



HAL
open science

Optimal level of human intracranial theta activity for behavioral switching in the subthalamo-medio-prefrontal circuit

Maëva Laquitaine, Mircea Polosan, Philippe Kahane, Stephan Chabardes, Jérôme Yelnik, Sara Fernandez-Vidal, Philippe Domenech, Julien Bastin

► To cite this version:

Maëva Laquitaine, Mircea Polosan, Philippe Kahane, Stephan Chabardes, Jérôme Yelnik, et al.. Optimal level of human intracranial theta activity for behavioral switching in the subthalamo-medio-prefrontal circuit. *Nature Communications*, 2024, 15 (1), pp.7827. 10.1038/s41467-024-52290-w . hal-04697212

HAL Id: hal-04697212

<https://hal.science/hal-04697212v1>

Submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Optimal level of human intracranial theta activity for behavioral switching in the subthalamo-medio-prefrontal circuit

Received: 30 August 2023

Accepted: 29 August 2024

Published online: 07 September 2024

 Check for updates

Maëva Laquitaine¹, Mircea Polosan², Philippe Kahane², Stephan Chabardes², Jérôme Yelnik³, Sara Fernandez-Vidal³, Philippe Domenech^{3,4,5} ✉ & Julien Bastin^{1,5} ✉

The ability to switch between rules associating stimuli and responses depend on a circuit including the dorsomedial prefrontal cortex (dmPFC) and the subthalamic nucleus (STN). However, the precise neural implementations of switching remain unclear. To address this issue, we recorded local field potentials from the STN and from the dmPFC of neuropsychiatric patients during behavioral switching. Drift-diffusion modeling revealed that switching is associated with a shift in the starting point of evidence accumulation. Theta activity increases in dmPFC and STN during successful switch trials, while temporally delayed and excessive levels of theta lead to premature switch errors. This seemingly opposing impact of increased theta in successful and unsuccessful switching is explained by a negative correlation between theta activity and the starting point. Together, these results shed a new light on the neural mechanisms underlying the rapid reconfiguration of stimulus-response associations, revealing a Goldilocks' effect of theta activity on switching behavior.

Behavioral switching allows us to rapidly adapt to changing demands and depends on a set of high-level cognitive functions from the anticipated application of abstract rules and the inhibition of no longer relevant choices to the selection and execution of the most adaptive response^{1,2}. However, the neuro-computational mechanisms underlying our ability to rapidly adjust our behavior in response to unpredictable and sudden environmental changes remains poorly understood.

Early evidence suggests that the hyperdirect pathway connecting the dorso-medial prefrontal cortex (dmPFC) to the subthalamic nucleus (STN) implements behavioral switching, action stopping or response conflict^{3–6}. Yet, previous studies have mainly focused on action stopping or speed-accuracy trade-off adjustments to “conflict”

(also referred to as choice difficulty or choice uncertainty)^{7–9}. Previous human intracranial studies showed that human dmPFC and STN neurons' firing rate, as well as theta (~5–10 Hz) prefronto-subthalamic activity, increase when the best response is more difficult to select and require longer decision time^{10–14}. Conversely, action stopping increases both neurons firing rate and beta-band activity (~15–30 Hz) in the same dmPFC-STN circuit^{15–22}. By contrast, non-invasive electrophysiological studies of task switching yield mixed findings in terms of temporal dynamics, frequency regime and putative brain circuits^{23–25}, highlighting the need for more direct human electrophysiological recordings to establish a precise mapping between prefronto-subthalamic neural dynamics and key cognitive processes underlying behavioral switching.

¹Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, GIN, F-38000 Grenoble, France. ²Univ. Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France. ³Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin center, F-91191 Gif/Yvette, France. ⁴Institut de Neuromodulation, Pole Hospitalo-Universitaire 15, Groupe Hospitalo-Universitaire Paris, Psychiatrie et Neurosciences, Université Paris Cité, Paris, France. ⁵These authors contributed equally: Philippe Domenech, Julien Bastin. ✉e-mail: philippe.domenech@inserm.fr; julien.bastin@inserm.fr

Drift diffusion models posit that response selection proceeds by accumulating evidence up to a decision threshold in a winner take all fashion. This class of model was also shown to be relevant in the context of rule switching to tease apart the behavioral variability arising from switching between set of rules and the selection of a new response, which seems to occur sequentially^{26,27}. Conversely, studies combining direct electrophysiological recordings with a drift diffusion modeling of perceptual choices revealed that decision-threshold adjustments to difficulty were associated with STN-dmPFC neural activity either in theta or beta bands, depending on the cognitive context^{8,28}. However, to the best of our knowledge, such a neuro-computational dissection has never been performed using invasive human electrophysiological recordings in the dmPFC-STN circuitry during behavioral switching. Moreover, the use of different experimental tasks in human neuroimaging and monkey single-cell literatures has made it difficult to compare the cognitive and neural processes at play during behavioral switch across techniques and species (but see ref. 29 for an exception).

Thus, critical questions about how neural activity relates to behavioral switching remain unaddressed: Do nodes in the dmPFC-STN hyperdirect circuit exhibit specific temporo-frequency signatures associated with switching between behaviors? Does neural activity within these brain regions relate to trial-by-trial fluctuations in behavioral performance, suggesting a close link between successful neural computations and participants' ability to implement behavioral switching? What are the neuro-computational principles underlying these brain-behavior associations?

To address these questions, we used intracerebral recordings collected in two groups of rare neuropsychiatric patients implanted with depth electrodes, which allowed us to record from the dmPFC and the STN (respectively), while these patients performed a behavioral switching paradigm previously used in monkeys^{3,4} to record multiunit activity from both the dmPFC and the STN. We show that behavioral switching depends on theta activity occurring in both dmPFC and STN. Critically, trial-by-trial fluctuations of theta activity negatively correlated with behavioral switching performance in both brain structures, which translated in a neural HDDM accounting for

observed RTs and performances as a modulation of the initial level of evidence by theta activity, prior to response selection, during switch trials. Our findings demonstrate that although theta activity in the dmPFC and in the STN increases on average during switch trials, too high an increase might paradoxically induce premature incorrect switch responses that are detrimental to adaptation.

Results

We administered a reactive behavioral switching paradigm to two groups of post-surgery neuropsychiatric patients while intracranial electroencephalographic activity (iEEG) was recorded, and to one group of healthy controls. The first group consisted of three patients with drug-resistant partial epilepsy implanted with linear depth macroelectrodes in the dorso-medial prefrontal cortex (dmPFC). The second group consisted of four patients with severe and drug-resistant OCD implanted with deep brain stimulation macroelectrodes in the subthalamic nucleus (STN). The third group consisted of ten healthy controls (from whom we collected only behavioral recordings). We report demographic and clinical data in Table S1, 2.

Each trial started with two colored squares (yellow/pink) randomly displayed on each side of a central white square (Fig. 1). After 500 ms, the white central square turned either pink or yellow (ongoing rule) to prompt the patient to press a button indicating on which side of the screen the cue matching the central square color was displayed (left/right bimanual response). This stimulus-response association rule pseudo-randomly changed every 2 to 6 trials. Overall, the task consisted of two to four sessions composed of 50 switch trials each (138 ± 49 switch trials per subject).

Patients were instructed to select the correct response as quickly and as accurately as they could. Hence, the task required to form a preliminary response by preemptively applying the previous rule as soon as the current stimulus pair was displayed. In switch trials (20% of trials), patients had to override this preliminary response to instantiate and apply the new rule, resulting in the selection of the alternative action. To promote speeded decisions, in line with the original studies in monkeys^{3,4}, we implemented a response time window for non-switch trials (excluding switch trials, which typically have longer durations).

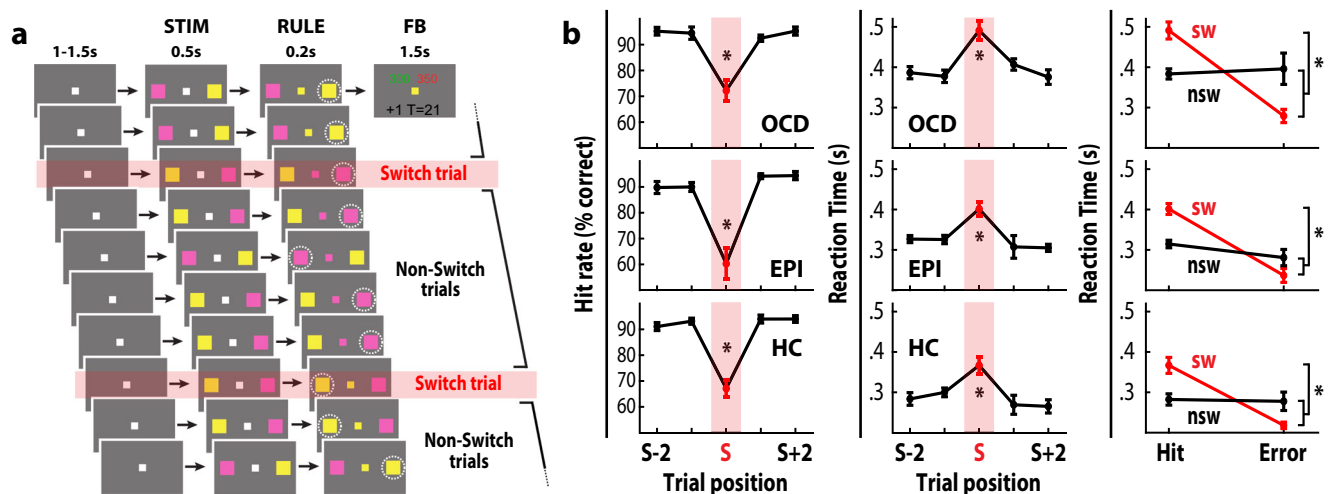


Fig. 1 | Behavioral task and results. a Task-switching paradigm. Patients had to indicate on which side the stimulus (STIM) matching the central square color (RULE) was. This response was followed by a feedback cue (FB) indicating whether it was correct and/or whether it fell within the allowed response time window. **b** Behavioral performances of patients with OCD (upper row; $n = 4$), epilepsy (EPI; middle row; $n = 3$) and of healthy controls (HC; bottom row; $n = 10$). Evolution of hit rates and reaction times averaged according to trial relative position to switch trials (left and central panels) or as a function of trial type (red: switch; black: non-switch) and

accuracy (hit or error). Error-bars indicate SEM between patients. Two-tailed paired t -tests showed that patients (OCD and EPI; $n = 7$) and healthy controls made more errors on switch trials (patients: $p = 1.09 \times 10^{-4}$; HC: $p = 5.73 \times 10^{-5}$), were longer during correct switch trials ($p = 9.66 \times 10^{-5}$; HC: $p = 7.61 \times 10^{-5}$) and their reaction times were also significantly faster during incorrect switch trials relative to incorrect non-switch trials ($p = 2.02 \times 10^{-2}$; HC: 2.29×10^{-2}). * $p < 0.05$. See also Table S3 for statistics across trials of each of the seven patients. S: switch (sw) trial; S-2 and S+2: non-switch (nsw) trial occurring two trials before (S-2) or after (S+2) a switch.

