to novel visual cues in a between-block design (Fig. 1b). First, they had to identify the spatial cue associated with the reward outcome. Once learned (reaching a predefined success criterion; 12–15 successful trials out of a maximum of the last 25 trials. Criterion selected randomly for each block to avoid identification of the temporal rule), the association changed (AC), and a new block began without any external sign. Thus, in the first trial of every block (AC trial), the NHPs experienced a prediction outcome mismatch, in which choosing the previously correct cue did not elicit a reward. Since our task is deterministic and our NHPs were excessively trained, they had to initiate directed exploration to increase their knowledge of the task and maximize their gain. The NHPs rapidly identified the AC and initiated directed exploration in 73% of the AC + 1 trials (Fig. 1c, bottom). During directed exploration, the NHPs showed fast learning dynamics expressing well-functioning WM, usually avoiding previously selected erroneous cues in search of the new association (Fig. 1c and Supplementary Fig. 1a, b). An optimal learning agent would find the correct response with an average of 1.5 trials. The NHPs found the cue associated with a rewarding outcome within 3.11 ± 0.04 trials (mean ± SEM). Once the new correct cue was identified, the NHPs exploited their newly acquired knowledge to maximize their gain, completing the learning phase (defined as three sequential successful trials) within 5.44 ± 0.05 trials. Post-learning and during the learning plateau, the NHPs achieved a mean success rate of 93% (Fig. 1c top). Their response times further reflected their cognitive effort, lengthening during the learning phase and shortening thereafter (Supplementary Fig. 1a-lower subplot). The NHPs performed a total of 1673 blocks during the neuronal recording sessions in the naïve state. The average number of trials per block was 18.8 trials (ranging between 13 and 40).

To complete the task successfully, our NHPs had to identify success, associate it with cue choice, carry the information across trials, and act accordingly. We, therefore, tested whether either brain area encodes reward by comparing their activities in response to rewards and reward omissions. We found that both brain regions elevate their firing rate (FR) in response to reward omissions (Fig. 2a-left and Supplementary Fig. 2b). This elevation is then carried to the subsequent trial, transiently before cue choice in the DLPFC, and spanning the entire subsequent trial in the GPe (Fig. 2a-right). Similarly, both regions elevate their FRs before the cue choice in exploratory (directed and random together) compared to non-exploratory (perseveration and exploitation together) trials (Fig. 2b).

Cue choice should be based on the success of the previous trial. Switching from the previous choice at the $n+1$ trial (directed or random exploration) or not (exploitation or perseveration) is dependent on the outcome of the n-trial (Fig. 1b). Assuming optimal behavioral policy, the NHPs should maintain the same choice (exploitation) following a successful trial and switch to a different cue (directed exploration) following an unsuccessful trial (Fig. 2c). Like all biological creatures, our NHPs were imperfect. They did not explore after every prediction outcome mismatch (perseveration errors occurred after 21.8% of unsuccessful trials). Neither did they exploit after every congruent prediction outcome (random exploration occurred after 2.25% of successful trials). Alternatively, they may employ random exploration to test the optimality of their current behavioral policy. Even though both areas elevated their FRs before choice following an unsuccessful trial (Fig. 2a), neither expressed a significant difference between directed exploration and perseveration (Fig. 2d and Supplementary Fig. 3 and 4). In contrast, the GPe elevated its activity before the choice of random exploration compared to exploitation in trials following success, whereas the DLPFC did not (Fig. 2e). In sum, both regions significantly elevated their FRs in trials following an unsuccessful trial leading to either directed exploration or perseveration, while the GPe also elevated its activity before random exploration choices (Fig. 2f).

To further disentangle the GPe and DLPFC correlates of the NHPs' behavior in the deterministic three-armed bandit task, we expanded our analysis from a single trial comparison to several trials and compared the neural dynamics to the behavioral ones during the learning phase (Fig. 3). We found that both regions encode the reward prediction error by their dynamic of discharge rates (Fig. 3a). Additionally, the dynamics of GPe activity highly correlated with switch probability (exploration), while DLPFC activity correlated only marginally (Fig. 3b). Switch probability (especially in the early part of the block) is strongly associated with the prediction-outcome relationship of the previous trial. Nonetheless, while directed exploration is driven by the prediction-outcome mismatch of the preceding trial, its direction is determined by new predictions based on the memory of past events' outcomes. A switch can either be successful or unsuccessful. While choosing successfully in the first trial after reversal is a sheer chance, switching successfully later in the block demands higher WM effort, remembering previously chosen incorrect trials and predicting the location of the correct stimulus cue. Hence, we compared the correlation between both regions' neural activity and all switch probability types: general, successful (associated with WM load), and unsuccessful switch probabilities (Fig. 3b-right). The DLPFC was highly correlated with the likelihood of a successful switch and did not correlate with the chance of an unsuccessful switch. The GPe, on the other hand, exhibited significant correlations with both successful and unsuccessful switch probabilities.

Learning dynamics depend on two key factors**:** exploration initiation and prediction accuracy—the faster these processes, the quicker the learning. To assess this, we compared neural activity with the learning slope (i.e., the derivative of the learning curve). While both regions correlated with learning speed, the DLPFC showed a much stronger correlation with the learning slope (Fig. 3c) than with switch probability (Fig. 3b). These findings suggest that DLPFC activity enhances memory efficiency, while GPe activity regulates exploratory behavior. Supporting this, GPe units responded similarly to all choice stimuli, whereas DLPFC units showed a more distinct response pattern for each stimulus (Supplementary Fig. 4a).

We then compared the activity dynamics of the two regions throughout the task. While their concurrent activities only marginally correlated, comparing their activities in subsequent trials (i.e., comparing the correlation between GPe activity in trial 'n' with DLPFC activity in trial 'n + 1') exhibited a strong correlation (Fig. 3d). These neuronal correlates show that GPe activity increases mainly after prediction-outcome mismatch and remains elevated in the following trial, as well as before directed and random exploratory actions. The DLPFC, on the other hand, increases its activity transiently, mainly before cue choice in subsequent trials associated with WM load and the prediction of the following optimal action process.

To directly examine DLPFC-GPe interactions, we analyzed concurrent local field potential (LFP) activity during task performance across all states. First, we examined naïve state LFP activity during the task in three frequency bands: theta (4–7 Hz), beta (13–30 Hz), and low gamma (30–40 Hz) (Supplementary Fig. 5). While all three bands showed increased activity in response to reward, only theta activity increased before cue choice in both brain regions (Fig. 4a, Supplementary Fig. 5). Cross-correlation analysis of the mean theta-band filtered LFP envelopes around cue choice revealed a strong correlation between the two regions without a significant lag (Fig. 4a, middle).

We applied Granger causality analysis to assess the causal influence between the two regions. This method quantifies how much of one region's LFP variance can be predicted by the other, providing a

