HHS Public Access

Author manuscript

Nat Biomed Eng. Author manuscript; available in PMC 2023 April 29.

Published in final edited form as:

Nat Biomed Eng. 2023 April ; 7(4): 576–588. doi:10.1038/s41551-021-00804-y.

Closed-loop enhancement and neural decoding of cognitive control in humans

Ishita Basu[1], **Ali Yousefi**[1,2], **Britni Crocker**[3], **Rina Zelmann**[3], **Angelique C Paulk**[3], **Noam Peled**[4], **Kristen K Ellard**[1], **Daniel S Weisholtz**[5], **G. Rees Cosgrove**[6], **Thilo Deckersbach**[1], **Uri T Eden**[7], **Emad N Eskandar**[8], **Darin D Dougherty**[1], **Sydney S Cash**[3,*], **Alik S Widge**[1,9,*,+]

[1]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA

[2]Present Address: Departments of Computer Science and Neuroscience, Worcester Polytechnic Institute, Worcester, MA

[3]Department of Neurology, Massachusetts General Hospital, Boston, MA

[4]Dept. of Radiology, MGH/HST Martinos Center for Biomedical Imaging and Harvard Medical School, Boston, MA

[5]Department of Neurology, Brigham & Womens Hospital, Boston, MA

[6]Department of Neurological Surgery, Brigham & Womens Hospital, Boston, MA

[7]Department of Mathematics and Statistics, Boston University, Boston, MA

[8]Department of Neurological Surgery, Massachusetts General Hospital, Boston, MA; Current address: Department of Neurological Surgery, Albert Einstein College of Medicine, Bronx, NY

[9]Current address: Department of Psychiatry, University of Minnesota, Minneapolis, MN

[+]**Correspondence and requests for materials** should be addressed to A.S.W. awidge@umn.edu.

[*]:Equal contribution as senior authors.

Author contributions

ASW, DDD, ENE, and SSC designed the study. IB, AY, BC, RZ, and UTE designed key software and tools required for data collection. KKE and TD selected the psychometric scales administered to participants and provided unpublished data related to norming of those questionnaires. ENE and GRC performed all surgical procedures. ASW, IB, BC, RZ, ACP, SSC, and DSW collected data with participants during acute seizure monitoring. ASW, IB, AY, ACP, and NP analyzed data. IB and ASW wrote the paper with substantial inputs from AY, RZ, ACP, and SSC. All authors had opportunities for critical input into and revision of the submitted manuscript, and approved its submission.

Reporting Summary.

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

Code availability

Analysis code is available at https://github.com/tne-lab/MSIT-Nature-Biomedical-Engineering. The closed-loop neurostimulation system has been released as open-source code and documented[47], and the neural decoding and state-space modelling engines have similarly been released for open download (https://github.com/TRANSFORM-DBS/Encoder-Decoder-Paper and https://github.com/Eden-Kramer-Lab/COMPASS).

Competing interests

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41551-021-00804-y.

Abstract

Deficits in cognitive control — that is, in the ability to withhold a default prepotent response in favour of a more adaptive choice — are common in depression, anxiety, addiction and in other mental disorders. Here, we report proof-of-concept evidence that, in participants undergoing intracranial epilepsy monitoring, closed-loop direct stimulation of the internal capsule or striatum, especially the dorsal sites, enhances the participants' cognitive control during a conflict task. We also show that closed-loop stimulation upon the detection of lapses in cognitive control produced larger behavioural changes than open-loop stimulation, and that task performance for single trials can be directly decoded from the activity of a small number of electrodes, via neural features that are compatible with existing closed-loop brain implants. Closed-loop enhancement of cognitive control might remediate underlying cognitive deficits and aid the treatment of severe mental disorders.