An incorrect feedback was given during non-switch trials whenever the response time exceeded a predefined limit. The limit was adjusted based on participants' performance after each switch trial: it increased by a 40 ms increment after incorrect switch trials and decreased by the same increment after correct switch trials, ensuring adaptive stringency. Note that this procedure did not censor observed RT distributions and maintained a consistent level of time pressure throughout the task as participants quickly reached a (fixed) floor value (300 ms).

Behavioral performance: the cost of switching

Given the striking similarity between OCD and epileptic patients' behavioral patterns (see Fig. 1), in the following, we report behavioral statistics pooled across the two groups of patients. It is noteworthy that these behavioral patterns closely mirrored those observed in healthy controls (Fig. 1b), while we also found consistent findings from ANOVAs followed by post-hoc tests performed separately for each group and even for each patient within the group (dmPFC group: Fig. S1; STN group: Fig. S2; see also Table S3 for statistics). First, we performed an analysis of variance (ANOVA) on mean reaction times (RTs), which demonstrated a significant interaction between trial type (switch vs. non-switch) and accuracy (hit vs. error) ($F_{(1,24)} = 15.12$; $p = 6.98 \times 10^{-4}$; Fig. 1b; repeated measures ANOVA). Post-hoc comparisons confirmed that RTs on correct trials were significantly longer on switch trials in comparison to non-switch trials (99 ± 11 ms; $t_{(6)} = 9.14$; $p = 9.66 \times 10^{-5}$; paired two-tailed Student's t-test). RTs were also significantly faster during incorrect switch trials relative to incorrect non-switch trials, supporting the idea that errors during switch trials resulted from a failure to override the preliminary response to apply the new rule (-86 ms \pm 28 ms; $t_{(6)} = -3.07$; $p = 2.02 \times 10^{-2}$; Fig. 1b; paired two-tailed Student's t-test). Unsurprisingly, patients also made significantly more errors on switch trials in comparison to non-switch trials (switch cost_(error) = 26 \pm 3%; $t_{(6)} = 8.09$; $p = 1.09 \times 10^{-4}$; paired two-tailed Student's t-test).

Behavioral switching: Neural activity in the dmPFC-STN network

To investigate whether neural activity in the dmPFC-STN network reflected behavioral switching, we first contrasted the power estimated in the time-frequency domain between correct switch (Sw/Hit) and correct non-switch (NoSw/Hit) trials across all recording sites within each brain region. dmPFC and STN LFP recordings were analyzed separately since these data came from distinct groups of patients (dmPFC: $n = 33$; STN: $n = 24$ recording sites). There was a significant increase in theta activity at rule onset when switching, both in the dmPFC (Sw/Hit vs NoSw/Hit, $[-156$ ms; $+664$ ms], $\text{sum}(t_{(23)}) = 232.7$, $p_c = 1.33 \times 10^{-4}$, Fig. 2a, b) and in the STN ($[4$ ms; $+1088$ ms], $\text{sum}(t_{(23)}) = 653.5$, $p_c = 1.00 \times 10^{-4}$, Fig. 3a, b), as well as an increase in high-gamma activity (dmPFC: $[+98$ ms; $+625$ ms], $\text{sum}(t_{(32)}) = 488.4$, $p_c = 1.67 \times 10^{-5}$, Fig. 2a; STN: $[-523$ ms; $+189$ ms], $\text{sum}(t_{(23)}) = 327.8$, $p_c = 5.58 \times 10^{-4}$, Fig. 3a). These switch-related changes in theta and high-gamma activity replicated when time-locking iEEG activity at response, which further revealed an increase in dmPFC's beta (15–30 Hz) activity (Sw/Hit vs NoSw/Hit; $[-631$ ms; 229 ms], $\text{sum}(t_{(23)}) = 743.6$; $p_c = 1.67 \times 10^{-5}$, Fig. 2a and Fig. S3).

Next, we focused our analyses on contacts exhibiting a differential activity when there was a change of stimulus-response mapping or not (i.e., contrasting switch vs. non-switch trials independently from accuracy) to investigate the precise time course of theta (5–10 Hz) or gamma (60–200 Hz) band activities in the dmPFC (Fig. 2) and in the STN (Fig. 3). In the dmPFC, we found that the timing of theta and gamma activities differed between switch hit and switch error trials. Theta activity peaked at response when patients failed to switch (0 ± 0.32 ms), but peaked significantly earlier when the switch was successful (Sw/Hit vs. Sw/Err, -279 ± 60 ms; $t_{(13)} = -4.6$, $p = 4.54 \times 10^{-4}$ paired two-tailed Student's t-test). Consistent with previous neuronal recordings in monkey dmPFC3, we observed a similar pattern in the

high-gamma band where activity peaked before the response for correct switches (-129 ± 26 ms) and after the response when it was missed ($+90 \pm 40$ ms; Sw/Hit vs. Sw/Err, -219 ± 42 ms; $t_{(8)} = -5.23$; $p = 7.93 \times 10^{-4}$; paired two-tailed Student's t-test, see Fig. 2e). This high-gamma activity peak in the dmPFC reliably followed the peak of theta activity (Sw/Hit_{theta} vs. Sw/Hit_{high-gamma}, -149 ± 44 ms; $t_{(21)} = -2.5$; $p = 2.02 \times 10^{-2}$ unpaired two-tailed Student's t-test). Furthermore, both theta and gamma activities were also significantly higher for incorrect than for correct switch trials (Theta band: Fig. 2c: Sw/Hit vs. Sw/Err, $[-299$ ms, $+522$ ms] relative to response, $\text{sum}(t_{(13)}) = 85.1$, $p_c = 1.67 \times 10^{-5}$; Gamma band: Fig. 2E, Sw/Hit vs. Sw/Err, $[-6$ ms, $+366$ ms] relative to response, $\text{sum}(t_{(8)}) = 57.7$, $p_c = 1.67 \times 10^{-5}$). Taken together, these results demonstrate that successful behavioral switching depends on a structured pattern of increased theta/high-gamma activity in the dmPFC from rule onset to response and precisely controlled in both time and amplitude. Consistent with this hypothesis, we found a significant amplitude-amplitude coupling between high gamma and theta activities in the dmPFC during correct switch trial (Fig. S4; $[0-500$ ms] time window post-rule onset; $t_{(13)} = 4.96$; $p = 2.61 \times 10^{-4}$; one sample two-tailed Student's t-test). This coupling was no longer significant during switch error trials ($t_{(13)} = 2.16$; $p = 5.04 \times 10^{-2} > 0.05$). Comparing directly the coupling strength for switch hit vs. switch error confirmed a higher theta-gamma coupling for correct switches ($t_{(26)} = 2.27$; $p = 3.14 \times 10^{-2}$; paired two-tailed Student's t-test).

In the STN of OCD patients, similar to what we observed in the dmPFC of epileptic patients, theta activity was significantly higher for incorrect than for correct switch trials (Sw/Hit vs. Sw/Err, see Fig. 3c, $[-0.04$ s, 0.39 s] relative to response; $\text{sum}(t_{(7)}) = 24.1$, $p_c = 1.91 \times 10^{-3}$). Moreover, theta activity in the STN peaked at response (Sw/Hit: -0.076 ± 0.042 s) when OCD patients successfully switched, but peaked after the response when they failed (Sw/Err: $+0.172 \pm 0.051$ ms; Sw/Hit vs. Sw/Err, -219 ± 39 ms; $t_{(7)} = -5.37$; $p = 1.00 \times 10^{-3}$; paired two-tailed Student's t-test), with a timing (relative to response) reminiscent of the dmPFC's high gamma activity observed in the group of epileptic patients (Sw/Hit vs. Sw/Err, -219 ± 39 ms; $t_{(8)} = -5.23$; $p = 7.93 \times 10^{-4}$; paired two-tailed Student's t-test). Finally, there was no difference in STN high gamma activity prior to/at the response between correct and incorrect switches, which reflected instead the upcoming motor response (Ipsilateral vs. Contralateral, Fig. S5). Instead, we found that theta increase for successful behavioral switches was larger in the STN ipsilateral to the newly selected response than in the contralateral STN, but not for unsuccessful switches (see Fig. S6, $t_{(7)} = 2.15$; $p = 0.03$), suggesting a direct role for STN theta oscillations in rule-switch execution. Recognizing that dmPFC and STN recordings were obtained from distinct patient groups, thereby precluding dmPFC-STN connectivity analyses, we thought that it remained interesting to compare theta dynamics across these patient groups and structures. Our analysis revealed a consistent temporal lag in rule-switch-related theta increase in the STN compared to the dmPFC (dmPFC-STN theta onsets relative to the rule: dmPFC: -0.419 ± 0.03 s; STN: -0.234 ± 0.08 s; $t_{(20)} = -2.57$, $p = 1.81 \times 10^{-2}$, see Fig. S7). Moreover, the theta increase peaked significantly later in the STN than in the dmPFC (peak latencies in dmPFC: $+0.199 \pm 0.074$ s vs. $+0.575 \pm 0.052$ s in the STN, $t_{(20)} = -3.54$, $p = 2.08 \times 10^{-3}$). Overall, this dmPFC-STN theta/high-gamma dynamic suggests that the dmPFC might drive STN neural activity when external cues trigger a rule-switch and casts the STN as an executive structure downstream the dmPFC along the hyperdirect pathway.