https://doi.org/10.1038/s41467-025-60044-5

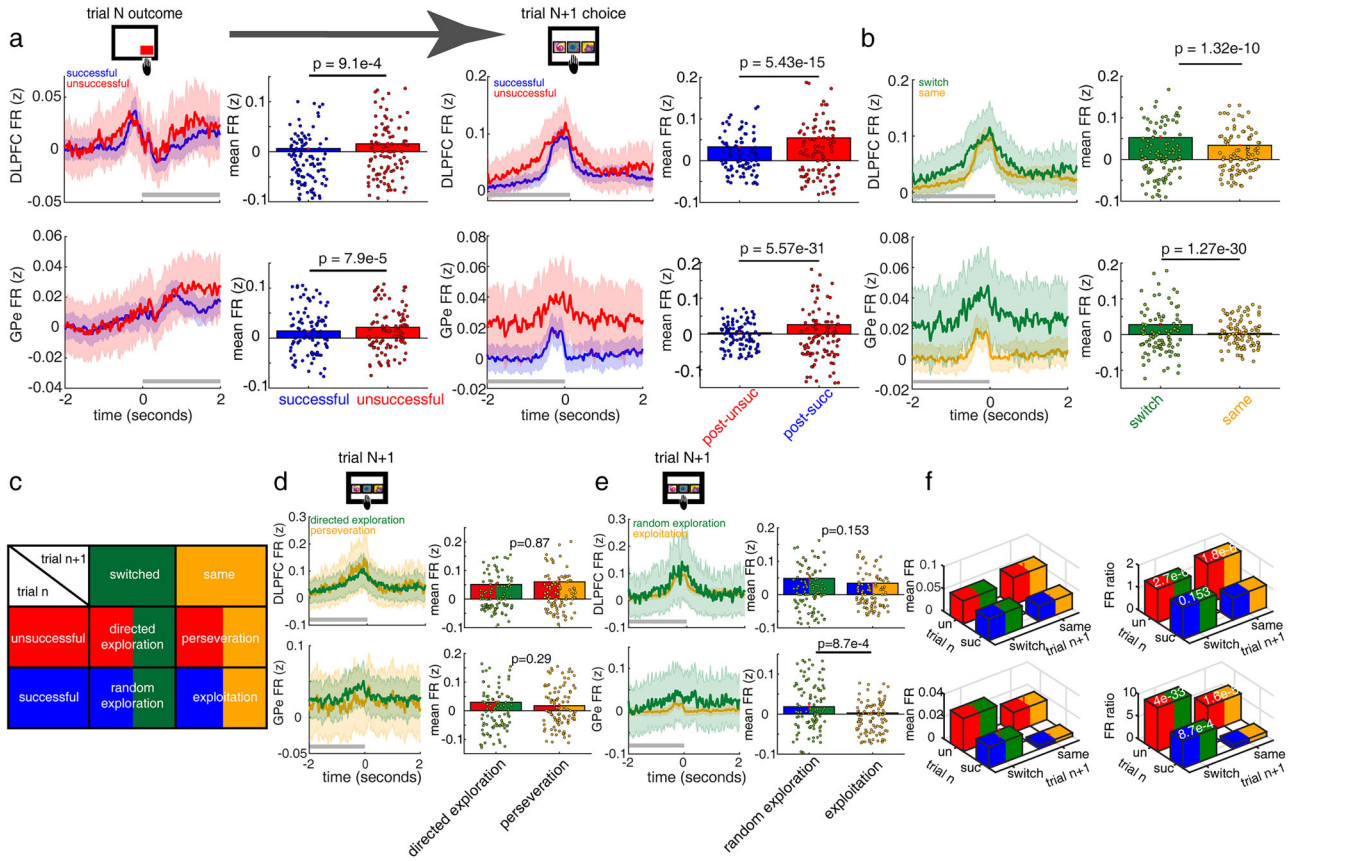


Fig. 2 | GPe and DLPFC encoding of exploration-exploitation behavior. **a** dorsolateral prefrontal cortex (DLPFC) (top, 12,035 recorded trials of 325 neurons) and external segment of the globus pallidus (GPe) (bottom, 17,327 recorded trials of 233 neurons) mean ± SEM z-normalized firing rates (FRs) around the reward outcome of trial *N* and subsequent cue choice in trial *N* + 1. The shaded gray area indicates the window for calculating the mean FR, as shown in the bar graphs on the right. Left—FRs around the reward outcome in trial N for successful (blue, reward) and unsuccessful (red, no reward) trials, with a z-score baseline from the two seconds before reward claiming. Right—FRs around cue choice in trial *N* + 1, with a z-score baseline from the two seconds before trial initiation. *p*-values are from two-tailed *t* tests comparing FRs. **b** DLPFC (top, 12,035 recorded trials of 325 neurons) and GPe (bottom, 17,327 recorded trials of 233 neurons) mean ± SEM z-normalized FRs around cue choice in exploratory trials (green, cue switch) and non-exploratory trials (orange, same cue as previous trial). *p*-values are from two-tailed *t* tests comparing FRs. **c** Trial type definitions: (1) Directed exploration—following unsuccessful trials with a choice switch. (2) Perseveration—following unsuccessful trials without a choice switch. (3) Random exploration—following successful trials with a choice switch. (4) Exploitation—following successful trials without a choice switch. Colors match those in panels (**a**) and (**b**). **d** Comparison of DLPFC (top, 1373 recorded trials of 325 neurons) and GPe (bottom, 2963 recorded trials of 233 neurons) FRs around choice selection in directed exploration and perseveration. Bar graphs show the mean ± SEM FR during the two seconds preceding cue choice. *p*-values (Bonferroni corrected for multiple comparisons) are from two-tailed *t* tests. **e** Comparison of DLPFC (top, 10,662 recorded trials of 325 neurons) and GPe (bottom, 14,364 recorded trials of 233 neurons) FRs in random exploration and exploitation. Bar graphs show the mean ± SEM FR during the two seconds before choice selection. *p*-values (Bonferroni corrected for multiple comparisons) are from two-tailed *t* tests. **f** Left—Mean ± SEM DLPFC and GPe FR leading to cue choice in the four trial types. Right—FR ratio relative to exploitation trials. *p*-values (Bonferroni corrected for multiple comparisons) are from two-tailed t-tests comparing FRs between random exploration, perseveration, directed exploration, and exploitation trials. Each bar chart is overlaid with 100 randomly selected data points falling within one standard deviation of the mean. For the full distribution of data points, please see Supplementary Fig. 3.

measure of directional connectivity. Additionally, performing this analysis in the frequency domain allows us to identify causal influences specific to distinct brain rhythms[41]. We analyzed the causality between the GPe and the DLPFC LFPs using a nonparametric variant of the Granger causality test[42]. This analysis revealed that while both regions exerted causal influence on each other, the GPe had a significantly stronger causal effect on the DLPFC than vice versa in the theta frequency band (Fig. 4a, bottom).

PCP Impairs E-E balance and learning while altering DLPFC-GPe neural dynamics and information flow

We then examined the effect of chronic, low-dose PCP treatment on neural activity and concurrent task performance. PCP was given by mini-osmotic pump at a daily dose of 1.68 mg/kg for 28 days. Under PCP administration, the NHPs exhibited noticeable negative symptoms. While in their shared cage, their social interactions were significantly reduced. They did not take part in grooming activities and mostly kept to themselves. Motorically, they were slower. Nonetheless, they maintained their coordination and ability to move, run, jump, and climb. Notably, PCP detrimentally affected the NHPs' ability to perform the task, resulting in reduced directed exploration, slower learning, increased random exploration, and a lower success rate plateau (Fig. 5a–c). The NHPs' tendency to initiate directed exploration in the AC + 1 trial slightly decreased from 73% in the naïve state to 69% under PCP administration. More robustly, their ability to produce reliable predictions based on past choices decreased to chance level during directed exploration. This performance did not improve with task progression and accumulation of knowledge, as indicated by their reduced ability to successfully switch during the learning phase (Fig. 5a, right inset, and Supplementary Fig. 6a). Consequently, the identification of the new correct response (Supplementary Fig. 6a-bottom), and the achievement of the learning criterion (Fig. 5b), were delayed. Overall, the NHPs showed a significant decrease in directed exploration behavior and a significant increase in random exploration behavior (Fig. 5c),

https://doi.org/10.1038/s41467-025-60044-5

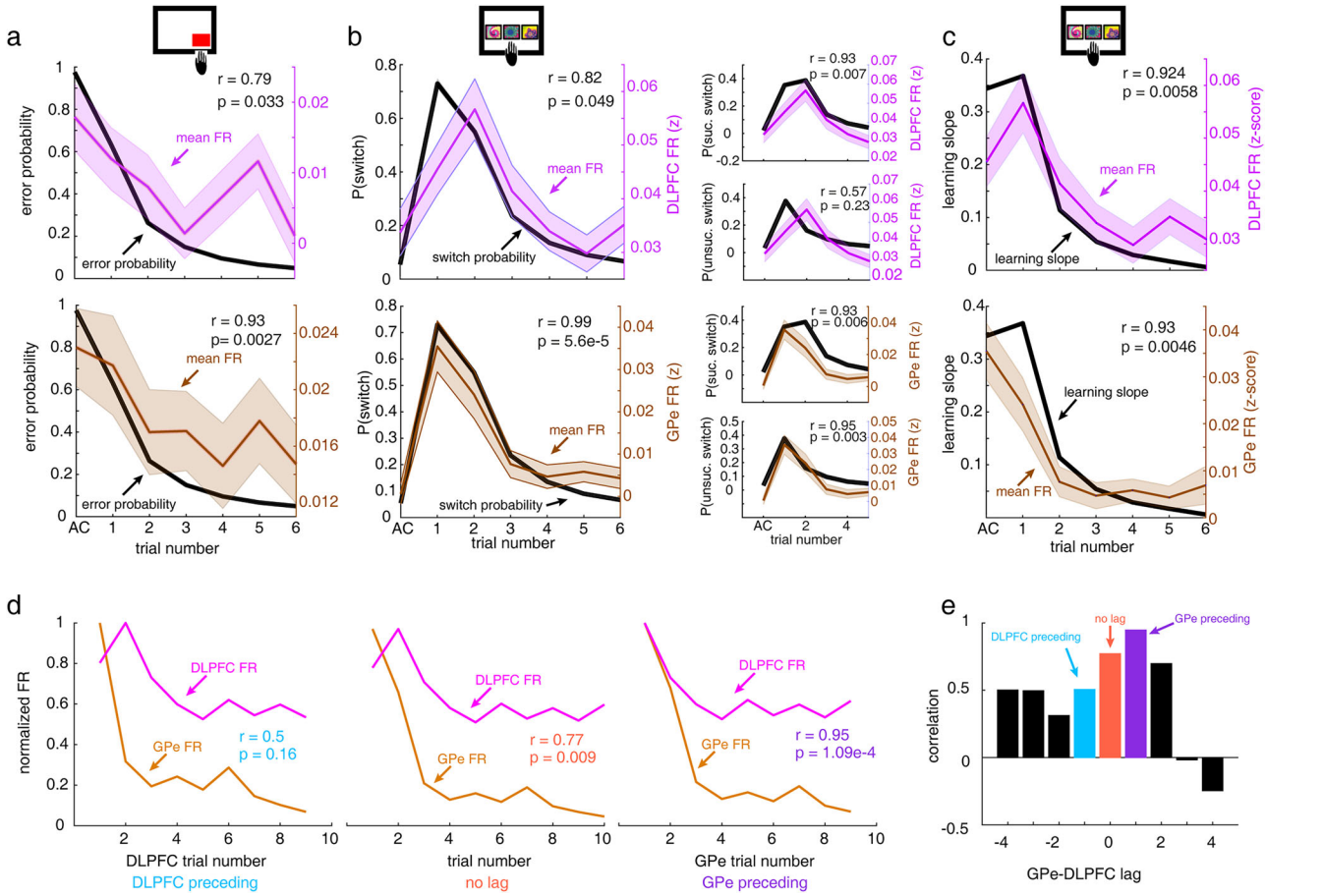


Fig. 3 | The dynamics of GPe activity correlate with exploratory behavior, while DLPFC activity lags and correlates with task knowledge. a mean ± SEM dorsolateral prefrontal cortex (DLPFC, top) and external segment of the globus pallidus (GPe, bottom) firing rate (FRs) of the two seconds ensuing reward outcome (purple and brown, respectively) with the probability of reward omission (black). Correlation values between switch probability and neural activity and their corresponding *p*-values are represented by 'r' and 'p,' respectively. **b** Mean ± SEM DLPFC (top) and GPe (bottom) FRs of the two seconds preceding choice selection with the non-human primates' (NHPs') probability of switching to a new key (left), the probability of switching to the new rewarded key (making a successful switch, finding the new association, right top figure), and the probability of switching to an unrewarded key (unsuccessful switch, right bottom figure). **c** Mean ± SEM DLPFC (top) and GPe (bottom) firing rate of the two seconds preceding choice selection with the NHPs' learning slope (black line). **d** Correlation between mean DLPFC and GPe firing rates (recorded during the two seconds leading to choice selection) during the first ten trials. Left—Comparing DLPFC activity in trial N with GPe activity in trial N + 1. Middle—Comparing DLPFC and GPe concurrent activities. Right—Comparing DLPFC activity in trial N + 1 with GPe activity in trial N. **e** Correlation values between DLPFC and GPe activities calculated from −4 (DLPFC activity precedes GPe activity by four trials) to +4 (GPe activity precedes DLPFC activity by four trials) trials lag. Bar colors correspond with the correlation color text of subplot (**d**). All correlation calculations were made using Pearson's correlation and *p*-values, Bonferroni corrected for multiple comparisons.

leading to slower learning and a reduced learning plateau success rate (Fig. 5a). Furthermore, the response times for stimulus choice selection increased throughout the block (Supplementary Fig. 6b), probably indicating the NHPs' growing uncertainty or reduced motivation. Response times for claiming reward outcomes no longer correlated with task performance (Supplementary Fig. 6c). Finally, following the cessation of PCP treatment, the NHPs' performance improved across all parameters (Fig. 5a–c and Supplementary Fig. 6).