Mental disorders are a leading source of medical economic burden[1]. Current therapies do not target the cause of these disorders: dysfunctional communication in distributed brain circuits[2,3]. Electrical brain stimulation can effectively modulate such circuits[4], but clinical neurostimulation trials in mental disorders have had mixed results[5,6], for two reasons. First, spatially targeting a circuit may not be enough. Symptom-related activity is only occasionally present, and effective therapies may require temporal specificity: a closed feedback loop that detects pathological activity and intervenes to restore more efficient function[7-9]. Second, mental illness is heterogeneous. A label such as "depression" likely describes several different brain disorders[8,10]. Rather than trying to detect or treat ill-specified symptoms such as mood[11,12], it may be more effective to target objective, rigorously measurable constructs, each representing a different component of cognitive or emotional processing[8,10,13]. In this dimensional approach, simultaneous impairments in multiple cognitive constructs act as "ingredients" to produce the clinical manifestation of disease[2,10]. An advantage of the dimensional approach is that individual constructs can be objectively measured, e.g. through performance on standardized tasks. That performance may be more easily reducible to specific brain circuits and stimulation targets.

One particularly promising construct is cognitive control, the ability to flexibly alter strategies or response styles as goals change[14-18]. First, cognitive control deficits are common across mental disorders, including depression, addiction, schizophrenia, and obsessive-compulsive disorder[19]. They are particularly relevant for mood and anxiety disorders, where patients are often unable to disengage from habitual, distress-driven behavior. Second, key aspects of control are readily measurable, using cognitive conflict tasks. Slower response times in these tasks signify difficulty exerting control to overcome response conflict, to the point that performance on such tasks is the most commonly used clinical and laboratory metric of cognitive control[19,20]. Third, control and conflict have a likely circuit substrate. Action selection in the face of competing options is well linked to interactions between multiple sub-regions of prefrontal cortex (PFC) and their counterparts in the striatum[21-23]. They similarly have a physiologic marker: many different conflict tasks evoke robust electrophysiologic signatures, namely theta (4-8 Hz) oscillations in PFC[24-26] and low-frequency synchrony among frontal and subcortical structures[19,27-29]. Fourth, that circuit substrate is well suited for manipulation through brain stimulation. PFC,

striatum, and basal ganglia are densely interconnected by white matter tracts that run largely in the internal capsule[30,31]. Those tracts follow relatively stereotyped trajectories[30-33], meaning that stimulation at a given capsular site will likely engage similar circuitry across individuals. To that point, deep brain stimulation (DBS) of the internal capsule, DBS of subthalamic nucleus, and transcranial current stimulation of the lateral PFC all improve performance on cognitive control tasks[18,34,35]. In other words, the scientific basis exists for direct remediation of cognitive control deficits, if these disparate findings can be integrated.

Three barriers have prevented that integration. First, there are no extant techniques for quantifying cognitive control performance in real time. Response time on conflict tasks, the standard metric, is also influenced by stochastic noise and often by changes in conflict level between subsequent task trials[36]. Second, there are no proven ways to rapidly alter that performance, i.e. to respond to and cancel out momentary lapses. Third, although there are known physiologic signatures of cognitive control when averaged across time, there are no such signatures for moment-to-moment fluctuation in control. Here, we demonstrate methods to overcome those barriers, through a proof of concept in participants undergoing stereotaxic electrode monitoring for epilepsy (Fig. 1A). We developed a method for quantifying conflict task performance at the single-trial level, while regressing out sources of undesired variability. This method is grounded in a state-space or latent-variable formalism, which generalizes the Kalman filtering approach that has been successful in other neural interface applications[37,38]. We showed that brief internal capsule stimulation rapidly enhances cognitive control. That enhancement was visible in raw task performance, in our latent variables, and in PFC theta oscillations. We then integrated those findings into a closed-loop controller that stimulated in response to lapses in cognitive control. Closed-loop stimulation produced larger performance changes than a corresponding open-loop paradigm. Finally, we showed that task performance, at the scale of single trials, can be decoded entirely from brain activity. Together, these findings provide proof of concept for a closed-loop system for treating cognitive control deficits.