Consistent with this view, we found that neural activity in the dmPFC further encoded the anticipation of upcoming rule switches: dmPFC's beta activity at stimuli onset increased in proportion with the probability of rule changes (dmPFC: $[-846$ ms; -534 ms], $\text{sum}(t_{(32)}) = 59.99$, $p_c = 9.33 \times 10^{-4}$, Fig. S8), 500 ms before the onset of the cue indicating that the rule had changed. Indeed, in our task, the

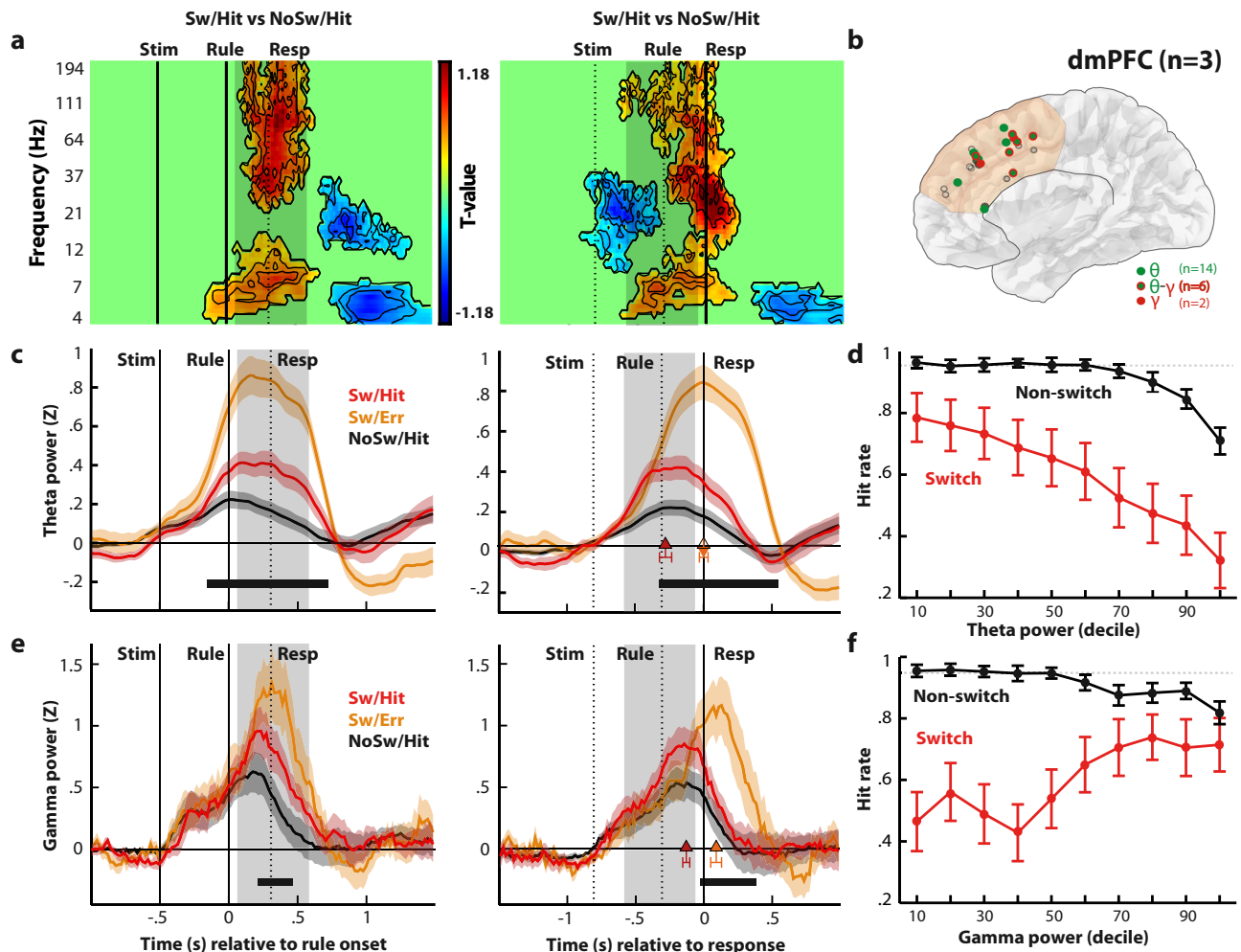


Fig. 2 | dmPFC neural activity during behavioral switching. **a** Time-frequency analysis of neural activity time locked to rule (left panels) or on response (right panels) onset across all dmPFC contact-pairs ($n = 33$). Warm (cold) colors indicate significant increases (decreases) of power ($p_c < 0.05$, FWE cluster-corrected). **b** Anatomical location of recording sites in the dmPFC plotted on a 3D reference brain. Each colored dot represents a contact-pair displaying a significant modulation in the theta (green) and/or high-gamma (red) during task-switching. **c** Time course of theta power (5–10 Hz) averaged across correct (red) switch, incorrect switch (orange) or correct non-switch (black) trials ($n = 14$ dmPFC contact-pairs). Bold traces indicate average activity and shaded areas correspond to SEM across contacts. The black horizontal bar at the bottom indicates time points for which the statistical contrast between incorrect and correct switch trials was significant ($p_c < 0.05$). Orange (red) vertical arrows at the bottom indicate average peak

latencies for error (hit) switch trials (averaged across contact-pairs; horizontal lines indicate SEM). Vertical grey shaded rectangles correspond to 95% confidence intervals of RT (left panels) or rule onset (right panels). **d** Neuro-psychometric curves depicting the relationship between hit rate and averaged theta power (across $n = 14$ dmPFC contact-pairs) binned into deciles separately during switch and non-switch trials. Error-bars correspond to SEM across contact-pairs. **e** Time course of high-gamma power (60–200 Hz; $n = 8$ dmPFC contact-pairs). Bold traces indicate average activity and shaded areas correspond to SEM across contacts. **f** Neuro-psychometric curves depicting the relationship between hit rate and averaged high-gamma power (across $n = 8$ dmPFC contact-pairs) binned into deciles separately during switch and non-switch trials. Error-bars correspond to SEM across contact-pairs. dmPFC: dorsomedial prefrontal cortex; Sw: switch; NoSw: non-switch.

probability of switch increased with the number of previous consecutive non-switch trials. Patients implicitly used this information to anticipate switches so that switch costs were large if a switch occurred after 2–3 non-switch trials (switch $\text{cost}_{(2\text{nsW})} = -43 \pm 13\%$, switch $\text{cost}_{(3\text{nsW})} = -37 \pm 7\%$). This effect disappeared if a switch occurred after 5–6 non-switch trials (switch $\text{cost}_{(5\text{nsW})} = -17 \pm 14\%$, switch $\text{cost}_{(6\text{nsW})} = 3.4 \pm 13\%$, see Fig. S8).

It is noteworthy that our main results were statistically robust and reproduced when using mixed-effect analyses to assess differences in theta activity between correct switch and non-switch trials, both in the dmPFC ($t_{(5455)} = 4.72$, $p = 2.4 \times 10^{-6}$; see also Fig. S9) and in the STN ($t_{(6244)} = 3.69$, $p = 0.00023$; see also Fig. S10), as well as between hit and error switch trials (dmPFC: $t_{(1284)} = 4.26$, $p = 2.2 \times 10^{-5}$; STN: $t_{(1332)} = 2.33$, $p = 0.019$). Finally, we tested whether baseline had an influence on subsequent theta dynamic in dmPFC and STN by comparing the time course of theta power across conditions separately for

high vs. low level of theta power prior rule-onset (Fig. S11; from -250ms to 0 ms; median split). This revealed that neither dmPFC (Fig. S11A–C) nor STN (Fig. S11D–F) baseline theta power significantly modulated the post-rule theta peak amplitude across conditions.

Trial-by-trial fluctuations in theta/high-gamma bands predict successful behavioral switching in dmPFC and in the STN

To explore the functional role of dmPFC-STN theta/high-gamma dynamic, we then investigated whether spontaneous trial-by-trial fluctuations in neural activity related to actual fluctuations in behavioral performance. To do so, we tested brain-behavior correlations between normalized theta or high-gamma activity and performances during switch and non-switch trials separately for each brain region and for each trial type (dmPFC: Fig. 2d–f; STN: Fig. 3d–f).