Comparing the neural dynamics and behavioral choices throughout the task allowed for a complementary understanding of the PCP-induced neuro-behavioral changes. Under the effect of PCP, DLPFC activity encoded the reward prediction error better but no longer correlated with the learning slope (Fig. 5d-top). On the other hand, GPe activity no longer reliably encoded the reward prediction error; however, it remained correlated with the probability of making exploratory actions, although with a slightly reduced correlation value (Fig. 5d bottom). Furthermore, DLPFC activity leading to cue choice, decreased, whereas GPe activity increased (Fig. 5e). Next, we compared the activities of both regions leading to cue choice in post-successful and post-unsuccessful trials (Fig. 5f). We found that DLPFC activity in both trial types decreased and, notably, the neural activity was no longer discriminative between post-successful and post-unsuccessful trials (Fig. 5f, and Supplementary Fig. 6d). Conversely, GPe activity remained discriminative between the two trial types (Supplementary Fig. 6d). While GPe activity during post-unsuccessful trials showed a non-significant increase, GPe activity in post-successful trials increased fourfold (Fig. 5f), corresponding to the rise in random exploration probability. This heightened GPe activity and exploratory behavior may reflect an altered decision-making process, potentially resembling the hypervigilance or heightened awareness observed in early-stage psychosis. These neuro-behavioral changes in GPe activity and exploration patterns support the correlation found in the naïve state between GPe activity and exploratory behavior. A ceiling effect of the increased GPe activity may contribute to the reduced correlation between switch probability and GPe activity under the PCP effect. We further explored the neural changes induced by PCP during rest and found a significant increase in LFP gamma activity (Supplementary Fig. 7) and a decrease in GPe 'pause'[43] duration, which persisted (Supplementary Fig. 8). These results are consistent with the observed effects of Ketamine[44] and REM sleep[45] on the GPe pausing activity.

https://doi.org/10.1038/s41467-025-60044-5

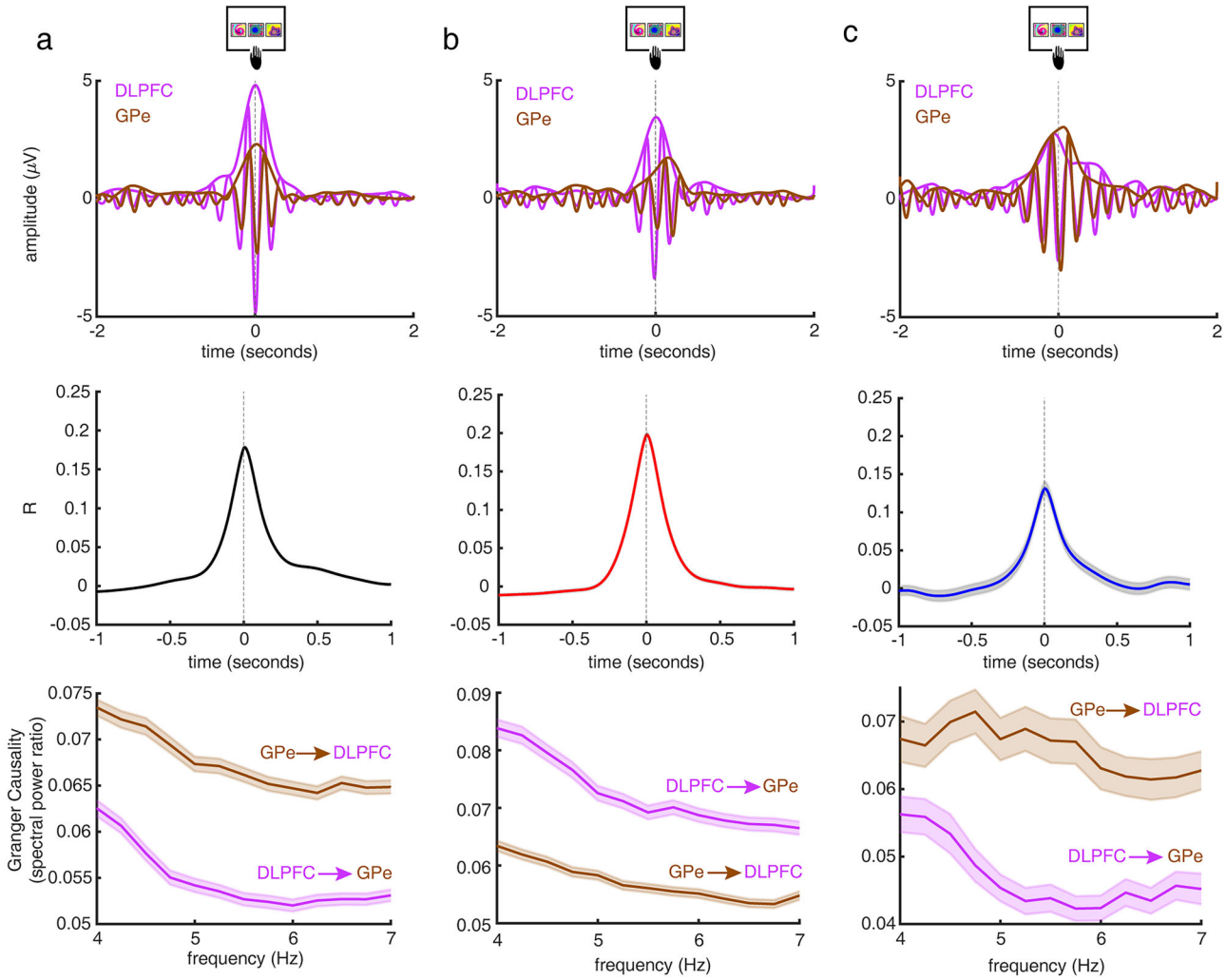


Fig. 4 | LFP theta activity around cue choice alters under PCP administration. **a** Analysis of local field potential (LFP) theta activity around cue choice in the naive state. Top—Mean theta-band (4–7 Hz) filtered LFP activity of the external segment of the globus pallidus (GPe, brown) and dorsolateral prefrontal cortex (DLPFC, purple), along with their corresponding envelopes. Middle—Mean ± SEM of the cross-correlation between the theta-band filtered envelopes of the DLPFC and GPe, computed across all individual envelope pairs. Bottom—Mean ± SEM of the Granger causality analysis of theta-band activity around cue choice (two seconds before until two seconds after) across the two brain regions. **b** The same as subplot a for the phencyclidine (PCP) period. **c** The same as subplot a for the post-PCP period.

Finally, We examined the effect of PCP on theta-band dynamics and connectivity between the DLPFC and GPe (Fig. 4b, c). PCP reduced and delayed GPe theta activity around cue choice (Fig. 4b top). Cross-correlation of theta-band filtered LFPs from the GPe and DLPFC showed zero lag (Fig. 4b middle), while Granger causality analysis revealed a stronger causal influence of the DLPFC on the GPe under PCP compared to the naïve state.

Discontinuation of PCP restored the NHPs' learning and neural dynamics around cue choice towards those observed during the naïve state (Fig. 5a–d). Concomitantly, GPe LFP theta oscillatory activity—linked to decision processes[46], attention, and WM[47,48], no longer lagged behind DLPFC activity and once again exerted a stronger causal influence on it (Fig. 4c). Interestingly, LFP gamma activity associated with attention and WM[47,48], but also with dissociative states such as following ketamine administrations and dreams[49,50], increased substantially in both regions (Supplementary Fig. 7). Compared to the naïve state, post-PCP GPe activity during post-unsuccessful trials significantly increased (Fig. 5f-left). On the other hand, GPe elevated activity during post-successful trials was found to be insignificant (Fig. 5f-right). These results correspond to the observed increase in directed exploration probability and contrary to the decrease in random exploration probability (Fig. 5c). Furthermore, DLPFC activity correlated again with the learning slope, further increasing its correlation with prediction-outcome mismatch probability, and the NHPs regained their ability to produce accurate predictions. GPe activity, on the other hand, did not regain its ability to encode the reward prediction error but regained its high correlation with exploratory choices. Behaviorally, post-PCP, the NHPs initiated directed exploration faster, making better predictions for finding the correct cues, thus resulting in faster learning.

In summary, PCP reduced directed exploration and increased random exploration while impairing the NHPs ability to predict the correct cue. Neuronally, DLPFC activity slightly decreased. Its correlation with reward omission probability increased but no longer correlated with the learning slope. On the other hand, GPe activity increased. It did not correlate with the probability of reward omission but remained correlated with the likelihood of exploratory actions. Furthermore, GPe activity increased significantly during trials following success, similar to the increase in random exploration and contrary to the decrease in directed exploration. Importantly, GPe LFP theta activity lagged, and the causal relationship between the two brain areas was reversed. Surprisingly, post-PCP, the NHPs' performance improved, making better predictions for the correct cue, increasing the probability of making directed