Results

Cognitive control and theta power enhancement through internal capsule stimulation

Participants performed a cognitive control task (the Multi-Source Interference Task (MSIT), Fig. 1B) while undergoing invasive multi-electrode monitoring. During task performance, we electrically stimulated in the internal capsule, at similar sites across participants (Fig. 1C). We collected 8,790 trials across 176 blocks from 21 participants – 12 without brain stimulation, 5 with both unstimulated and open-loop stimulation sessions, 1 with only open-loop stimulation, and 3 with unstimulated and closed-loop stimulation sessions. Dropping incorrect responses and non-responses excluded 345 trials (3.92% of total; 8,445 trials retained in analysis). In open-loop experiments, a random 50% of trials received brief, task-linked stimulation (Fig. 1D).

MSIT engaged cognitive control: participants were 216 ms slower on high-conflict trials ($N=21$, $p<0.001$, $t=33.62$, Wald test on GLM coefficient, Fig. 2A). Conflict increased task-related theta power in the posterior cingulate ($p<0.02$, $t=3.24$) and dorsolateral prefrontal cortex ($p<0.001$, $t=4.31$) (Fig. 2B). Capsular stimulation enhanced both cognitive control

and its electrophysiologic correlates, with dorsal sites showing stronger effects. Right dorsal ($p<0.001$, $t=-4.28$), left dorsal ($p<0.01$, $t=-2.65$) and right ventral ($p<0.05$, $t=-2.64$) capsular stimulation all significantly decreased reaction time (RT, Fig. 2C, Supplementary Table 3a) without impairing accuracy (Extended Data Fig. 2A). RTs under dorsal stimulation were faster than with ventral stimulation on both sides, with right dorsal being the overall most effective (Supplementary Table 3a). These findings mirror the capsular topography, where more dorsal sites are enriched in fibers originating in dorsal PFC[30,31], which in turn is associated with cognitive control during conflict tasks[19,20,28,39]. Consistent with that topography, stimulation in dorsal capsule sites outside of task performance propagated mainly to ACC and DLPFC (Extended Data Fig. 3). RT improvements were not explained by practice or regression to the mean; RT in the final non-stimulated (NS1) block was significantly higher than the initial block (Supplementary Fig. 1A).

There was no evidence for an interaction between stimulation and conflict level (AIC: −449.27 for a model without an interaction term vs. −445.72 with interaction). To assess stimulation's effect on theta, we analyzed artifact-free trials interleaved within stimulated blocks (NS2, Fig. 1D) and compared these to blocks without stimulation (NS1). Left dorsal and right ventral capsular stimulation significantly increased theta power in NS2 compared to NS1 trials (LD: $p= 0.0428$, RV: $p=0.0006$, FDR corrected, Fig. 2D-E, Supplementary Table 3b). Right dorsal capsular stimulation also increased theta but did not reach significance ($p= 0.1733$, FDR corrected). Theta increases were present in many PFC channels and were specific to behaviorally effective stimulation (Supplementary Fig. 2). Theta increases also could not be explained as regression to the mean or practice effects, as theta power attenuated over the experimental session in the absence of stimulation (Supplementary Fig. 1B).

Closed-loop stimulation based on a state-space model efficiently enhances cognitive control

We next sought to quantify task performance and capsular stimulation effects at a trial-to-trial level. We achieved this with a state-space latent variable model (Fig. 3A-B). This model assumes that each trial's RT can be modeled as the combination of a baseline or expected RT for all trial types (x_{base}), a specific slowing on high conflict trials ($x_{conflict}$), and a log-Gaussian noise process. This formulation allows rapid tracking of stimulation-induced changes (see Methods). We verified that the two-process model captured the majority of variance in the RT data (Supplementary Fig. 3), that it converged in all participants (Supplementary Fig. 4), and that response accuracy did not carry additional information (Supplementary Fig. 5). Stimulation in the dorsal capsule improved overall performance (x_{base}, Fig. 3C) and reduced conflict effects (Fig. 3D). Right dorsal capsular stimulation again had the largest effects. Ventral stimulation significantly reduced $x_{conflict}$, but not x_{base}. The observed differences could not be explained by block-to-block carry-over or other persistent effects, and x_{base} rapidly increased once stimulation ceased (Supplementary Fig. 6).