During switch trials, there was a negative correlation between theta activity trial-by-trial fluctuations and performances, both in the

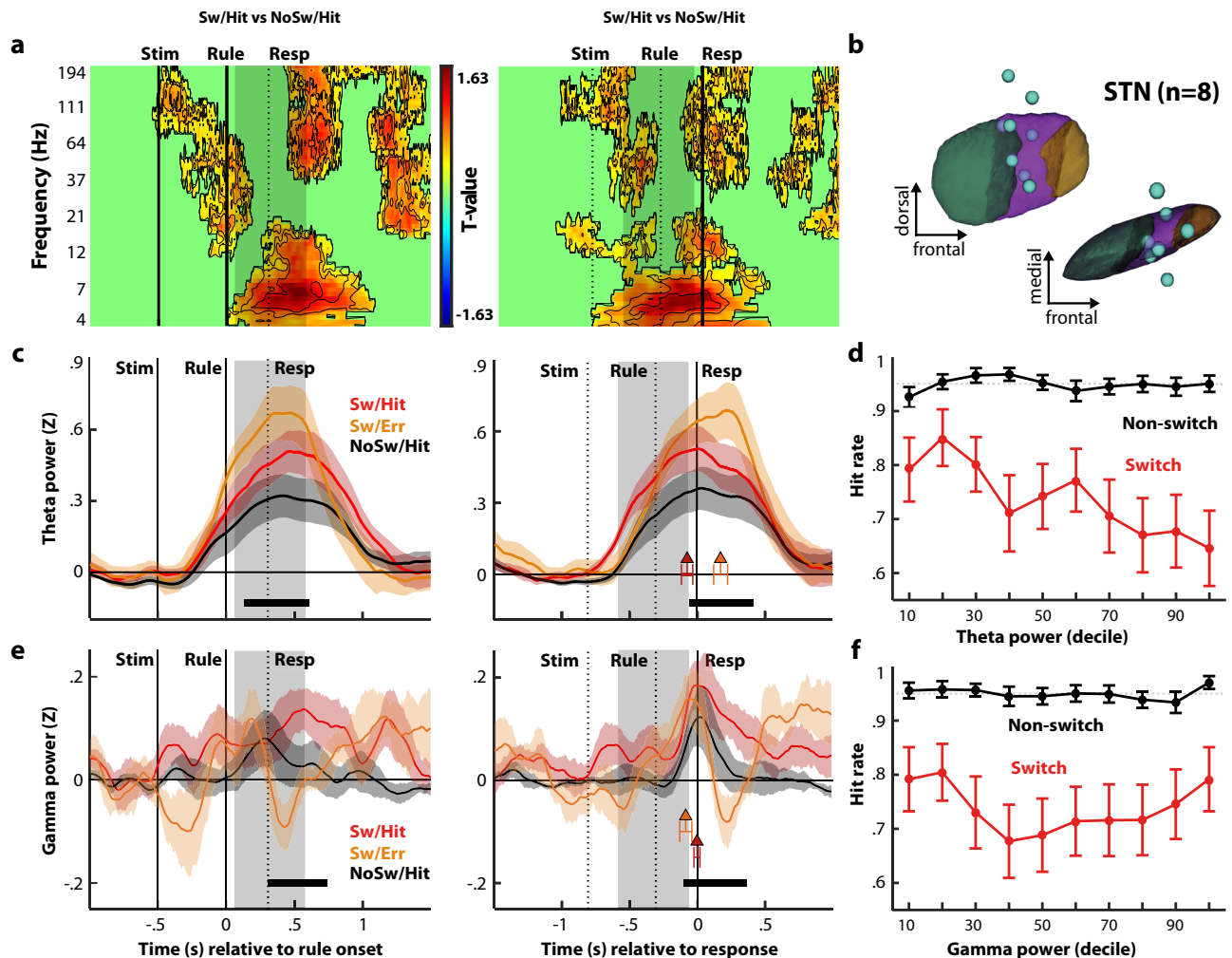


Fig. 3 | STN neural activity during behavioral switching. **a** Time-frequency analysis of neural activity time locked on rule (left panels) or on response (right panels) onset (contrasts across all STN contact-pairs; $n = 24$; $p_c < 0.05$ FWE cluster-corrected). **b** Anatomical location of STN recording locations with higher behavioral switch related theta increase (blue dots) displaying switch-related theta activity. Each contact-pair is plotted on a 3D reconstruction of STN sensorimotor (in green), associative (in purple) and limbic (in brown) territories. **c** Time course of theta power (5–10 Hz) averaged across correct (red) switch, incorrect switch (orange) or correct non-switch (black) trials ($n = 8$ STN contact-pairs). Bold traces indicate average activity and shaded areas correspond to SEM across contacts. The black horizontal rectangle at the bottom indicates time points for which the statistical contrast between incorrect and correct switch trials was significant ($p_c < 0.05$). Orange (red) vertical arrows at the bottom indicate average peak latencies for error

(hit) switch trials (average across contact-pairs; horizontal lines indicate SEM). Vertical grey shaded rectangles correspond to 95% confidence intervals of RT (left) or rule onset (right). **d** Neuro-psychometric curves depicting the relationship between hit rate and averaged theta power (across $n = 8$ STN contact-pairs) binned into deciles separately during switch and non-switch trials. Error-bars correspond to SEM across contact-pairs. **e** Time course of high-gamma power (60–200 Hz; $n = 8$ STN contact-pairs). Bold traces indicate average activity and shaded areas correspond to SEM across contacts. **f** Neuro-psychometric curves depicting the relationship between hit rate and averaged high-gamma power (across $n = 8$ STN contact-pairs) binned into deciles separately during switch and non-switch trials. Error-bars correspond to SEM across contact-pairs. STN subthalamic nucleus. Sw switch; NoSw: non-switch.

dmPFC and the STN (dmPFC: Fig. 2d; $\beta_{\text{hit-switch}} = -0.33 \pm 0.06$; $t_{(13)} = -5.67$, $p = 7.65 \times 10^{-5}$; STN: Fig. 3d; $\beta_{\text{hit-switch}} = -0.13 \pm 0.03$; $t_{(7)} = -4.24$, $p = 3.82 \times 10^{-3}$). However, during non-switch trials, there was no influence of STN theta activity trial-by-trial fluctuations on performances, while a negative correlation between theta activity and performances was still found in the dmPFC (theta: $\beta_{\text{hit-nonswitch}} = -0.14 \pm 0.02$; $t_{(13)} = -6.19$, $p = 3.29 \times 10^{-5} < 0.01$), suggesting a functional decoupling of STN and dmPFC theta activities during non-switch trials. Interestingly, we also found significant positive correlations between performances and pre-response high-gamma activity in the dmPFC during switch trials (Fig. 2f; $\beta_{\text{hit-switch}} = 0.18 \pm 0.05$; $t_{(7)} = 3.32$, $p = 1.06 \times 10^{-2}$), whereas this correlation was negative during non-switch trials (Fig. 2f; $\beta_{\text{hit-nonswitch}} = -0.09 \pm 0.01$; $t_{(7)} = -5.66$, $p = 4.75 \times 10^{-4} < 0.01$). Thus, during non-switch trials, there was a negative correlation (i.e., like what was observed in its theta band and

in pre-response beta, see Fig. S12), whereas this correlation became strongly positive during switch trials, suggesting that dmPFC pre-response high gamma activity reflects a neural process selective to behavioral-switching.

Theta activity is negatively associated with the starting point during behavioral switching in the dmPFC and in the STN

Finally, to further understand how these trial-by-trial fluctuations in neural activity influenced behavior, we used a hierarchical drift-diffusion model (HDDM³⁰) to disentangle behavioral variability arising from switching between rules vs. selecting a new response. At the behavioral level, drift diffusion models (DDMs) are commonly used to model the dynamics of action selection as an accumulation of evidence over time until a certain threshold is reached and a response is triggered³¹ (see Fig. 4a). DDMs have four main parameters²⁶: the time of

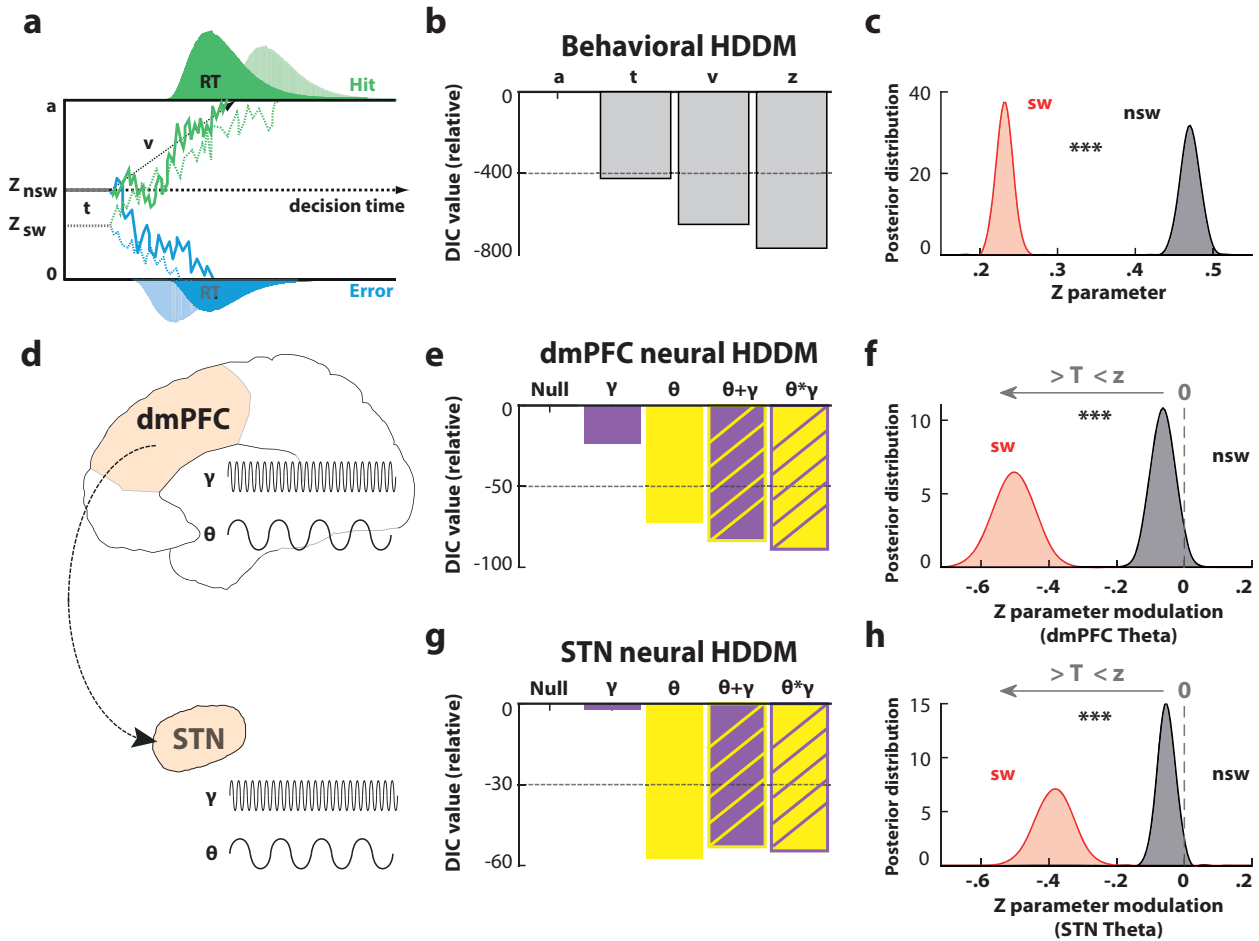


Fig. 4 | Behavioral switching modeling across epileptic and OCD participants. **a** Drift diffusion model. Response execution is preceded by an accumulation of evidence increasing sequentially in favor of one of the two options until a boundary is reached. The decision processes start at stimulus onset and at an initial level depending on the subject's prior beliefs. The initial level turns out to be different per trial type. **b** Behavioral model comparison. Relative value of the deviance information criterion (DIC) per model when considering the decision threshold model *a* as a reference. **c** Behavioral model posteriors. Probability density of the initial level of evidence values for switch and non-switch trials. The initial level of evidence is lower during switch trials in accordance with subjects' belief of an upcoming non-switch trial. The statistical difference between switch and non-switch posteriors was significant (significance from posterior probabilities:

$p < 0.001$). **d** Spontaneous trial-by-trial activities included in the neural HDDM (hierarchical drift diffusion model) model space. Neural model comparison for dorsomedial prefrontal cortex (dmPFC; **e**) and subthalamic nucleus (STN; **g**). Relative value of the DIC per model with a model based on a normally distributed noise as a reference. Neural model posteriors for dmPFC (**f**) and STN (**h**). Probability density for the effect of theta power on the initial level of evidence *z*. A negative (resp. positive) regression coefficient means *z* decreases (resp. increases) when theta power increases. Here, the regression coefficient is negative for non-switch trials and strongly negative for switch trials. The statistical difference between switch and non-switch regression coefficient was significant (significance from posterior probabilities: $p < 0.001$). *** $p < 0.001$.

non-decision *t*, the initial level of evidence *z*, a drift rate *v* (the accumulation speed of the information relevant to action selection) and thresholds *a*, which controls the speed-accuracy tradeoff of the selection process and accounts well for its type I errors^{9,28}. In our task, we expected to find a shift of the starting point for switch trials, (1) as correct switches were associated with slower RTs and failed switches with faster RTs (compared to correct non-switch responses; see Fig. 1b), (2) as dmPFC encoded a priori the likelihood of upcoming switches and (3) as early behavioral studies suggested that, during task switching, reconfiguring the set of rules seemed to occur before the selection of a new response^{26,27}. To test this hypothesis, we first fitted a hierarchical drift diffusion model to RT distributions of correct and incorrect responses from all patients³¹ (i.e., combining the behavioral data of OCD and epileptic patients), and tested which model parameter best accounted for observed switch costs (MI model space, also included all possible parameter pairs, see Methods).

We found that switch costs were best captured by a shift of the starting point (*z* parameter) toward the lower (erroneous) model

boundary (Fig. 4b). Moreover, the sampled posterior distribution of *z* parameter indicated that HDDM starting point was significantly lower during switch trials ($z_{sw} = 0.23 \pm 0.06$) compared to non-switch trials ($z_{nsw} = 0.44 \pm 0.06$; $p < 0.001$; Fig. 4c). We found identical results when testing which model best accounted for healthy subjects' behavior (see Fig. S13). Posterior predictive checks confirmed that this HDDM model reproduced all key behavioral patterns observed after stimulus-response mapping changes (see Fig. 1b), for each individual in each group (dmPFC Epileptic group; see Fig. S14; STN OCD group; see Fig. S15).

Next, we tested whether dmPFC and STN neural activity further modulated that initial level evidence (*z*). We reasoned that using band-specific neural fluctuations to adjust trial-by-trial HDDM's starting point would significantly improve our model fit only if it reflected the switches between rules. Hence, we built a second neural HDDM model space consisting of HDDMs with starting point modulations by all the switch-related bands identified in our previous analysis of STN and dmPFC neural activity, as well as linear and

multiplicative interactions with these frequency bands (M2 model space, see Methods, Fig. 4d).

In the dmPFC, the neural HDDM including theta fluctuations as a modulator of the starting point outperformed other single-band models and was not significantly outperformed by more complex neural HDDMs including an interaction term between theta and high-gamma neural activities (see Fig. 4e). Consistent with our other findings in the STN, the neural HDDM model including theta fluctuations as a modulator of the starting point significantly outperformed all the other neural HDDM tested (Fig. 4f). Figure 4f–h show the posterior distribution of the regression coefficient between HDDM starting point and residual trial-by-trial power in the theta band. The significant shift of the posterior distribution toward negative correlation coefficients between the starting point and theta power residuals (dmPFC: Fig. 4f; STN: Fig. 4h, both $p < 0.01$) is consistent with our previous observation that, although theta activity in the dmPFC-STN network increases on average during switch trials, too high an increase becomes detrimental to performances. Interestingly, this negative correlation between the starting point and residual trial-by-trial theta power fluctuations during switch trials was also observed as early as 250 ms before rule onset in both the dmPFC and in the STN (see Fig. S16), at a moment when there was no difference on average between correct and erroneous switch. Thus, these results are consistent with the role of a prior modulation, which is usually associated with the starting point parameter.

Some key brain regions, known to exhibit rule-switch activity are not mentioned in this study because an inherent limitation to intracranial EEG recordings is their limited spatial sampling. Note however that the intracranial recordings from the epileptic group allowed us to explore brain activity beyond the hyperdirect dmPFC-STN circuitry. For example, we found that broadband gamma (60–200 Hz) and beta (15–35 Hz) power in the anterior insular cortex was higher during correct switch trials than during non-switch trials (Fig. S17). The main differences with the dmPFC were that (1) there was no modulation of theta band activity in alns and (2) the increase in both beta and broadband gamma amplitude during incorrect behavioral switches was only observed after the response, thus lying out of our temporal window of interest (in this study, we focused on the exploration of the cognitive processes related to switch between rules).

Discussion

In this study, we used human iEEG recordings in either the dmPFC or in the STN to investigate the neural mechanisms associated with behavioral switching. Our findings reveal new insights into (1) how temporally coordinated neural patterns in the theta/gamma frequency bands in the dmPFC-STN hyperdirect pathway support adaptive behavioral switching (2) to the extent that trial-by-trial neural fluctuations of these neural patterns were predictive of fluctuations in observed behavior. (3) We also demonstrate a mechanism for behavioral switching casting switch costs as a modulation of the starting point in a drift-diffusion process tied, in turn, to a modulation of theta power oscillations in the dmPFC-STN circuit. In the following sections, we will delve into these three lines of results in detail.

By contrasting neural activity between successful trials with and without behavioral switching (Figs. 2 and 3), we identified a significant increase of theta and high-gamma activity in the dmPFC-STN circuit between rule onset and response, i.e. when cognitive processes underlying task-switching are theoretically expected to occur. When contrasting successful and failed switches, we further found that switching accuracy did not only depend on the magnitude of theta and gamma activities: increases of theta (in the STN and dmPFC) and gamma (only in the dmPFC) band activities were also found to peak before the response during successful switches and at the time of the response or later during failed switches (Figs. 2 and 3, panels C and E). Previous monkey electrophysiological studies also reported that

neuronal spiking increased in the dmPFC-STN circuit when switching^{3,4}, at timings similar to the increase in high-gamma activity observed in our study for successful and failed switches. Taken together, this suggests that a plausible neurophysiological mechanism underlying our findings would be that, in the dmPFC-STN circuit, increased theta band activity may control the timing of the increase in neuronal spiking activity³², reflected in our data as an increased activity in the high-gamma just before response. Interestingly, whereas recent iEEG recordings during task switching or stroop paradigms focused on high gamma oscillations^{29,33}, these studies did neither investigate the relative timing of theta and gamma activities, nor how these neural proxies relate to behavior. In this study, we found that high-gamma peak activity lagged significantly behind theta's during successful switch trials, suggesting that dissociable temporal components of switching might be multiplexed by distinct frequency regimes in the dmPFC. This view is also consistent with the amplitude-amplitude coupling between theta and gamma activities observed in the dmPFC during correct behavioral switch, but not during erroneous switch. An alternative interpretation of the difference of latency observed between theta vs. high gamma peak activity would be that high-gamma increase would partly depend on the preceding theta activity, in line with the extensive literature related to the theta-gamma neural coding schemes underlying several cognitive processes³². The modulation of theta amplitude found in the dmPFC and in the STN during switching is consistent with previous EEG studies in healthy participants^{23,25} and with previous iEEG studies on executive control functions recording LFP across several cortical or subcortical targets^{34,35,10,13}. Critically, we also found that theta activity magnitude further increased during unsuccessful trials compared to correct switches in both brain structures. Yet, unlike the large body of studies that previously interpreted theta increases during correct and incorrect trials as if these would reflect two independent executive control functions, such as conflict monitoring (in case of success) or error monitoring^{34,35} (in case of failure), we propose an alternative view. We identified a single mechanism explaining parsimoniously the pattern of theta oscillations observed in either the dmPFC or the STN: following rule change, theta oscillations amplitude lowers the initial level of evidence available in favor of the new correct response (Fig. 4) such that success would only occur for an optimal level of theta oscillations in the circuit.