https://doi.org/10.1038/s41467-025-60044-5

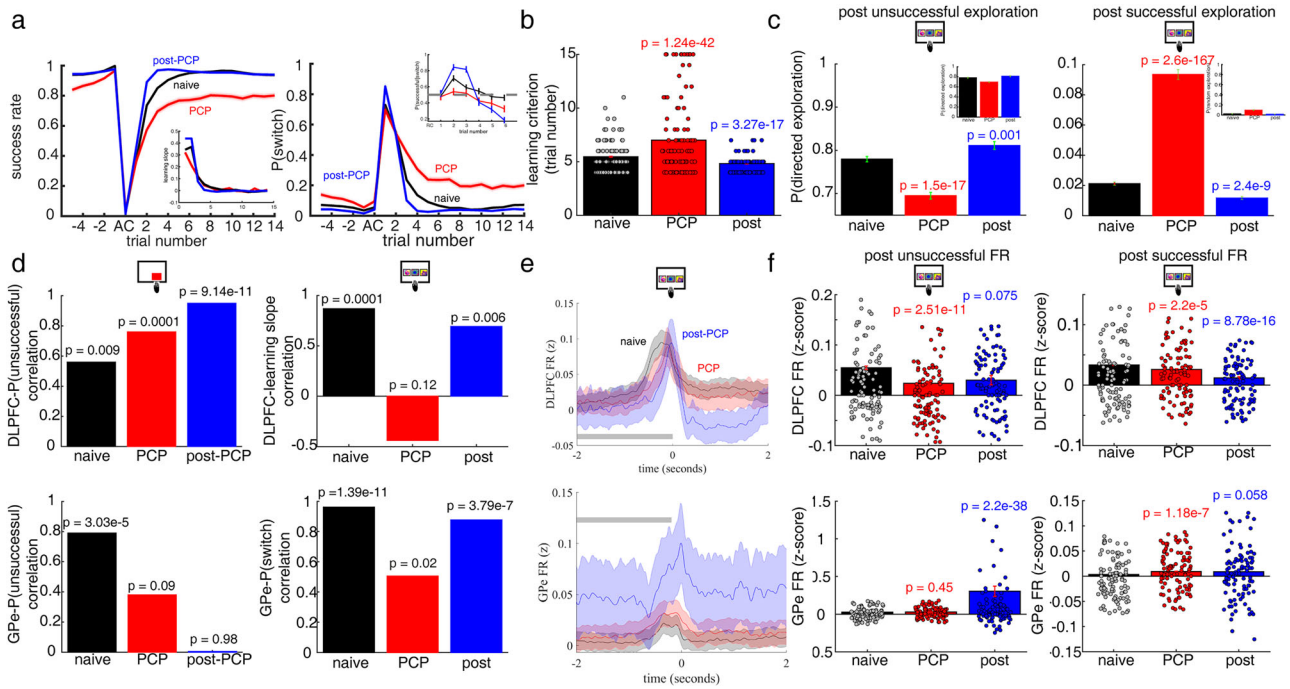


Fig. 5 | PCP robustly affects behavior, DLPFC, and GPe activity. a Left—non-human primates' (NHPs) mean ± SEM learning curves in the naïve (black), phencyclidine (PCP, red), and post-PCP (blue) conditions. The inset shows associated learning slopes. Right—Average switch probability throughout the task, with the inset showing the probability of choosing the correct stimulus in the first six trials following a switch (a dashed gray line marks the chance level). **b** Learning criterion: average and SEM with 100 randomly selected data points overlaid. of the trial number when learning was achieved, defined as three consecutive successful trials. *p*-values represent the significance (Bonferroni corrected) of the two-tailed *t* test comparing learning criterion under PCP effect and post-PCP with the naïve state. **c** Left—Directed exploration probability, mean ± SEM (switch probability after an unsuccessful trial, with a total of 3202 trials in the naïve state, 5040 under the PCP effect, and 355 after PCP withdrawal). Right—Random exploration probability (switch probability after a successful trial, total of 18,869 trials in the naïve state, 13,159 under PCP effect, and 2715 after PCP withdrawal). Insets show data on a 0-1 *Y*-axis. *p*-values represent the significance (Bonferroni corrected) of the two-tailed t-test comparing directed and random exploration probabilities in PCP and post-PCP conditions to the naïve state. **d** Left—Pearson's correlation of the dorsolateral prefrontal cortex (DLPFC, top) and external segment of the globus pallidus (GPe, bottom) activity with the probability of reward omission. Right—Correlation of DLPFC activity with learning slope (top) and GPe activity with switch probability (bottom). A total of 14,332 trials in the GPe and 10,708 trials in the DLPFC were recorded in the naïve state; 8291 trials in the GPe and 5034 trials in the DLPFC were recorded under the PCP effect; and 1788 trials in the GPe and 1858 trials in the DLPFC were recorded after PCP withdrawal. **e** DLPFC (top) and GPe (bottom) mean ± SEM FR around choice selection (time zero). **f** Left: Mean ± SEM activity in DLPFC (top) and GPe (bottom) during the 2 s preceding choice selection (shaded gray in e) following unsuccessful trials, across naïve (black), PCP (red), and post-PCP (blue) conditions. Right: Same, for successful trials. *p*-values from Bonferroni-corrected two-sample t tests compare each condition to naïve. Trial counts for unsuccessful trials: naïve—1373 DLPFC/2,963 GPe (325/233 neurons); PCP—1749 DLPFC/3291 GPe (178/149 neurons); post-PCP—163 DLPFC/192 GPe (45/33 neurons). For successful trials: naïve—10,662 DLPFC/14,364 GPe; PCP—4804 DLPFC/8355 GPe; post-PCP—1464 DLPFC/1251 GPe. Each bar chart is overlaid with 100 randomly selected data points falling within one standard deviation of the mean. For the full distribution of data points, please see Supplementary Fig. 3.

exploration and reducing the likelihood of making random exploration. Neuronally, the causal relationship between the two areas returned to its naïve pattern. The DLPFC regained its correlation with the learning slope, while the GPe showed a substantial increase in activity, strongly correlating with switch probability but not with reward omissions. These findings, shown separately for each NHP (Supplementary Fig. 9), led us to hypothesize that GPe activity is imperative for maintaining balanced exploratory strategies by modulating attentional control over access to WM in the DLPFC[23]. Due to these findings and since the GPe has a low density of NMDA-R[51], we propose that the increase in GPe activity under PCP administration was intensified by DLPFC dysfunction and impaired WM and predictive ability (e.g., compensatory mechanism). The subsequent rise in GPe activity post-PCP and the restoration of predictive skills (WM) led to improved behavioral outcomes.

Computational modeling

We initially employed a basic reinforcement learning (RL) model in which expected state-action values $V(s,a)$ are updated based on reinforcement history for each stimulus-action pair. However, even at the highest learning rate constant ($\alpha = 1$), the model failed to learn fast enough (Supplementary Fig. 10a). To overcome this limitation, we used features attributed to WM in the following models. First, we used a full state-action update (FSA) model. This model updates all possible stimulus-action pairs, unlike the basic RL model, which only updates the chosen pair. Here, WM encodes observed events, inferring values for unchosen stimuli based on the selected choice. This memory retention immediately influences behavior, accelerating learning and aligning with the NHPs' behavioral results (Supplementary Fig. 10a top). A different feature of WM that may be beneficial for learning speed is forgetfulness. We, therefore, implemented a decay in the model (ϕ) so that after the RL update, all state-action values are degraded at the beginning of the following trial[52,53]. Adding a forgetting process (without a full state-action update) to the basic model improved learning speed, replicating the NHPs' behavioral results (Supplementary Fig. 10a bottom).

Adaptive learning models - Next, we aimed to model the high correlation observed between GPe activity and switch probability. We hypothesized that a switch should occur when the surprise is high or when knowledge is scarce. Therefore, we set the learning rate parameter α_t to represent the model's surprise level and examined its relationship with recorded GPe activity (See the methods section for

the complete mathematical derivation):

$$\alpha_t = C \cdot \left(\frac{1 - P(r_t|s,a)}{1 + P(r_t|s,a)}\right)^{\sigma} \quad (1)$$

where C is a constant, $C \in (0,1]$ allowing to shift α_t baseline level and σ is the SD of choice probabilities. Thus, when knowledge about the correct cue is low or when the level of surprise is high, α_t increases.

Adaptive forgetful and full state-action update models - These models are similar to the basic forgetful and full state-action update (FSA) models, with one key difference: the step-size parameter α_t changes dynamically based on the need for learning. Specifically, α_t adjusts in response to the model's "surprise" and knowledge of the current task state. This learning rate controls how much stimulus values are updated and hence the pace of learning. Higher α_t values correspond to faster, more vigorous learning, whereas lower values correspond to slower or even minimal learning. The state-action update role for the adaptive forgetful model is therefore

$$V(s_t, a_t) \leftarrow V(s_t, a_t) + \alpha_t \times (r_t - V(s_t, a_t)) \quad (2)$$

And for the adaptive FSA model is:

$$\begin{aligned} V(s_t^{\text{chosen}}, a_t) &\leftarrow V(s_t^{\text{chosen}}, a_t) + \alpha_t \times (r_t - V(s_t^{\text{chosen}}, a_t)) \\ V(s_t^{\text{unchosen}}, a_t) &\leftarrow V(s_t^{\text{unchosen}}, a_t) + \alpha_t (1 - R_t - V(s_t^{\text{unchosen}}, a_t)) \end{aligned} \quad (3)$$

In this framework, α_t acts as a modulator of attention, effectively combining elements of Mackintosh and Pearce-Hall's attention theories[54–56]. Attention is allocated based on the understanding that spatial cues reliably predict reward outcomes, aligning with Mackintosh's theory, which assigns greater attention to well-established predictors. Conversely, when knowledge about the correct cue is low or when a surprising outcome occurs, attention increases, as proposed by Pearce-Hall's theory, directing focus toward stimuli with uncertain or unpredictable associations. These models are biologically plausible, as they conserve the agent's energy and minimize the risk of learning insignificant or erroneous information.

Reinforcement learning + working memory, combined model—this model integrates basic RL and WM models. To capture the transient nature of WM, we simulated it using a forgetful RL model with a learning rate of one[52]. Here, an observed event, when retained, can immediately and profoundly influence behavior but remains accessible only for a relatively short duration[52,53]. Since the set size is small (three stimuli) and does not change, we assume no capacity limitation. The action selection of the RL and WM models was determined using the SoftMax function $p_{wm}(a) = \text{softmax}(T_{wm}, V_{wm}), p_{RL}(a) = \text{softmax}(T_{RL}, V_{RL})$. The probability that the RL or the WM component governs action selection is fixed by the surprise measure α_t.

$$p(a) = \alpha_t \times p_{\text{WM}} + (1 - \alpha_t) \times p_{\text{RL}} \quad (4)$$

Thus, in situations requiring cognitive flexibility—when knowledge is limited or surprise is high—the WM component will dominate the action selection process. Conversely, when surprise is low and sufficient knowledge has been accumulated, the RL component will take precedence in determining actions.

First, we tested whether α_t acts similarly to the GPe recordings in each of the models. We, therefore, 'forced' the models to replicate the NHPs' behavioral choices and calculated the resultant α_t values. We found that α_t values highly correlate with the NHPs' GPe discharge rate in all models (Fig. 6a, c top, Supplementary Fig. 10c). Next, we assessed whether the models could reproduce the observed neuro-behavioral outcomes when faced with the behavioral task (i.e., the models were now free to make their own choices to solve the task). Indeed, all three models showed a high degree of correlation with the NHPs' behavioral data (Fig. 6a, c, middle, Supplementary Fig. 10c). Moreover, the models' α_t value was highly correlated with the learning slope (Supplementary Fig. 10b) and switch probability (Fig. 6a, b, bottom), as was the NHPs' GPe activity (Fig. 3b). This suggests that the models' α_t value may play a similar role in learning and decision-making as the NHPs' GPe activity.

Subsequently, we investigated the three adaptive models' capacity to explain the neuro-behavioral changes observed following PCP administration. For this purpose, we increased the pseudo-temperature T value in the adaptive FSA model and the forgetfulness ϕ value in the adaptive forgetful and combined models and examined the changes in the models' behavior and α_t value throughout the task (Fig. 6b, d left columns and Supplementary Fig. 10d). This single parameter change caused all three models to learn slower (Fig. 6b, d left columns top) and increased random exploration probability together with elevated α_t values (Fig. 6b, d left columns bottom) following successful trials (mirroring the neurobehavioral results under PCP). Conversely, directed exploration probability decreased in all models, while α_t values following an unsuccessful trial did not change monotonically (Fig. 6b, d, middle, Supplementary Fig. 10d). These results again mirror the NHPs' neurobehavioral results, where the directed exploration decreased under PCP, but GPe activity did not significantly change (Fig. 5c, f).

Finally, we investigated the models' capacity to explain the neurobehavioral changes observed following the cessation of PCP administration. We restored T and ϕ to their original values and systemically increased α_t modulator value, C, thus increasing its 'baseline' level linearly (simulating the increase in GPe discharge rate). Consistent with the NHPs' post-PCP results, all models showed that an increase in the basal level of α_t enhances learning speed (Fig. 6b, d, top right columns) and increases the likelihood of directed exploration, accompanied by a rise in α_t after unsuccessful trials (Fig. 6b, middle right columns), mirroring the neurobehavioral findings (Fig. 5a–c, f). On the other hand, random exploration probability decreased, accompanied by an increase in α_t value in the FSA and forgetful models (Fig. 6b bottom right columns, Supplementary Fig. 10d), mimicking the NHPs' results (Fig. 5c, f), but increased in the combined model (Fig. 6b bottom right column).

All three models show that α_t basal level positively correlates with learning speed and directed exploration probability. Two of the three further show that intensifying α_t is accompanied by a decrease in random exploration probability, while the combined model predicts the opposite result. Notably, the adaptive models indicate that enhancing GPe activity, e.g., by low-frequency electrical stimulation of the GPe, may confer benefits for task performance in our NHPs with PCP-induced cognitive schizophrenia-like symptoms.