We applied capsular stimulation under closed-loop control in 3 further participants. We estimated x_{base} in real time and triggered stimulation during control lapses, i.e. when

x_{base} increased beyond an experimenter-determined threshold (Fig. 4A). As predicted, conditioning stimulation on x_{base} specifically improved that variable (Fig. 4B) without enhancing $x_{conflict}$ (Fig. 4C). Closed-loop stimulation was more effective than open-loop. Stimulation of the right ventral capsule, which did not have significant effects in open-loop tests (Fig. 3C), now significantly reduced x_{base} ($p<0.01$, permutation test, Fig. 4B). At both dorsal stimulation sites, closed-loop stimulation reduced x_{base} significantly more than open-loop stimulation ($p<0.001$, permutation test, Fig. 4B). There were again no accuracy effects (Extended Data Fig. 2B). Closed-loop stimulation's effect was manifest in raw RT data for right dorsal capsule stimulation ($p<0.001$, permutation test, Extended Data Fig. 4). The effects cannot be explained by regression to the mean (Supplementary Fig. 7). The closed-loop algorithm did stimulate more often that its open-loop counterpart (22-29 stimulated trials per block), but this alone did not explain the increased behavior effect. Closed-loop stimulation also appeared more efficient than open-loop in terms of performance gain for the applied energy. It produced a greater change in x_{base} per stimulation in the right ventral and dorsal capsule (Fig. 4D), although this did not reach the pre-determined significance threshold (RV: $p=0.207$, RD: $p=0.293$).

A few participants reported that improving objective task performance also improved their subjective well-being. Although participants could not directly identify when stimulation was on or off, they perceived when they were performing more fluidly (Supplementary Table 4). Two participants who had previously reported difficulty with effortful self-control noted relief of their usual anxiety (Supplementary Table 5). No participant reported negative emotional effects during any stimulation experiment.

Neural decoding of cognitive states for closed-loop control

To demonstrate that cognitive control lapses could be remediated outside of a controlled, structured task setting, we developed decoders to read out cognitive control from LFP. For each participant, we estimated an encoding model (Fig. 5A) to map cognitive states to LFP power. State variables were linearly related to neural features (Supplementary Fig. 8), at a level exceeding chance (Supplementary Fig. 9). The confidence intervals of cognitive states decoded from LFP and estimated from behavior largely overlapped (Fig. 5B, x_{base}: 84.02 ± 15.8% overlap, $x_{conflict}$: 83.17 ± 16.3% overlap). Decoding used relatively few power features in each participant (Fig. 5C; 11.75 ± 6.63 features for x_{base} and 11.27 ± 6.74 for $x_{conflict}$). Decoding weighted brain regions commonly implicated in cognitive control. x_{base} was encoded primarily in dlPFC (8-30, 130-200 Hz), vlPFC (15-30, 65-110 Hz), and temporal cortex (multiple bands). $x_{conflict}$ was similarly encoded in dlPFC (most bands), vlPFC (8-55 Hz), temporal cortex (broadband), and amygdala (8-55 Hz).

Decoding was also possible during intermittent brain stimulation (x_{base}: 83.15 ± 9.04%, $x_{conflict}$: 84.43 ± 9.6% of trials overlapping the confidence interval of the behavioral estimate). Stimulation did not meaningfully alter the number of features needed for decoding (x_{base}: 11.11 ± 4.73 vs. 11.75 ± 6.63 features; $x_{conflict}$: 12.55 ± 4.24 vs. 11.27 ± 6.74 features; all $p > 0.5$, unpaired t-test). It did, however, decrease the number of cortical regions that encoded either x_{base} or $x_{conflict}$ (Fig. 5D). This encoding may have transferred to the dorsal striatum (caudate), which showed increased encoding across

frequency bands, although this did not reach our pre-specified significance threshold for x_{base} (x_{base}t=−0.3980, p=0.6908, $x_{conflict}$: t= 5.3947, p<0.001, paired t-test between encoder-model coefficients on NS1 and NS2 trials across participants).