To clarify the respective role of theta and gamma band activities in the prefronto-subthalamic circuit, we examined how the theta and gamma trial-by-trial amplitude fluctuations were associated with behavioral variability across trials, separately for the dmPFC and the STN. The initial observation was counter-intuitive, since we found that theta activity was negatively associated with hit rate on switch trials in both the dmPFC and the STN. In other words, switching was associated with an increased theta activity observed when contrasting switch vs. non-switch trials, but trial-to-trial analyses revealed that the greater the increase, the lower the switch performance (Figs. 2d and 3d). Strikingly, there was no such clear-cut relationship between theta trial-by-trial fluctuations and performance for non-switch trials, highlighting its functional dependence to the cognitive context. This finding echoes a previous study showing complex non-linear modulations of decision threshold by STN theta oscillatory activity during simple perceptual decision-making, with a reversal of the relationship as a function of the level of perceived choice difficulty⁸. At the behavioral level, we found that switch trials (compared to non-switch trials) were best captured by a shift of the starting point (z parameter) toward the lower (erroneous) model boundary in a HDDM³¹. This result contrasts with previous demonstration in patients with Parkinson's disease during which STN oscillations were found to modulate the threshold during difficult decisions involving perceptual or response "conflicts"^{8,28,36}. Differences in the cognitive processes studied are likely to explain this discrepancy, as suggested by the global slowing down reported in these studies for both correct and erroneous responses when conflict was

high. In our task, the erroneous response at switch trials is the response one would have selected if the rule had not been changed. In the context of a reactive behavioral switching paradigm inducing reliable increase of RTs when successfully switching between rules, but also premature/fast guess errors when unsuccessful, the fact that the key HDDM parameter associated with behavioral switches was the initial level of evidence is consistent with the idea that subjects were able to prepare the correct response ahead of knowing the rule to be used and, at switch trials, had to inhibit (or reset) this pre-selected response, switch to an alternative set of rules and then reassess their choice according to it, as previously suggested^{26,27}. Interestingly, we also found that switch probability scaled with pre-rule onset beta activity in the dmPFC (but not in the STN), suggesting that this structure actively monitored and predicted these cognitive events.

At the neural level, we compared several neurally-informed HDDM, in which we tested the benefit of adding either theta or high gamma trial-by-trial activity, or more complex additive or multiplicative models, as factors further modulating HDDM starting point (against no additional neural information). This analysis revealed that across the dmPFC-STN circuit, the most parsimonious mechanism explaining our behavioral results was that trial-by-trial theta fluctuations in this circuit negatively scaled with the starting point parameter of the HDDM. Moreover, this negative correlation was much larger for switch trials (compared to non-switch trials). In other words, when theta power increased at switch trials, it pushed the starting point further toward the lower (erroneous) boundary of the HDDM. This mechanism simply explains how a moderate theta increase allowed patients to buy some time to form their decision when switching to a new set of rules (thus predicting longer correct responses), but also why incorrect switches were more frequent and characterized by very fast reaction times, as more theta activity was observed on average during those trials and its effect on the starting point was amplified, pushing it further toward the lower (erroneous) boundary. The proposed mechanism echoes previous findings from neuroimaging studies in healthy participants suggesting that neural activity in the dmPFC-STN loop may be associated with decision threshold changes captured by drift diffusion models^{9,37} and from studies reporting that applying deep brain stimulation to the subthalamic nucleus of patients with Parkinson's disease yield complex changes in the regulation of the decision threshold^{8,28,36,38,39}. To conclude, note that within the theoretical framework of DDMs, only a shift in a starting point can generate such a pattern of RTs combining slower correct and faster erroneous responses for switch trials, as well as an increased error rate (all relative to non-switch trials). Explaining such behavioral patterns is the original motivation behind the introduction of changes (and variability) in DDM's starting point⁴⁰. Hence, it is especially meaningful that using trial-by-trial neural variability actually improve the HDDM fits (see Fig. 4e–g).

An obvious limitation of this study is that our data were collected from patients with neuropsychiatric disorders so that disease-related factors could have influenced the neural or cognitive processes of interest. However, it was reassuring to observe that the behavior of patients and healthy subjects was remarkably similar while the neural results were consistent with the results of the animal literature^{3,4}. Another limitation is that the dmPFC and STN were recorded in two different groups of subjects, so that direct comparisons between brain regions should be interpreted with caution. That said, differences in pathology or recording modality between the two groups cannot explain the almost identical behavioral and neural patterns we found within each group. The only direct comparison we report pertains to the latencies of theta activity, with earlier onset/peak observed in the dmPFC group compared to the STN group. This is in line with the literature⁴ and with computational models of the cortico-basal ganglia architecture thought to underlie behavioral switching⁴¹. In addition, some key brain regions consistently found to exhibit task-switching

signals are absent in our results because an inherent limitation to intracranial EEG recordings is their limited spatial sampling. Yet, we could observe that the anterior insula differentially contributes to behavioral switching compared to the dmPFC (with later timings and a distinct frequency regimes), confirming the interest of intracerebral data to distinguish the functional role of the dmPFC and the anterior insula³⁵, which are regions often co-activated in functional neuroimaging studies so that their respective functions are still debated. Examining more precisely the dynamics of task-switching signals in an extended fronto-insular-parietal task-switching network would require additional experiments across a much larger population of implanted epileptic patients performing a task-switching paradigm.

To conclude, we used intracerebral recordings in humans to specify the neural and computational mechanisms through which humans switch between behaviors: theta activity in the dmPFC-STN circuit adjusts the initial level of evidence such that an optimal level of activity is required to successfully reconfigure the task-set (i.e., stimulus-response mapping rule, here). Our results may prove useful in the development of new deep brain stimulation procedures. Recent studies have demonstrated that stimulating the STN when a pathological level of beta oscillation is detected in the sensorimotor territory of the STN is clinically effective in the context of Parkinson's disease⁴². Nevertheless, the non-monotonic functional associations between theta/gamma activity and complex cognitive processes revealed in our study indicate that a threshold-based closed-loop neurostimulation strategy applied to theta oscillations may not always be sufficient in the context of neuropsychiatric diseases.

Methods

Patients and electrode implantation

The study was approved by local ethics committees (CPP 09-CHUG-12, study 0907 and 2011- A00083-38). All patients volunteered to participate and provided written informed consent prior to participation.

OCD participants. We included four patients (1 male; mean age 36 ± 3 years old) undergoing bilateral surgical implantation of DBS electrodes with 4 contact leads of 1.5 mm height and 0.5 mm apart (Model 3389; Medtronic, Minneapolis, Minnesota, US) implanted in the STN to treat severe treatment-resistant OCD. These patients were enrolled for STN-DBS according to standard inclusion/exclusion criteria⁴³ and were kept on medication during the experiment (detailed clinical and demographic data are reported in Table S1-a).

Epileptic participants. Three epileptic patients (no males; mean age 32 ± 2 years old; clinical and demographic details in Table S1-b) suffering from drug-resistant focal epilepsy participated in the study. These patients underwent intracranial cortical recordings by means of stereotactically implanted depth macroelectrodes for therapeutic purposes. All target structures were selected according to clinical considerations independent from this study. We recruited patients with at least one electrode in the dorsomedial prefrontal cortex (dmPFC). Nine to twenty semi-rigid macroelectrodes were implanted for each patient. Each macroelectrode had a diameter of 0.8 mm and, depending on the target structure, contained 6–18 contact leads of 2 mm wide and 1.5–4.1 mm apart (Dixi Medical, Besançon, France).

STN intracranial recordings

We performed intracranial EEG recording from quadripolar DBS electrodes bilaterally implanted in the STNs of four severe and treatment resistant OCD patients using standard clinical procedures (Neurosurgery Department of Grenoble University Hospital) 2–4 days after surgery. LFPs were recorded using temporary lead extensions connected to an EEG acquisition system (Micromed SD MRI, bandwidth 0.15–600 Hz, sampling rate 2048 Hz) before the definitive connection between the electrodes and the stimulation device. Each DBS electrode

consisted of 4 contacts with a length of 1.5 mm, separated by 0.5 mm (macro-electrode 3389, Medtronic, Minneapolis, US). We used the tip of the left electrode as reference and computed offline the bipolar derivations between the 3 adjacent pairs of contacts of each electrode to minimize the contribution of sources outside the STN to LFPs. In this study, we only report results based on bipolar derivations because it allows for better signal artifacts removal and achieve high spatial resolution (about 3 mm) by canceling out contributions of distant sources, which spread equally to adjacent recording sites.

STN electrode localization

We reconstructed STN contact locations using a 3D histologic atlas of the basal ganglia warped onto a pre-surgery T1 weighted MRI co-registered to a post-surgery 3D helical TDM⁴⁴. By analogy with tract tracing data from previous monkey studies, this atlas distinguishes 3 functional territories in the STN⁴⁵: posterior/sensorimotor (green in Fig. 2b), intermediate/associative (purple), and anterior/limbic parts (brown). STN contacts shown in Fig. 2b are in between the two contacts used for computing the bipolar derivations.

dmPFC intracranial recordings

Cortical recordings were performed using video-EEG monitoring systems that allowed for simultaneous recording of 128 depth-EEG channels sampled at 512 Hz. Acquisitions were made with the Micro-med (Treviso, Italy) system and online band-pass filtering from 0.1 to 200 Hz. Data were acquired using a referential montage with a reference electrode chosen in the white matter. Before analysis, all signals were re-referenced to their nearest neighbor on the same electrode, yielding a bipolar montage.

dmPFC electrode localization

The electrode contacts were localized and anatomically labeled using the *IntrAnat Electrodes* software⁴⁵. Briefly, the pre-operative anatomical MRI (3D T1 contrast) and the post-operative image with the sEEG electrodes (3D T1 MRI or CT scan), obtained for each patient, were co-registered using a rigid-body transformation computed by the *Statistical Parametric Mapping 12 (SPM12)*⁴⁶ software. The gray and white matter volumes were segmented from the pre-implantation MRI using *Morphologist* as included in *BrainVisa*. The electrode contact positions were computed in the native and MNI referential using the spatial normalization of SPM12 software. Coordinates of recording sites were then computed as the mean of the MNI coordinates of the two contacts composing the bipole. For each patient, cortical parcels were obtained for the *MarsAtlas*⁴⁶. Each electrode contact was assigned to the gray or white matter and to an anatomical parcel by taking the most frequent voxel label in a sphere of 3 mm radius around each contact center. Note that data from both hemispheres were collapsed to improve statistical power. Contact-pairs were assigned to the dmPFC region of interest if they belonged to either the caudal dorsomedial PFC or to the rostromedial PFC in the *MarsAtlas* parcellation scheme.