GPe stimulation restores PCP effected directed and random exploration imbalance

To test the prediction of the adaptive models and investigate the causal relationship between GPe activity and behavioral performance under the influence of PCP, we conducted macro-electrode stimulation of the GPe during the NHPs' behavioral task performance. We administered either a high-frequency (130 Hz, neuronal activity dampener[57–59]) or low-frequency (13 Hz, activity enhancer[60–62]) long-duration stimulation to the GPe of the NHPs under PCP influence. In line with the prediction of the adaptive models, the frequency of stimulation had a significant impact on task performance. High-frequency stimulation leading to poorer performance and low-frequency stimulation resulting in improvement (Fig. 7a). Decreasing GPe activity by long-duration 130 Hz stimulation reduced cognitive flexibility and the probability of directed exploration, while increasing GPe activity by low-frequency stimulation enhanced it (Fig. 7b middle). Furthermore, the NHPs' learning dynamics improved when GPe

https://doi.org/10.1038/s41467-025-60044-5

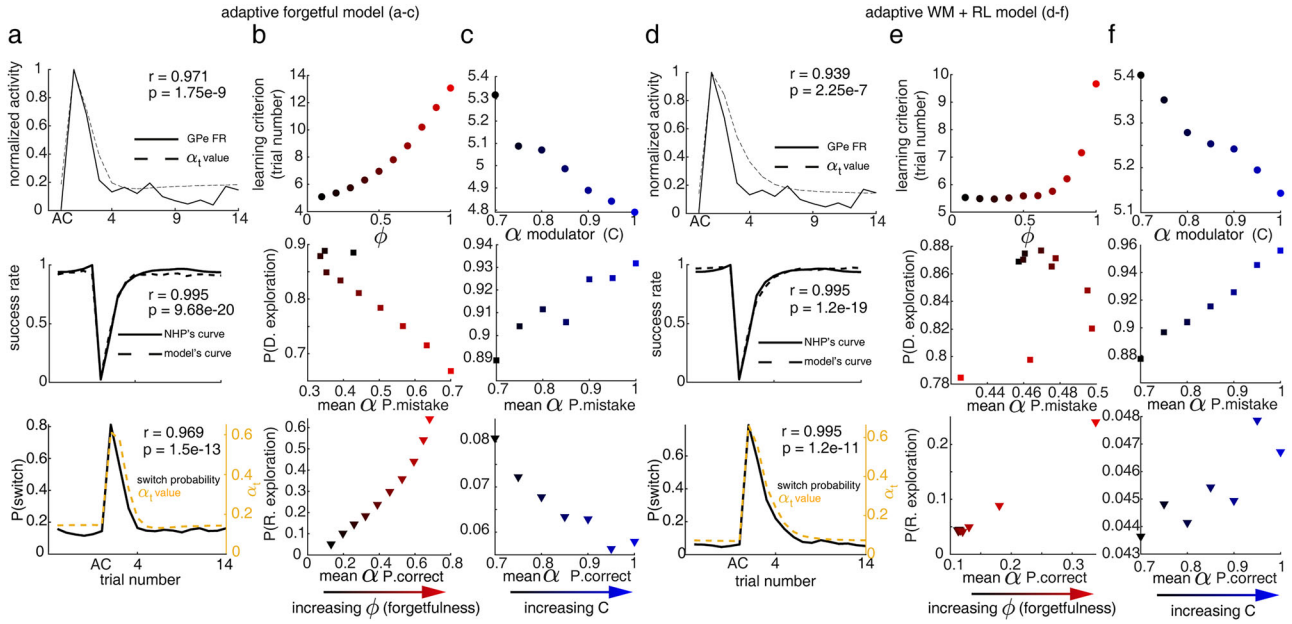


Fig. 6 | Adaptive reinforcement learning models replicate neuro-behavioral results and predict potential benefits for GPe stimulation under PCP administration. a, **b** Adaptive forgetful model. **a** Top—Comparing the non-human primates' (NHPs) normalized recorded activity of the external segment of the globus pallidus (GPe) firing rate (FR, solid line) with the normalized surprise measure α_t (dashed line) calculated based on the NHPs' choices. r and p indicate the correlation coefficient and *p*-values using Pearson's correlation. Middle—Comparing the models' and the NHPs' learning curves. Bottom - Comparing α_t value with the models' switch probability. **b** Simulating the phencyclidine (PCP) state by increasing the model's forgetfulness (ϕ). Simulated parameters color graded from black to red ($\phi = 0.1 - 1$). Top, learning criterion. Middle—α_t value after an unsuccessful trial compared with the probability for directed exploration. Bottom—α_t value after a successful trial compared with the probability for random exploration. **c** The same as (**b**), this time increasing the value of *C*, the α_t modulating parameter. **d**–**f** The same as (**a**–**c**), here showing the results of the adaptive combined (WM + RL) model. WM working memory, RL reinforcement learning.

activity was enhanced by 13 Hz stimulation, leading to faster learning (Fig. 7a). In line with the FSA update and forgetful models, decreasing GPe activity by 130 Hz stimulation increased random exploration, whereas increasing GPe activity by 13 Hz stimulation decreased random exploration (Fig. 7b).

The behavioral effects of the stimulation, shown separately on each NHP (Supplementary Fig. 11), highlight the modulatory role of GPe activity in cognitive flexibility and E-E balance. As GPe activity intensified, cognitive flexibility improved, and exploration patterns became more accurate, leading to an increased probability of initiating directed exploration and a decreased likelihood of performing random exploration. Conversely, reducing GPe activity resulted in lower cognitive flexibility and less precise exploration patterns. This aligns with the classical rate models of the basal ganglia that predict a reduction in the GPe discharge rate following Parkinson's dopamine depletion[63]. Lower cognitive flexibility is indeed a common symptom in all stages of Parkinson's disease. Thus, our stimulation studies with bidirectional effects on GPe activity and the resulting opposite effect on directed and random exploration probabilities suggest that GPe serves a critical function in controlling cognitive flexibility and E-E balance[33].

Discussion

Exploration is an action aimed at learning through experience of unfamiliar modalities. It can be directed by information seeking demanded by environmental changes or randomly generated by internal decision noise. Cognitive flexibility and exploration strategies are intertwined and imperative for learning, and both require well-grounded WM. The BG and the DLPFC are believed to have significant roles in both cognitive flexibility[9–11], WM[23–25,64] and exploration strategies[17–21,65]. The DLPFC, the center of executive functions, and the GPe, through GPe-subthalamic and other BG main-axis loops, have been suggested as the primary hubs controlling the E-E balance[17–21,65]. Additionally, GPe may dynamically regulate WM, stored in cortical areas, according to the changing environment and the agent's needs[23–25]. This gating is probably maintained by reducing the dimensionality of the information sent from the whole cortex to the BG and back to the frontal cortex[18,23]. Our results support the attention-gating hypothesis[23–25], suggesting that the GPe plays a role beyond being a switch for exploration. The GPe may help focus attention on relevant information and suppress irrelevant information, which is essential for accurate exploratory strategies.

Subjects with schizophrenia have dysfunctional cortical and GPe activities[29–32]. They also manifest decreased cognitive flexibility and diminished drive for directed exploration but excessive drive for random exploration[27,28]. Recently, the E-E balance has been suggested to be a holistic and ecologically valid framework to resolve some of the apparent paradoxes that have emerged within schizophrenia research[28]. A reduction in directed exploration has been linked to greater anhedonia[66], as patients are less likely to discover that alternative actions could lead to more rewarding outcomes, reinforcing their diminished motivation and pleasure. Conversely, an increase in random exploration has been associated with positive symptoms, particularly delusional thinking[28]. Excessive exploration may lead to over-association between unrelated stimuli, contributing to the formation and persistence of delusions.

To better understand the neural underpinnings of these processes in the healthy state and under PCP effect causing a similar shift in the E-E balance, we used a deterministic three-armed bandit task. We intentionally chose a deterministic task instead of a probabilistic one to better differentiate the two exploration types. In this setting, post-unsuccessful exploration is aimed at information seeking, and post-successful exploration is most likely generated by internal noise (be it an unintentional mistake, forgetfulness, or other). Here, we show that increased GPe activity enhances cognitive flexibility and improves exploratory strategies (both directed and random).

https://doi.org/10.1038/s41467-025-60044-5

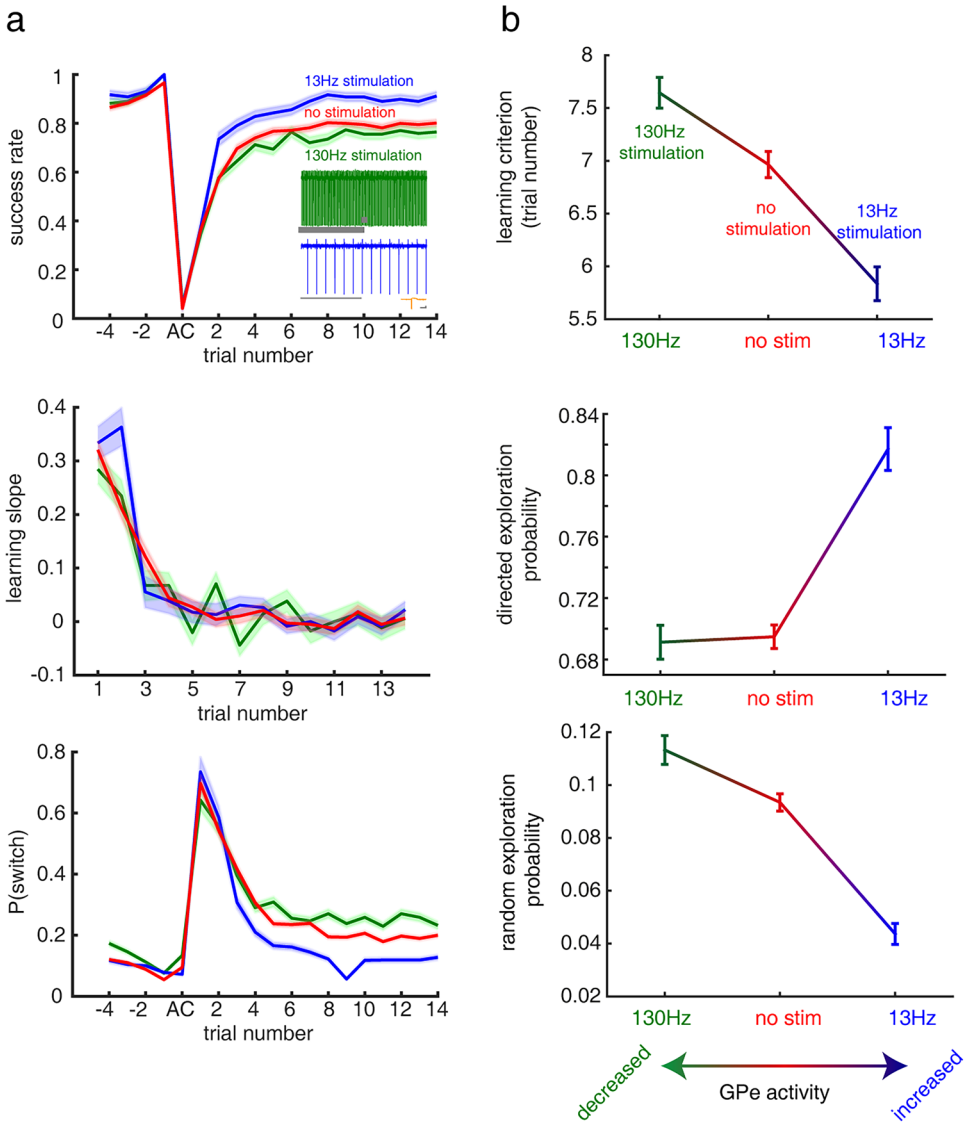


Fig. 7 | GPe low-frequency macro stimulation improves task performance under PCP administration, whereas high-frequency stimulation hampers it. Behavioral performance analysis was carried out under phencyclidine (PCP) administration. Red - No stimulation, blue—130 Hz continuous stimulation (activity dampener), and green—13 Hz continuous stimulation (activity enhancer). **a**Top—Mean ± SEM learning curve. Inset shows a one-second recording during 130 Hz stimulation (green) and during 13 Hz stimulation (blue), the gray bar shows 500 ms duration and a recording of one stimulation epoch (yellow). Middle—learning slope. Bottom—Switch probability. **b** Top—Mean ± SEM learning criterion (achieving three consecutive correct choices) during 130 Hz stimulation (left, blue, total of 343 blocks), no stimulation (middle, red, total of 770 blocks), and 13 Hz stimulation (right, green, total of 236 blocks). Middle—the non-human primates' (NHPs) Mean ± SEM probability of making directed exploration (i.e., to switch their choice after a prediction-outcome mismatch). A total of 1762 trials during 130 Hz stimulation, 3,611 trials without stimulation and 771 trials with 13 Hz stimulation. Bottom—The NHPs' Mean ± SEM likelihood of making random exploration (i.e., to switch their choice after congruent prediction-outcome). A total of 3345 trials under 130 Hz stimulation, 7939 trials without stimulation and 2681 with 13 Hz stimulation.