Discussion

Cognitive control is impaired in numerous mental disorders[8,13,19,40]. We augmented one aspect of human cognitive control, performance on a cognitive conflict task, by intermittent closed-loop stimulation of the internal capsule. The effects were detectable in both manifest data (raw reaction times) and derived variables. This brief intervention produced positive emotional effects in a small subset of patients who reported relevant pre-existing impairments. This may suggest a future utility of cognitive control as a psychiatric treatment target, independent of a patient's diagnostic label. To that point, we previously showed that improvements in similar aspects of cognitive control may be a mechanism of action of clinical deep brain stimulation[18]. Further, we could separate and alter components of cognitive control. We enhanced baseline performance (expected RT on all trial types) without driving the conflict response (specific slowing and expected RT on high-conflict trials) in the same direction. This suggests that these two processes could be targeted separately, e.g. if only one was impaired in a hypothetical future patient. Both states could be decoded with 11 LFP spectral features per participant, from a mean of 6 brain regions. Existing human-grade implants may be able to implement these or similar decoders[41].

Our result both converge with and diverge from past studies of cognitive control. Anatomically, we found the largest effects from stimulation in the dorsal capsule. This part of the capsule contains corticofugal axons originating mainly in lateral PFC and dorsomedial PFC or cingulate[31]. We observed that lateral PFC in particular showed a theta power enhancement during behaviorally effective stimulation. This fits with the central role that those PFC regions likely play in cognitive control. We obtained our largest effects by stimulating in areas where we could modulate the relevant PFC sub-regions through retrograde activation. Our physiologic results support that retrograde model – effective stimulation increased PFC theta power, the most commonly reported correlate of control. Surprisingly, however, our decoders did not strongly select theta-band features when given the full range of LFP activity to consider. This may be because prior studies linking cognitive control to theta used EEG[24] and were analyzed on a trial-averaged basis. The mean theta effect across trials, although robust, may not relate to the trial-to-trial performance. Further, LFP signals from a few millimeters of tissue may not correlate well with EEG, which records synchronized signals from broader swaths of cortex. Additionally, state variables were strongly encoded outside the canonical PFC-cingulate cognitive control network[19,39], particularly in temporal lobe. Hippocampus and amygdala are important in human anxiety-like behavior[42] and integrate signals from multiple prefrontal regions[43]. These structures may play a greater role in cognitive control than has been recognized.

Effective stimulation shifted encoding from frontotemporal cortex to the dorsal striatum or caudate. This is consistent with models of cortico-striatal function that consider the dorso-medial striatum, including the caudate, to be part of a loop-structured circuit serving flexibility and exploration[19,21]. That encoding may also explain why dorsal capsular

stimulation was more behaviorally effective, as the stimulation field likely overlapped both caudate and incoming fibers from dorsal PFC. It is less clear why stimulation slightly decreased the overall encoding of control-related variables, yet improved task performance. One possibility is the hypothesized link between cognitive control and anxiety[44]. Humans may normally exert more cognitive control than we need to, i.e. over-weighting risk. Our performance enhancement could be explained as a reduction of that “excess” control. This is concordant with the two participants who reported that stimulation allowed them to shift attention away from anxious, self-focused processing. We note that this is also a very different subjective effect than reports from clinical DBS of the capsule or striatum, which usually causes euphoria[45,46]. It seems unlikely that such positive mood changes drove faster RTs, because the majority of participants reported no subjective mood alteration.

Because participants had seizure vulnerability, we were limited to a single stimulation session between the completion of clinical monitoring and electrode removal. We could not retest the same stimulation multiple times in the same participant, nor could we attempt a variety of setpoints to determine the limits of this closed-loop approach. We were limited to intermittent stimulation to reduce seizure risk, and could not directly test this intermittent paradigm against the continuous DBS of prior studies. Electrodes were placed along individualized trajectories, and their location within the capsule varied substantially (Supplementary Fig. 10). In this context, the consistent effect across participants is likely explained by the relatively consistent topographic structure of the capsule across individuals and species[30,31]. Electrodes used only for recording had a similar limitation. Demonstration of a consistent, repeatable within-subject effect would be critical for turning these results into a therapy. The nearly identical effects between this study and our prior work with psychiatric DBS patients[18] do suggest that the effect is robust. Further, because there is known variability of fiber placement within the capsule despite the topographic map[32], we might expect much larger effects in a clinical scenario where we could precisely optimize each electrode’s placement to engage a target of interest.