Behavioral task

We administered a task-switching paradigm adapted from Isoda and Hikosaka (2007; Fig. 1a). The task included four to five sessions of 200 trials. Each trial started with two colored squares (yellow/pink) randomly displayed on each side (left/right) of a white central square for 500 ms before revealing the ongoing rule (white central square turning either yellow or pink during 200 ms to indicate the current target color). Patients indicated their response (left/right with a gamepad Logitech F310S) within a 1000 ms interval by pressing one of two buttons with the index of the corresponding hand. Presentation of visual stimuli and acquisition of behavioral data were performed on a PC (19 inch TFT monitor with a refresh rate of 60 Hz) using presentation software (v14.1; Neurobehavioral systems, Albany, CA). In addition, a time pressure was used to encourage speeded response during

non-switch trials; it was initially set to 500 ms. This time window was adaptively adjusted by 40 ms steps according to patients' responses: it was shortened after successful switches, increased after failed switches and could not be inferior to 300 ms. Trials ended with a 1.5 s feedback screen displayed immediately after the response or after 1000 ms in the absence of response to indicate the response time and score of the patient using green fonts if the response was correct or a red font if it was incorrect. The score was incremented (decremented) by 1 after each successful (erroneous) non-switch trial and by 5 after each successful (erroneous) switch trial. This procedure ensured that speed and accuracy were balanced: time pressure promoted speeded response while our reward policy promoted high accuracy throughout the task. In between trials, a white square was displayed centrally for fixation (1–1.5 s). The rule (yellow/pink) randomly changed after 2 to 6 trials, so that there were 50 switch trials (and 200 non-switch trials) per session. We trained the patient prior to the recording session. First, we provided written instructions and, if necessary, reformulated them orally to ensure they were understood. Then, patients performed a training session of 20 trials (5 switch and 15 non-switch trials) that was eventually followed by a longer training session (maximum 200 trials) to ensure that the patient felt comfortable with the task. The electrophysiological acquisition system was synchronized to the system used to run the behavioral task via TTL pulses signaling the onset of the task events.

Behavioral analyses

Statistical analyses were performed with Matlab Statistical Toolbox (Matlab R2018a, The MathWorks, Inc., USA). All means are reported along with the standard error of the mean. Trials were excluded from further behavioral and electrophysiological analyses if patients responded before rule onset, if they took more than one second to respond or in the case of response omission.

Intracranial EEG analyses

iEEG data were first epoched from –3000 ms before rule onset to 3000 ms after rule onset. Time-frequency analyses were carried out using the *Fieldtrip* toolbox for Matlab. Spectral analyses were analyzed in two distinct frequency ranges using a multitaper approach. For the lower frequencies (4–32 Hz), we used adaptive temporal window lengths shifted in steps of twenty ms while fixing the number of cycles (6 cycles and 3 tapers per window. Conversely, for higher frequencies (32–200 Hz), we used a time window length of 187.5 ms shifted in steps of twenty ms with an adaptive number of tappers (4–31 tappers: the number of tappers decreased with increasing frequency). This approach uses a constant number of cycles across frequencies up to 32 Hz (hence a time window whose duration decreases when frequency increases), and a fixed time window with an increasing number of tappers above 32 Hz to obtain more precise power estimates by adaptively increasing smoothing at high frequencies. Spectral power was log-transformed to improve Gaussianity of the data, which allowed us to use standard parametric tests to assess the statistical significance of the observed effects. Theta (5–10 Hz) or high-gamma (60–200 Hz) activity time-courses were extracted by averaging spectral powers (in dB). Theta and gamma activities were normalized (Z-scored) using a [–1500 to –500 ms before rule onset]. We used the full time-frequency analysis to choose upper and lower boundaries for the frequency bands that were identical for dmPFC and for STN recordings, and consistent with previous intracranial studies^{47–50}.

Analyses of switch hit compared to non-switch hit trials in the time-frequency domain (Figs. 2a, 3a). For each contact, we computed the unpaired difference in frequency powers at each time step (in the –3000 to 3000 ms relative to rule onset) for every frequency (4–200 Hz) between correct switch vs/ correct non-switch trials. We then computed the corresponding T-values, which we averaged across contacts and participants for each region of interest (dmPFC or STN).

Statistical inferences were performed with a corrected for multiple comparison (corrected p values, noted $p_c < 0.05$; Bonferroni-Holmes Family-Wise-Error corrections at the cluster level; $n = 150$ permutations within contacts, $n = 50000$ permutations between contacts).

Selection of task-responsive contact-pairs

For each frequency band of interest (theta or high-gamma), we first selected contact-pairs exhibiting a significant switch-related activity. Switch responsive contact-pairs were identified by (1) extracting the (5–10 Hz) or (60–200 Hz) single-trial power averaged in the time windows during which there was a significant cluster-corrected increase of power time locked to the rule during correct switch trials relative to non-switch trials (dmPFC: [–150 to 560 ms]; STN: [0–1100 ms]). We next used these power estimates to contrast for each contact theta or high-gamma power during switch vs. non-switch trials (student unpaired t-test thresholded at $p < 0.05$). All dmPFC contact pairs displaying a significant effect for each frequency band were kept for further analyses because of the spatial sampling heterogeneity inherent to sEEG recordings from epileptic patients. In contrast, all OCD patients were implanted bilaterally within the STN so that we kept one recording site per hemisphere for the STN (choosing the highest T-value among the three contact-pair for each hemisphere), in line with previous human local field potential studies.

Analyses of switch error compared to switch hit trials in theta or high-gamma bands (Figs. 2c, e and 3c, e). For each task-responsive contact, we next computed the unpaired difference in powers at each time step (in the –3000 to 3000 ms relative to rule onset) between incorrect switch vs/ correct switch trials. We then computed the corresponding T-values, which we averaged across contacts and participants for each region of interest (dmPFC or STN). Statistical inferences were performed with a corrected for multiple comparison (corrected p values, noted $p_c < 0.05$; Bonferroni-Holmes Family-Wise-Error corrections at the cluster level; $n = 150$ permutations within contacts, $n = 50,000$ permutations between contacts).

Analyses of theta or gamma activity expressed in percentile and hit rate within each trial type (Figs. 2d, f and 3d, f). For each contact located within our regions of interest, we first sorted theta (or high-gamma) activity into 10 bins. Across the trials within each bin, we then computed for each trial type the accuracy. We then regressed neural activity (expressed in percentiles) and behavioral accuracy separately for each trial type (switch or non-switch trials) for each contact to extract a beta estimate for each contact and each frequency band of interest. We then tested across all contact-pairs within a given ROI (dmPFC or STN) and for each frequency (theta or high-gamma) and for each trial-type (switch or non-switch) whether the regression estimates (slopes) were reliably different from the chance level (two-tailed one-sample t-tests).

Hierarchical drift diffusion models (HDDM)

HDDM with behavioral data. We fit participant's choices with the drift diffusion model considering the upper and lower boundaries are accounted for hit and error trials. We used Python 2.7 and the HDDM³⁰ module to launch all the model fitting. The first step of our HDDM analysis was to model behavioral results. For this purpose, we considered each parameter fluctuation as a possible explanation of the difference of reaction time and error rate observed between switch and non-switch trials. Once the DIC compares, it results that a fluctuation of the initial level of evidence is the best explanation of RT and ER difference. We express it as follow:

$$z = k_0 + k_1 \text{Task (M1) also described with patsy as : } z \sim 1 + C(\text{condition}) \quad (1)$$

when considering a , t , v constant and independent to task condition (switch or non-switch). Here, k_0 and k_1 are coefficients with k_0 equivalent to the intercept and Task refers to the trial type (0 for non-switch, 1 for

switch). Markov chain Monte Carlo sampling was used for Bayesian approximation of the posterior distribution of model parameters by generating 150,000 samples of which the first 100,000 samples were discarded. Thus we obtain a final chain of 50,000 samples and the posterior of all parameters $[a, t, v, z]$ for each subject and for the group. We checked traces of model parameters, their autocorrelation and the model convergence. To test the significance of parameter differences across conditions (i.e., switch vs. non-switch trials), we computed the proportion of samples in which starting point for switch condition was lower than in non-switch condition over the full simulated posterior distribution. Model parameters were judged significantly different from 0 if $\geq 95\%$ of the samples drawn from the full posterior were below/above from zero. Even though such posterior probabilities are distinct from classical parametric statistics (e.g., in a t-test), they can be interpreted in a similar manner^{28,30}.

HDDM with neural data. In a second step, we analyzed neuro-computational effects during the task. We consider the neural fluctuation through trials and add this information to the previous model. We use the mean value of the power during a specific time window as reflecting the neural state of each trial. We define the initial level of evidence as follow:

$$z = k_0 + k_1 \theta_{\left(\frac{\text{ms}}{\text{Hz}}\right)} \text{ (M2) also described with patsy as : } z \sim 1 + \theta : C(\text{condition}) \quad (2)$$

where k_0 and k_1 are coefficients, θ the mean power for the frequency range^{5–10} Hz. k_0 acts as an intercept and represents the reference value of the initial level of evidence calculated for all the trials. k_1 is the coefficient of regression associated with θ which is assigned a certain distribution for switch trials and another for the non-switch trials. Here, we generated a chain of 150,000 samples of which the first 100 000 samples were discarded. Thus, we obtain a chain of 50,000 samples. This time, we calculated the posterior of all parameters for the group only to facilitate model convergence. Indeed, we widely increase the complexity of the model by adding the neural coefficient of regression. To calculate θ value, we used the theta single trial estimation means during the significant increase time window (Figs. 2c and 4c). It gives us a theta estimation per trials which we normalized by subtracting the mean value and dividing by the standard deviation per subject and per trial type^{39,42}.