The central location of the GPe within the basal ganglia led us to explore two hypotheses using computational models. The RL + WM combined model posited that the GPe acts as a gatekeeper between the striatal-based RL habitual system and the goal-directed system of the DLPFC. In contrast, the adaptive forgetful and the FSA models treated the GPe as modulating attention, without distinguishing between these two learning systems. Both models replicated the results well with one key difference. While the combined model predicted a positively correlated increase in random exploration with α_t elevation, the forgetful model predicted a negatively correlated decrease. We found that increasing the activity of the GPe using low-frequency stimulation[60–62] decreased random exploration probability, thus, favoring the forgetful or the FSA models. Of note, the exact mechanism and effect of DBS are under open discussion.

Our findings underscore the critical role of GPe activity in regulating cognitive flexibility and the E-E balance. Consistent with the predictions of adaptive models, the effects of GPe stimulation were frequency-dependent: high-frequency (130 Hz) stimulation impaired task performance, reducing cognitive flexibility and directed exploration while increasing random exploration, whereas low-frequency (13 Hz) stimulation had the opposite effect, enhancing learning dynamics and promoting more accurate exploration patterns. However, a limitation of our study is that we could not examine DLPFC neural dynamics during macro-stimulation due to recording artifacts, preventing simultaneous assessment of its role in these effects.

Given that schizophrenia is characterized by reduced directed exploration, contributing to anhedonia[66], and increased random exploration, which is linked to delusional thinking[28], GPe stimulation could offer a potential therapeutic strategy. By low-frequency

stimulation GPe activity, directed exploration may be restored, helping patients better recognize rewarding alternatives, while excessive random exploration may be suppressed, reducing over-association between unrelated stimuli that contribute to delusions. These results suggest that targeted modulation of GPe activity could help rebalance the E-E trade-off in schizophrenia, improving cognitive flexibility and adaptive decision-making. In the future, it would be valuable to investigate GPe activity using an alternative schizophrenia model to further validate these findings.

Schizophrenia affects an estimated 21 million people worldwide (0.3–0.5% of adults)[67], with up to 34% being treatment-resistant[68] and having limited options after first-line therapies fail. This has driven research into potential DBS targets, including the nucleus accumbens, hippocampus, GPi, dorsomedial thalamus, and medial septal nucleus[69], based on DBS success in depression and obsessive-compulsive disorder. Here, we show that low-frequency GPe stimulation improves the E-E balance disrupted by PCP, aligning with studies suggesting the GPe-GPi lamina as an optimal DBS target for Parkinson's[70] and GPe stimulation for insomnia[71]. Future research should investigate theta activity-specific parameters as biomarkers, similar to adaptive DBS in Parkinson's[72] and obsessive-compulsive disorder[73]. Still, our findings support exploring GPe DBS for treatment-resistant schizophrenia, potentially as part of a multi-electrode strategy paired with target for positive symptoms (e.g., the substantia nigra or accumbens).

Materials and Methods

Animal training and behavioral tasks

Data were obtained from two female vervet monkeys (Cercopithecus aethiops, monkeys K and R) weighing 3.5–4 kg. All data were pooled for both monkeys for all analyses. Care and surgical procedures were in accordance with the National Research Council Guide for the Care and Use of Laboratory Animals[74] and the Hebrew University guidelines for the use and care of laboratory animals in research. The study was approved (MD-15-14412-5) and supervised by the institutional animal care and use committee of the Hebrew University and Hadassah Medical Center. The Hebrew University is an Association for Assessment and Accreditation of Laboratory Animal Care internationally accredited institute.

The behavioral paradigm used was a multiblock three-choice reversal learning task (Fig. 1b). The NHPs used their right (contralateral to the recording hemisphere) hands to touch stimuli presented on a screen that was located ~16 cm from their heads (Elo 1939L 19-inch open-frame touch-monitor; Elo Touch Solutions Limited). Three square fractal images were randomly selected as stimuli in each daily session out of 10 possible images. The stimuli were presented in fixed positions throughout the session, on the white screen's left, right, and center. Each trial began with a presentation of a black horizontal box on the lower right corner of the screen (Fig. 1b). To initiate a trial, the NHPs touched the black rectangle, which then disappeared. Two seconds later, the three fractal stimuli appeared. All three stimuli disappeared after touching (choosing) one of the three. Two seconds later, a red horizontal box appeared on the bottom of the screen. Touching this box was followed by three simultaneous results: The box would disappear, a banana-flavored liquid reward was either delivered or not (according to the NHPs' choice), and an ~80 ms auditory stimulus was played, obscuring any acoustic artifacts of the food pump. The sound played independently of the trial's outcome. Trials were aborted if no choice was made within 30 s or the red box was not touched within 30 s. All trials (correct, incorrect, and aborted) were followed by a variable intertrial interval (ITI) lasting 5–8 s. For each block, only one of the three stimuli was deterministically rewarded. No reward was delivered for the choice of other stimuli. The NHPs had no guiding information pointing them toward the correct stimulus. Still, eventually, learning was established, and the probability of choosing the correct stimulus reached a plateau (Fig. 1c). The criterion for learning was reached once the monkey chose the rewarded stimulus for 12–15 trials out of the last 25 (the criterion was randomly selected per block). Once this happened, an un-cued switch in the reward stimulus's identity occurred (i.e., reversal), and a new block started. Each daily session ended once the monkeys no longer initiated trials. After long periods in which the monkeys did not work, the experimenter occasionally delivered a free reward to re-motivate them. The behavioral paradigm was designed and run using the Psychophysics toolbox (Brainard et al., 1997) for MATLAB (The MathWorks, Inc.). Monkeys were trained for 5–6 days per week and were allowed free access to water in their home cages. Supplementary food was delivered when the monkeys did not reach the predefined daily calorie minimum. NHPs were given free access to food on the weekends.

Surgery and MRI

The NHPs were fully trained on the task (4–5 months) before the recording chamber was implanted. After the training period, they were operated on under full anesthesia and sterile conditions. In the surgery, an MRI-compatible Cilux head holder (Crist Instrument) and a square Cilux recording chamber (AlphaOmega) with a 27 mm (inner) side, located above a burr hole in the skull, were attached to the heads of the monkeys. The recording chamber was attached to the skull, tilted ~45° laterally in the coronal plane, with its center targeted at the stereotaxic coordinates of the left GPe (Fig. 1a). All surgical procedures were performed under aseptic conditions and general isoflurane and N2O deep anesthesia. Analgesia and antibiotics were administered during surgery and continued postoperatively. Recording began after a postoperative recovery period of several days. We estimated the stereotaxic coordinates of the recording target using MRI scans. The MRI scan (General Electric 3 tesla system, T2 sequence) was performed under i.m. Domitor and ketamine moderate sedation. Upon experiment completion, all surgical attachments were removed from NHPs. The NHPs were then rehabilitated and placed at the Israeli Primate Sanctuary.

Recording and data acquisition

During recording sessions, the heads of the monkeys were immobilized, and simultaneously, up to eight glass-coated tungsten microelectrodes (impedance 0.15–1 MΩ at 1000 Hz), confined within a cylindrical guide (1.65-mm inner diameter) were advanced separately (EPS; Alpha- Omega Engineering) into the GPe and the DLPFC (four to the GPe and four to the DLPFC). The electrical activity was amplified with a gain of 20, then filtered using hardware Butterworth filters (high-passed at 0.075 Hz, two poles; low-passed at 10,000 Hz, three poles) and finally sampled at 44.6 kHz (SnR; Alpha-Omega Engineering). Cortical and GPe units were identified by stereotaxic and MRI coordinates, electrophysiological hallmarks of the encountered structures along the penetration, and unique characteristics, such as GPe neurons' high FRs and pausing behavior[75]. Neuronal activity was sorted and classified online using a template-matching algorithm (SnR; Alpha-Omega Engineering). Each entry position was registered according to the chamber's X-Y coordinates, and brain structures were identified electro-physiologically, creating a 3D map of the desired brain column beneath the chamber. Cells were selected for recording as a function of their isolation quality and optimal signal-to-noise ratio[76]. For our analysis, we included only GPe and DLPFC neurons with isolation scores exceeding 0.8 and 0.7, respectively. We recorded a total of 325 DLPFC and 233 GPe neurons in the naïve state, 178 DLPFC and 149 GPe neurons in the PCP state, and 45 DLPFC and 33 GPe neurons after PCP withdrawal.

Analysis of behavior

To evaluate the NHPs' exploratory behavior, we categorically divided exploration patterns into exploration-favorable (following a

prediction-outcome mismatch) and exploitation-favorable (without such mismatch) trials. The former are trials that followed an unsuccessful trial (i.e., post-mistake) in which exploration (i.e., switch of choice) will lead to reward, and the latter are trials that followed a successful trial (i.e., post-correct) in which exploitation (i.e., choosing the same stimulus as before) will lead to reward. Each category was further subdivided into two groups, resulting in four options. (1) Directed exploration—trials following an unsuccessful trial in which the NHPs switched their choice (i.e., explored). (2) Perseveration—trials following an unsuccessful trial in which the NHPs did not switch their choice. (3) Random exploration—trials following a successful trial in which the NHPs' switched their choice. (4) exploitation –trials following a successful trial in which the NHPs did not switch their choice.

We also evaluated the learning dynamics throughout the task, looking at five main parameters. The learning curve, defined as the NHPs' success rate (i.e., the probability of choosing successfully), the learning slope (the derivative to the learning curve), switch probability (i.e., the probability of choosing a different cue than the choice of the previous trial), which was further divided to successful switch probability (i.e., the probability to switch successfully—finding the correct cue) and unsuccessful switch probability (i.e., the probability of switching unsuccessfully). We further analyzed the NHPs' likelihood of success if they changed their choice (i.e., the probability of success given that a switch was made). This analysis provides more precise information about the NHPs' memory of the previous states (50% chance is a random choice).