We enhanced x_{base}, which reflects overall attentional focus. $x_{conflict}$ corresponds to the more immediate effect of conflict and the difficulty of executing control. In a clinical setting, either might be disrupted, and we may need to apply closed-loop control to both simultaneously. Here, when we controlled x_{base}, $x_{conflict}$ significantly increased. These two states are not inherently anti-correlated, because both were reduced by open-loop stimulation (contrast Fig. 3C-D against Fig. 4B-C), but it remains to be shown that x_{base} can be altered without effects on other cognitive variables. Similarly, our two-state model follows one specific theory, where control is allocated reactively in response to decision conflict. There are other theoretical models of cognitive control measurement and allocation[39], and it would be important to explore how our stimulation approaches do or do not alter variables in those frameworks. For instance, alternate frameworks place a heavy emphasis on a participant’s motivation to exert control, sometimes framed as expected value. Our results could be explained by an increase in that expected value, and dissecting such questions will require more sophisticated tasks. Finally, decoding might perform better on network-level measures rather than the LFP power features we used, based on recent studies linking connectivity metrics to cognitive control[19,27-29]. At present, connectivity is difficult to estimate in real time or on a therapeutic device, although approaches are emerging[47] .

Closely related, both this study and our prior work operationalized cognitive control through the MSIT, which is only one of several tasks commonly used to measure control. This raises the question of whether our results will generalize to other cognitive control metrics. Three factors suggest they should. First, there were substantial differences in task and stimulation details between this and our prior study, but consistent behavioral effects. Second, MSIT is primarily a combination of two other paradigms, the Simon and Flanker tasks[48]. Prior studies that separated these two conflict types showed consistent neural effects between them[36,49], suggesting that the same circuitry is involved regardless of the specific stimuli used to evoke conflict. Third, and closely related, a wide variety of cognitive conflict tasks consistently evoke theta-band frontal oscillations, implying a common mechanism of cognitive control across tasks[19,24,25,29,44,49]. We observed and modulated those oscillations, suggesting that we engaged circuits that are related to cognitive control generally, not to our specific task. These are reasons for hope, but definitive demonstration that this approach applies to the broadest definitions of cognitive control will require testing across multiple tasks, settings, and possibly species. There is an open question of whether modulating this task-related activity can improve cognitive control in more “real world” situations, which will need to be tested as part of future clinical studies.

Translating these findings to practical clinical use will require overcoming both technical and ethical hurdles. Technically, there is substantial interest in the idea of cognitive control as a cross-cutting symptom of numerous psychiatric disorders[19,40,50], and thought leaders in psychiatry are urging the development of treatments that directly target decisional or cognitive dysfunctions[51,52]. Another group recently reported closed-loop control of impulsivity, a close cousin of cognitive control[53], and has advanced that approach into clinical trials[54]. We and others have argued that cognitive remediation is a mechanism of multiple existing neurostimulation therapies, but is not recognized as such[8,35,55]. There are not, however, well-established diagnostic and monitoring scales for cognitive control or most other cognitive constructs. There are promising first steps[40,56], but establishing clinically valid assessments would be important. It would similarly be both important and interesting to understand how these results might change with a different task that assesses different aspects of cognitive control, e.g. a more explicit extra-dimensional set shift[35]. Given that control depends on a distributed fronto-striatal network[19,57], there are likely multiple access points into that network. Broader explorations, including in animal models[53], may clarify which stimulation targets are most useful in which situations.