Reporting summary

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

Data availability

The behavior and neural data generated in this study are available from the Figshare database [source data: <https://doi.org/10.6084/m9.figshare.26363740>].

Code availability

The custom codes used to generate the figures and statistics are available from the lead contact (JB) upon request.

References

1. Monsell, S. Task switching. *Trends Cogn. Sci.* **7**, 134–140 (2003).
2. Sakai, K. Task Set and Prefrontal Cortex. *Annu. Rev. Neurosci.* **31**, 219–245 (2008).
3. Isoda, M. & Hikosaka, O. Switching from automatic to controlled action by monkey medial frontal cortex. *Nat. Neurosci.* **10**, 240–248 (2007).
4. Isoda, M. & Hikosaka, O. Role for Subthalamic Nucleus Neurons in Switching from Automatic to Controlled Eye Movement. *J. Neurosci.* **28**, 7209–7218 (2008).

5. Greenhouse, I., Gould, S., Houser, M. & Aron, A. R. Stimulation of contacts in ventral but not dorsal subthalamic nucleus normalizes response switching in Parkinson's disease. *Neuropsychologia* **51**, 1302–1309 (2013).
6. Pasquereau, B. & Turner, R. S. A selective role for ventromedial subthalamic nucleus in inhibitory control. *eLife* **6**, e31627 (2017).
7. Aron, A. R. & Poldrack, R. A. Cortical and Subcortical Contributions to Stop Signal Response Inhibition: Role of the Subthalamic Nucleus. *J. Neurosci.* **26**, 2424–2433 (2006).
8. Herz, D. M., Zavala, B. A., Bogacz, R. & Brown, P. Neural Correlates of Decision Thresholds in the Human Subthalamic Nucleus. *Curr. Biol.* **26**, 916–920 (2016).
9. Forstmann, B. U. et al. Striatum and pre-SMA facilitate decision-making under time pressure. *Proc. Natl Acad. Sci. USA.* **105**, 17538–17542 (2008).
10. Zavala, B. et al. Cognitive control involves theta power within trials and beta power across trials in the prefrontal-subthalamic network. *Brain* **141**, 3361–3376 (2018).
11. Sheth, S. A. et al. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* **488**, 218–221 (2012).
12. Zavala, B. et al. Human Subthalamic Nucleus Theta and Beta Oscillations Entrain Neuronal Firing During Sensorimotor Conflict. *Cereb. Cortex* **27**, 496–508 (2017).
13. Kelley, R. et al. A human prefrontal-subthalamic circuit for cognitive control. *Brain* **141**, 205–216 (2018).
14. Smith, E. H. et al. Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nat. Neurosci.* **22**, 1883–1891 (2019).
15. Kühn, A. A. et al. Event-related beta desynchronization in human subthalamic nucleus correlates with motor performance. *Brain* **127**, 735–746 (2004).
16. Swann, N. et al. Intracranial EEG Reveals a Time- and Frequency-Specific Role for the Right Inferior Frontal Gyrus and Primary Motor Cortex in Stopping Initiated Responses. *J. Neurosci.* **29**, 12675–12685 (2009).
17. Swann, N. C. et al. Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: Electrophysiological responses and functional and structural connectivity. *NeuroImage* **59**, 2860–2870 (2012).
18. Bastin, J. et al. Inhibitory control and error monitoring by human subthalamic neurons. *Transl. Psychiatry* **4**, e439 (2014).
19. Benis, D. et al. Subthalamic nucleus activity dissociates proactive and reactive inhibition in patients with Parkinson's disease. *NeuroImage* **91**, 273–281 (2014).
20. Benis, D. et al. Response inhibition rapidly increases single-neuron responses in the subthalamic nucleus of patients with Parkinson's disease. *Cortex* **84**, 111–123 (2016).
21. Wessel, J. R., Waller, D. A. & Greenlee, J. D. Non-selective inhibition of inappropriate motor-tendencies during response-conflict by a fronto-subthalamic mechanism. *eLife* **8**, e42959 (2019).
22. Mosher, C. P., Mamelak, A. N., Malekmohammadi, M., Pouratian, N. & Rutishauser, U. Distinct roles of dorsal and ventral subthalamic neurons in action selection and cancellation. *Neuron* **109**, 869–881.e6 (2021).
23. Cooper, P. S. et al. Frontal theta predicts specific cognitive control-induced behavioural changes beyond general reaction time slowing. *NeuroImage* **189**, 130–140 (2019).
24. Proskovec, A. L., Wiesman, A. I. & Wilson, T. W. The strength of alpha and gamma oscillations predicts behavioral switch costs. *NeuroImage* **188**, 274–281 (2019).
25. McKewen, M. et al. Dissociable theta networks underlie the switch and mixing costs during TASK SWITCHING. *Hum. Brain Mapp.* **42**, 4643–4657 (2021).
26. Schmitz, F. & Voss, A. Decomposing task-switching costs with the diffusion model. *J. Exp. Psychol.: Hum. Percept. Perform.* **38**, 222–250 (2012).
27. Sinha, N., Brown, J. T. G. & Carpenter, R. H. S. Task Switching as a Two-Stage Decision Process. *J. Neurophysiol.* **95**, 3146–3153 (2006).
28. Herz, D. M. et al. Distinct mechanisms mediate speed-accuracy adjustments in cortico-subthalamic networks. *eLife* **6**, e21481 (2017).
29. ter Wal, M. et al. Human stereoEEG recordings reveal network dynamics of decision-making in a rule-switching task. *Nat. Commun.* **11**, 3075 (2020).
30. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front. Neuroinform.* **7**, (2013).
31. Pedersen, M. L., Frank, M. J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychon. Bull. Rev.* **24**, 1234–1251 (2017).
32. Lisman, J. E. & Jensen, O. The Theta-Gamma Neural Code. *Neuron* **77**, 1002–1016 (2013).
33. Tang, H. et al. Cascade of neural processing orchestrates cognitive control in human frontal cortex. *eLife* **5**, e12352 (2016).
34. Fu, Z. et al. Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. *Neuron* **101**, 165–177.e5 (2019).
35. Bastin, J. et al. Direct Recordings from Human Anterior Insula Reveal its Leading Role within the Error-Monitoring Network. *Cereb. Cortex* **27**, 1545–1557 (2017).
36. Cavanagh, J. F. et al. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat. Neurosci.* **14**, 1462–1467 (2011).
37. Mansfield, E. L., Karayanidis, F., Jamadar, S., Heathcote, A. & Forstmann, B. U. Adjustments of Response Threshold during Task Switching: A Model-Based Functional Magnetic Resonance Imaging Study. *J. Neurosci.* **31**, 14688–14692 (2011).
38. Herz, D. M. et al. Mechanisms Underlying Decision-Making as Revealed by Deep-Brain Stimulation in Patients with Parkinson's Disease. *Curr. Biol.* **28**, 1169–1178.e6 (2018).
39. Green, N. et al. Reduction of Influence of Task Difficulty on Perceptual Decision Making by STN Deep Brain Stimulation. *Curr. Biol.* **23**, 1681–1684 (2013).
40. Ratcliff, R. & McKoon, G. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Comput.* **20**, 873–922 (2008).
41. Wiecki, T. V. & Frank, M. J. A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Rev.* **120**, 329–355 (2013).
42. Tinkhauser, G. et al. The modulatory effect of adaptive deep brain stimulation on beta bursts in Parkinson's disease. *Brain* **140**, 1053–1067 (2017).
43. Mallet, L. et al. Subthalamic Nucleus Stimulation in Severe Obsessive-Compulsive Disorder. *N. Engl. J. Med.* **359**, 2121–2134 (2008).
44. Yelnik, J. et al. A three-dimensional, histological and deformable atlas of the human basal ganglia. I. Atlas construction based on immunohistochemical and MRI data. *NeuroImage* **34**, 618–638 (2007).
45. Bardin, E. et al. A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease: Clinical article. *JNS* **110**, 208–219 (2009).
46. Auzias, G., Coulon, O. & Brovelli, A. MarsAtlas: A cortical parcellation atlas for functional mapping: MarsAtlas. *Hum. Brain Mapp.* **37**, 1573–1592 (2016).

47. Lopez-Persem, A. et al. Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nat. Neurosci.* **23**, 664–675 (2020).
48. Cecchi, R. et al. Intracerebral mechanisms explaining the impact of incidental feedback on mood state and risky choice. *eLife* **11**, e72440 (2022).
49. Domenech, P., Rheims, S. & Koechlin, E. Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* **369**, eabb0184 (2020).
50. Collomb-Clerc, A. et al. Human thalamic low-frequency oscillations correlate with expected value and outcomes during reinforcement learning. *Nat. Commun.* **14**, 6534 (2023).

Acknowledgements

This work benefited from University Grenoble Alpes 'Investissements d'Avenir' program (ANR-17-CE37-0018; ANR-18-CE28-0016; ANR-22-CE17-0057; ANR-23-CE17-0070) awarded to JB; PK and MP and from an 'Agence Nationale de la Recherche' young investigator grant awarded to PD (ANR-21-CE37-0004-01). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank all the patients who participated to the study and Grenoble Teaching Hospital epilepsy, psychiatry and neurosurgery clinical teams for their support. We thank Dr. E. Boorman and N. Kolling for their helpful comments on an early version of this manuscript and Dr. M. Laubach for his suggestion about the Goldilocks' expression.

Author contributions

J.B. designed the experiment and collected the data. J.B., P.D., J.Y., S.F.V. and M.L. provided preprocessing scripts and anatomical location of dmPFC/STN sites. M.L., J.B. and P.D. performed the data analysis. M.P., P.K., and S.C. did the clinical investigation. M.L., P.D. and J.B. wrote the manuscript. All the authors discussed the results and commented on the manuscript.

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/s41467-024-52290-w>.

Correspondence and requests for materials should be addressed to Philippe Domenech or Julien Bastin.

Peer review information *Nature Communications* thanks Ishita Basu, Vincenzo Fiore, Alekhya Mandali and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024