Lastly, we evaluated the NHPs' response times. Response times were defined as the time it took the NHPs to 1) choose one of the three stimuli and 2) claim the reward outcome by pressing the red rectangle (i.e., reaction plus movement time). All data analysis was performed using MATLAB (The MathWorks, Inc.).

Analysis of single-unit firing activity

Data analysis was conducted only for units that were stably held, well-isolated [isolation score (24) ≥ 0.8 in the GPe and ≥ 0.7 in the DLPFC] and unquestionably identified as either cortical or GPe units. Only GPe units that fired ≥20 spikes per second on average were included (thus focusing on the prototypical high-frequency discharge population). Continuous traces showing GPe spiking activity (Fig. 1a) were obtained using a two-pole Butterworth filter (1000–3000 Hz). Resting neural activity was recorded during the last three seconds of the inter-trial interval (Supplementary Fig. 7). To calculate FR during the trial, we defined a two-second baseline period to measure changes from resting activity. This baseline was used to generate z-scored dynamic temporal patterns of GPe and DLPFC activity (PSTHs). We divided each second into ten nonoverlapping bins and calculated the spike count per bin. This data, per trial, was used as a baseline for the PSTHs. For each time point of the PSTH, the baseline discharge rate for that trial was subtracted from the count per bin, and the difference was divided by the baseline SD. We used the last two seconds of the ITI as the baseline for all epochs (except for neuronal activity in response to reward/reward omission). The baseline discharge rate for z-score calculation of neuronal activity in response to reward and reward omissions (Figs. 2a and 3a) was the two seconds preceding reward claiming (i.e., two seconds before pressing of the red box). The obtained values were smoothed in all cases using a 50ms-wide Gaussian Kernel (MATLAB's filtfilt function).

Chronic, low-dose PCP administration

A mini-osmotic pump (Alzet osmotic pump model 2ML4) was implanted subcutaneously between the scapulae of both NHPs, providing a continuous daily dose of 1.68 mg/kg for 28 days. Both NHPs were implanted twice with at least one month washout period in between. Post-PCP recordings started at least two weeks after pump removals.

Macro-stimulation

Using the 3D map constructed during our recording sessions, we chose a desired stimulation position, maximizing the GPe location. We then placed a microgrid attached to the chamber's inner wall. Using a specially designed guiding tower (AlphaOmega) we inserted a (recording) micro-electrode to the desired location, marked the GPe borders, and identified its center. Placing the concentric macro-electrode (microProbes CEAX 200) into the head mount allowed us to insert it into the desired location and stabilize it to the grid using dental cement. The macro-electrode remained connected and immobilized to the chamber. No sooner than three days after implantation, experimentation resumed. The NHPs were given either 13 Hz or 130 Hz macro-stimulation, on alternating days, during the behavioral task while we recorded their performance.

Data analysis

Statistical analyses were performed using MATLAB version R2022b and are described in the figure legends. The letter 'r' represents the correlation coefficient for correlation analyses. The corresponding probability value, P, is the probability of getting 'r' as large as the observed value by random chance when the true correlation is zero and was computed by transforming the correlation to create a t-statistic. When there were multiple comparisons, the Bonferroni correction was used (i.e., the obtained P-value was multiplied by the number of comparisons) before significance testing ($P < 0.05$) and is presented as such. All data presented met the assumptions of the statistical test used.

Computational modeling

Basic RL model - We initially employed a standard reinforcement learning (RL) model in which expected state-action values $V(s,a)$ are updated based on reinforcement history for each stimulus (s) and action (a). The value for the selected action given the stimulus is updated based on the observation of the trial's reward outcome, r_t, as a function of the prediction error between the expected and observed reward

$$V(s_t,a_t) \leftarrow V(s_t,a_t) + \alpha \times (r_t - V(s_t,a_t)) \tag{5}$$

In this model, the learning rate α is constant. Choices are generated probabilistically as a function of the difference in state-action values using the SoftMax probabilistic choice function:

$$p(\text{choose } s_i) = \frac{e^{\frac{V(s_i)}{T_{RL}}}}{\sum_{j=1}^{3} e^{\frac{V(s_j)}{T_{RL}}}} \tag{6}$$

Such that the 'temperature' T sets the level of noise in the decision process, with large T corresponding to high decision noise with nearly random decision and small T corresponding to low decision noise and near-deterministic (greedy) choice of the highest-value option. Full state-action (FSA) update model - Unlike the simple RL model, where learning is confined to the action taken and its immediate reward, we simulated WM as encoding an observed event, incorporating inferences about the unchosen stimuli based on the choice made (full state-action update). Thus, when retained in memory, it immediately and significantly influences behavior.

$$\begin{aligned} V_{WM}(s_t^{\text{chosen}},a_t) &\leftarrow V_{WM}(s_t^{\text{chosen}},a_t) + \alpha \times (r_t - V_{WM}(s_t^{\text{chosen}},a_t)) \\ V_{WM}(s_t^{\text{unchosen}},a_t) &\leftarrow V_{WM}(s_t^{\text{unchosen}},a_t) + \alpha(1 - R_t - V_{WM}(s_t^{\text{unchosen}},a_t)) \end{aligned} \tag{7}$$

Forgetful RL model - A different feature of WM that may be beneficial for learning speed is forgetfulness. We, therefore, implemented a decay in the model so that after the RL update, all state-

action values are degraded at the beginning of the following trial[52,53]. Here, state-action value learning and action selection are identical to the basic RL model with the addition of a forgetting process over time. At the beginning of every trial, the state-action values of all stimulus-action pairs are decayed towards their initial values.

$$V(s,a) = V(s,a) + \phi \times (V_0 - V(s,a)) \quad (8)$$

Where $V_0 = \frac{1}{n}$ is the initial state-action value for all pairs and n represents the number of presented stimuli to choose from. ϕ controls the degree of forgetfulness. Adding a forgetting process to the basic model improved learning speed, replicating the NHPs' behavioral results (Supplementary Fig. 10a).

Adaptive parameter α_t derivation

We defined $P = p(r_t|s,a)$ as the model assigned probability of receiving the reward value r_t given a choice (we assumed negligible noise in the process of perceiving reward). For example, assume that $p(r_t = 1|s_1, a_1) = 0.98$ and S_1 was chosen. If $r_t = 1$, meaning, a reward was given then $P = 0.98$. But if a reward was not given, then $P = 1 - p(r_t|s_1,a_1) = p(r_t|s_2,a_2) + p(r_t|s_3,a_3) = 0.02$.

We then use Shannon entropy to evaluate surprise, given a choice, and the perceived reward, $-\log(P)$. To normalize the surprise value to the range $[0,1]$ we use the following function $f_{\text{surprise}}(x)$:

$$f_{\text{surprise}}(x) = \frac{e^x - 1}{e^x + 1}$$

$$x = -\log(P)$$

$$f_{\text{surprise}}(-\log(P)) = \frac{e^{-\log(P)} - 1}{e^{-\log(P)} + 1} = \frac{\frac{1}{P} - 1}{\frac{1}{P} + 1} = \frac{1 - P}{1 + P}$$

$$f_{\text{surprise}}(P = 1) = 0 \le f_{\text{surprise}}(P) \le 1 = f_{\text{surprise}}(P = 0) \quad (9)$$

Therefore, as the mismatch between the prediction and the outcome is small ($P \to 1$) surprise is low, and $f_{surprise}$ value decreases. But when the mismatch between the prediction and the outcome is large ($P \to 0$) the surprise is high, and $f_{surprise}$ value increases.

To increase its value in situations of least knowledge (i.e., $V(S_1) \cong V(S_2) \cong V(S_3)$) we raised $f_{surprise}$ by the power of the standard deviation (SD) of the value of their corresponding probabilities, $std(p(r_t|s_1,a_1), p(r_t|s_2,a_2), p(r_t|s_3,a_3))$. Such that when the SD is low, the value will increase, and when it is high, it will not. Finally, receiving the following parameter α_t

$$\alpha_t = C \cdot \left(f_{surprise}(P)\right)^{\sigma} \quad (10)$$

Where C is a constant, $C \in (0,1]$ allowing to shift α_t base line level and σ is the SD of choice probabilities.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated in this study have been deposited in the figshare database, https://doi.org/10.6084/m9.figshare.27021373.

Code availability

The code for the reinforcement learning model is available on the Code Ocean platform, https://doi.org/10.24433/CO.3920364.v1.

References

1. Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the exploration-exploitation dilemma. *J. Exp. Psychol. Gen.* **143**, 2074–2081 (2014).
2. Bouchacourt, F., Tafazoli, S., Mattar, M. G., Buschman, T. J. & Daw, N. D. Interplay between rule learning and rule switching in a perceptual categorization task. *bioRxiv* https://doi.org/10.1101/2022.01.29.478330 (2022).
3. Wilson, R. C., Bonawitz, E., Costa, V. D. & Ebitz, R. B. Balancing exploration and exploitation with information and randomization. *Curr. Opin. Behav. Sci.* **38**, 49–56 (2021).
4. Gershman, S. J. Deconstructing the human algorithms for exploration. *Cognition* **173**, 34–42 (2018).
5. Sadeghiyeh, H. et al. Temporal discounting correlates with directed exploration but not with random exploration. *Sci. Rep.* **10**, 4020 (2020).
6. Barto, A. & Sutton R. S. *Reinforcement learning: an introduction*. The MIT Press (2018).
7. Balleine, B. W. & O'Doherty, J. P. Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* **35**, 48–69 (2010).
8. Rusu, S. I. & Pennartz, C. M. A. Learning, memory and consolidation mechanisms for behavioral control in hierarchically organized cortico-basal ganglia systems. *Hippocampus* **30**, 73–98 (2020).
9. Van Schouwenburg, M. R. et al. Cognitive flexibility depends on white matter microstructure of the basal ganglia. *Neuropsychologia* **53**, 171–177 (2014).
10. Vatansever, D., Manktelow, A. E., Sahakian, B. J., Menon, D. K. & Stamatakis, E. A. Cognitive flexibility: a default network and basal ganglia connectivity perspective. *Brain Connect.* **6**, 201–207 (2016).
11. Eslinger, P. J. & Grattan, L. M. Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia* **31**, 17–28 (1993).
12. Parent, A. & Hazrati, L. N. Functional anatomy of the basal ganglia. *Rev. Neurol.* **25**, S121–S128 (1997).
13. Haber, S. N. The primate basal ganglia: Parallel and integrative networks. *J. Chem. Neuroanat.* **26**, 317–330 (2003).
14. Groenewegen, H. J., Berendse, H. W., Wolters, J. & Lohman, A. The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization. *Prog Brain Res.* **85**, 95–116 (1990).
15. Hollerman, J. R., Tremblay, L. & Schultz, W. Involvement of basal ganglia and orbitofrontal cortex in goal-directed behavior. *Prog. Brain Res.* **126**, 193–215 (2000).
16. Redgrave, P. et al. Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nat. Rev. Neurosci.* **11**, 760–772 (2010).
17. Chersi, F., Mirolli, M., Pezzulo, G. & Baldassarre, G. A spiking neuron model of the cortico-basal ganglia circuits for goal-directed and habitual action learning. *Neural Netw.* **41**, 212–224 (2013).
18. Maith, O., Baladron, J., Einhäuser, W. & Hamker, F. H. Exploration behavior after reversals is predicted by STN-GPe synaptic plasticity in a basal ganglia model. *iScience* **26** (2023).
19. Suryanarayana, S. M., Hellgren Kotaleski, J., Grillner, S. & Gurney, K. N. Roles for globus pallidus externa revealed in a computational model of action selection in the basal ganglia. *Neural Netw.* **109**, 113–136 (2019).
20. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
21. Zajkowski, W., Kossut, M. & Wilson, R. C. A causal role for right frontopolar cortex in directed, but not random, exploration. In: *Proc. ICCM 2017—15th Int. Conf. on Cognitive Modeling* 79–84 https://doi.org/10.1101/127704 (2017).

22. Kojima, S., Kao, M. H., Doupe, A. J. & Brainard, M. S. The avian basal ganglia are a source of rapid behavioral variation that enables vocal motor exploration. *J. Neurosci.* **38**, 9635–9647 (2018).
23. Bar-Gad, I., Morris, G. & Bergman, H. Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog. Neurobiol.* **71**, 439–473 (2003).
24. Frank, M. J., Loughry, B. & Reilly, R. C. O. Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect Behav Neurosci.* **1**, 137–160 (2001).
25. McNab, F. & Klingberg, T. Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.* **11**, 103–107 (2008).
26. Schlagenhauf, F. et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* **89**, 171–180 (2014).
27. Cathomas, F. et al. Increased random exploration in schizophrenia is associated with inflammation. *npj Schizophr*. **7** (2021).
28. Speers, L. J. & Bilkey, D. K. Maladaptive explore/exploit trade-offs in schizophrenia. *Trends Neurosci.* **46**, 341–354 (2023).
29. Weinberger, D. R., Berman, K. F. & Zec, R. F. Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: i. regional cerebral blood flow evidence. *Arch. Gen. Psychiatry* **43**, 114–124 (1986).
30. Perlstein, W. M., Carter, C. S., Noll, D. C. & Cohen, J. D. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am. J. Psychiatry* **158**, 1105–1113 (2001).
31. Galeno, R., Molina, M., Guirao, M. & Isoardi, R. Severity of negative symptoms in schizophrenia correlated to hyperactivity of the left globus pallidus and the right claustrum. A PET Study. *World J. Biol. Psychiatry* **5**, 20–25 (2004).
32. Spinks, R. et al. Globus pallidus volume is related to symptom severity in neuroleptic naive patients with schizophrenia. *Schizophr. Res.* **73**, 229–233 (2005).
33. Mandali, A., Rengaswamy, M., Srinivasa Chakravarthy, V. & Moustafa, A. A. A spiking Basal Ganglia model of synchrony, exploration and decision making. *Front. Neurosci.* **9**, 1–21 (2015).
34. Li, F. & Tsien, J. Z. Memory and the NMDA receptors. *N. Engl. J. Med.* **361**, 302 (2009).
35. Cadinu, D. et al. NMDA receptor antagonist rodent models for cognition in schizophrenia and identification of novel drug treatments, an update. *Neuropharmacology* **142**, 41–62 (2018).
36. Blackman, R. K., Macdonald, A. W. & Chafee, M. V. Effects of ketamine on context-processing performance in monkeys: A new animal model of cognitive deficits in schizophrenia. *Neuropsychopharmacology* **38**, 2090–2100 (2013).
37. Bubeníková-Valešová, V., Horáček, J., Vrajová, M. & Höschl, C. Models of schizophrenia in humans and animals based on inhibition of NMDA receptors. *Neurosci. Biobehav. Rev.* **32**, 1014–1023 (2008).
38. Jentsch, J. D. & Roth, R. H. The neuropsychopharmacology of phencyclidine: From NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology* **20**, 201–225 (1999).
39. Uno, Y. & Coyle, J. T. Glutamate hypothesis in schizophrenia. *Psychiatry Clin. Neurosci.* **73**, 204–215 (2019).
40. Wise S. P. *Cortical Evolution in Primates: What Primates Are, What Primates Were, and Why the Cortex Changed* (Oxford University Press, 2024).
41. Friston, K., Moran, R. & Seth, A. K. Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* **23**, 172–178 (2013).
42. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci.* **2011**, 156869 (2011).
43. Elias, S. et al. Statistical properties of pauses of the high-frequency discharge neurons in the external segment of the globus pallidus. *J. Neurosci.* **27**, 2525–2538 (2007).
44. Slovik, M. et al. Ketamine induced converged synchronous gamma oscillations in the cortico-basal ganglia network of nonhuman primates. *J. Neurophysiol.* **118**, 917–931 (2017).
45. Mizrahi-Kliger, A. D., Kaplan, A., Israel, Z. & Bergman, H. Desynchronization of slow oscillations in the basal ganglia during natural sleep. *Proc. Natl Acad. Sci. USA* **115**, E4274–E4283 (2018).
46. N. R. G. et al. Basal ganglia components have distinct computational roles in decision-making dynamics under conflict and uncertainty. 1–24 https://doi.org/10.1371/journal.pbio.3002978 (2025).
47. Lisman, J. Working memory: The importance of theta and gamma oscillations. *Curr. Biol.* **20**, R490–R492 (2010).
48. Sederberg, P. B., Kahana, M. J., Howard, M. W., Donner, E. J. & Madsen, J. R. Theta and gamma oscillations during encoding predict subsequent recall. *J. Neurosci.* **23**, 10809–10814 (2003).
49. Lazarewicz, M. T. et al. Ketamine modulates theta and gamma oscillations. *J. Cogn. Neurosci.* **22**, 1452–1464 (2010).
50. Ursino, M. & Pirazzini, G. Theta–gamma coupling as a ubiquitous brain mechanism: implications for memory, attention, dreaming, imagination, and consciousness. *Curr. Opin. Behav. Sci.* **59**, 101433 (2024).
51. RAVENSCROFT, P. & BROTCHIE, J. NMDA receptors in the basal ganglia. *J. Anat.* **196**, 577–585 (2000).
52. Collins, A. G. E. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).
53. Rmus, M. et al. Age-related differences in prefrontal glutamate are associated with increased working memory decay that gives the appearance of learning deficits. *Elife* **12**, e85243 (2023).
54. Mackintosh, N. J. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol. Rev.* **82**, 276–298 (1975).
55. Pearce, J. M. & Hall, G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
56. Pearce, J. M. & Mackintosh, N. J. in *Attention and Associative Learning: from Brain to Behaviour* (eds Mitchell, C. J. & Le Pelley, M. E.) 11–39 (Oxford University Press, 2010).
57. Dostrovsky, J. O. & Lozano, A. M. Mechanisms of deep brain stimulation. *Mov. Disord.* **17** (2002).
58. Benabid, A. L., Pollak, P., Louveau, A., Henry, S. & De Rougemont, J. Combined (thalamotomy and stimulation) stereotactic surgery of the vim thalamic nucleus for bilateral parkinson disease. *Stereotact. Funct. Neurosurg.* **50**, 344–346 (1987).
59. Benabid, A. L. et al. Chronic electrical stimulation of the ventralis intermedius nucleus of the thalamus and of other nuclei as a treatment for Parkinson's disease. *Tech. Neurosurg.* **5**, 5–30 (1999).
60. Montgomery, E. B. & Baker, K. B. Mechanisms of deep brain stimulation and future technical developments. *Neurol. Res.* **22**, 259–266 (2000).
61. Khanna, P. et al. Low-frequency stimulation enhances ensemble cofiring and dexterity after stroke. *Cell* **184**, 912–930.e20 (2021).
62. Oza, C. S., Brocker, D. T., Behrend, C. E. & Grill, W. M. Patterned low-frequency deep brain stimulation induces motor deficits and modulates cortex-basal ganglia neural activity in healthy rats. *J. Neurophysiol.* **120**, 2410–2422 (2018).
63. Carpenter, Malcolm B. & Jayaraman, A. eds. in *The Basal Ganglia II: Structure and Function—Current Concepts* (Springer US, 2013).
64. Curtis, C. E. & D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* **7**, 415–423 (2003).
65. Kalva, S. K., Rengaswamy, M., Chakravarthy, V. S. & Gupte, N. On the neural substrates for exploratory dynamics in basal ganglia: A model. *Neural Netw.* **32**, 65–73 (2012).
66. Strauss, G. P. et al. Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects

of negative symptoms in schizophrenia. *Biol. Psychiatry* **69**, 424–431 (2011).
67. Charlson, F. J. et al. Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophr. Bull.* **44**, 1195–1203 (2018).
68. Potkin, S. G. et al. The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. *npj Schizophr.* **6** (2020).
69. Nucifora, F. C., Woznica, E., Lee, B. J., Cascella, N. & Sawa, A. Treatment resistant schizophrenia: Clinical, biological, and therapeutic perspectives. *Neurobiol. Dis.* **131**, 104257 (2019).
70. Holland, M. T. et al. Identifying the therapeutic zone in globus pallidus deep brain stimulation for Parkinson's disease. *J. Neurosurg.* **138**, 329–336 (2023).
71. Castillo, P. R. et al. Globus Pallidus Externus Deep Brain Stimulation Treats Insomnia in a Patient With Parkinson Disease. *Mayo Clin. Proc.* **95**, 419–422 (2020).
72. Asch, N. et al. Independently together: subthalamic theta and beta opposite roles in predicting Parkinson's tremor. *Brain Commun.* 1–13 https://doi.org/10.1093/braincomms/fcaa074 (2020).
73. Rappel, P. et al. Subthalamic theta activity: A novel human subcortical biomarker for obsessive compulsive disorder. *Transl. Psychiatry* **8** (2018).
74. National Research Council, Division on Earth, Life Studies, Institute for Laboratory Animal Research, Committee for the Update of the Guide for the Care, and Use of Laboratory Animals. "Guide for the care and use of laboratory animals." (2010).
75. DeLong, M. R. Activity of pallidal neurons during movement. *J. Neurophysiol.* **34**, 414–427 (1971).
76. Joshua, M., Elias, S., Levine, O. & Bergman, H. Quantifying the isolation quality of extracellularly recorded action potentials. *J. Neurosci. Methods* **163**, 267–282 (2007).

Acknowledgements

The authors thank Carmel R. Auerbach Asch for the fruitful discussions and feedback on our manuscript, and all members of the Bergman lab for helpful discussions. We gratefully acknowledge the following funding sources: The ISF Breakthrough Research program (Grant NO: 1738/22) to H.B. and the Collaborative Research Center TRR295, Germany (Project number 424778381) to H.B.

Author contributions

N.A. and H.B. conceived the research, designed the experiments, and wrote the manuscript. N.A. performed the in vivo experiments (electrophysiological and behavioral recordings), the surgical procedure, data analysis, statistics, and mathematical modeling. N.R. and U.W.R. performed the in vivo experiments. A.M. performed STA analysis. R.P. supports the data analysis and the mathematical modeling. ZI performed the surgical procedure. H.B. supervised the work.

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/s41467-025-60044-5.

Correspondence and requests for materials should be addressed to Nir Asch.

Peer review information *Nature Communications* thanks Charles Mikell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.