Based on two participants reporting subjective well-being during our experiments, it seems possible that improved control might be directly reflected in standard mood or anxiety scales, but this would be a very noisy metric. Objective measures of cognitive control would also be important to ensure that the first use of such a technology rests on strong ethical grounds. Patients have a strong interest in closed-loop neurotechnologies[58], but there are broad societal concerns about altering personalities or authentic selves through such technologies[59,60]. Given that we could improve task performance in patients with no overt cognitive impairment, these results also raise concern for attempts at cognitive enhancement in healthy humans. Given the numerous concerns surrounding potential enhancement[61], there should be careful ethical consultation before this approach is used outside of a rigorous and restricted research setting.

In summary, we have developed methods for real-time monitoring of human cognitive control-task performance, detection of lapses, and closed-loop remediation of those lapses. We produced potentially useful cognitive and emotional effects with far less energy than conventional deep brain stimulation. The same framework could also be applied to other cognitive or emotional problems, e.g. monitoring and enhancement of learning[62,63] or emotion dysregulation[64,65]. Although substantial technology gaps remain before these results can be directly applied in the clinic, and the evidence base needs to be built for cognition as a primary focus of treatment, our results could be the basis of a highly specific approach for intervention in human neuropsychiatric disease. In theory, and with further development, this approach could address a wider range of disorders than existing neurostimulation.

Methods

Experimental Design

Twenty-one participants (age range: 19-57, mean age: 35, female: 12/21, left handed: 5/21) with long-standing pharmaco-resistant complex partial seizures were voluntarily enrolled after fully informed consent according to NIH and Army HRPO guidelines. Consent was obtained by a member of the study staff who was not the participant's primary clinician. Study procedures were conducted while participants underwent inpatient intracranial monitoring for seizure localization at Massachusetts General Hospital or Brigham & Women's Hospital. The electrode implants were solely made on clinical grounds and not tailored for research purposes. Informed consent and all other ethical aspects of the study were approved and monitored by the Partners Institutional Review Board. Participants were informed that their involvement in the study would not alter their clinical treatment in any way, and that they could withdraw at any time without jeopardizing their clinical care. This was an exhaustive sample of all participants who were available, consented to research, and had electrodes in brain regions of interest. There was no pre-planned power calculation. Data collection ceased at the end of a continuous 4-year period once study funding lapsed. Detailed clinical information on participants is in Supplementary Table 1.

The core hypothesis, that internal capsule stimulation would enhance cognitive control (shorten response times in a cognitive control task without altering error rates) was pre-specified based on our prior work[18]. The analyses of behavioral and electrophysiological data, including the state-space and neural decoding modeling described below, were similarly pre-planned. The analyses described up through main text Fig. 2 were specified to replicate the prior study and demonstrate that its effects were robust to changes in study population and stimulation details. Participants were part of a larger multi-modal study of the neural basis of mental illness[13]. They performed multiple other cognitive tasks both before and during their hospital stay and completed self-report scales related to emotional and cognitive difficulties. The list of tasks and scales and our strategies for comparing study participants against a normative database are described in [13]. All are well-validated assessments that have previously been used in large-scale studies. The specific comparison of these scales against participants' experience of subjective improvement was not pre-planned, as we did not know in advance that participants would report these effects.

Since we did not have a pre-specified stopping rule, we performed a *post hoc* sensitivity analysis in G*Power 3.1[66] to verify that we were adequately powered for the pre-specified hypothesis that capsular stimulation would improve cognitive control. The regression coefficients reported in Tables S2 and interpreted as significant in the text correspond to a partial R^2 of 0.0045 for the RT model and 0.01 for the theta-power model. These correspond to effect sizes f^2 of 0.0045 and 0.01, respectively. With the sample counts or degrees of freedom as in those tables, we have over 87% power for the detected effect sizes.

Behavior Paradigm – Multi Source Interference Task (MSIT)

Participants performed the Multi-Source Interference Task (MSIT) [48] with simultaneous recordings of behavior and local field potentials (LFPs) from both cortical and subcortical brain structures. MSIT is a cognitive control task known to induce statistically robust subject-level effects, at both the behavioral and neural level[18,48,67]. These relatively large effect sizes amplified our ability to detect stimulation-induced differences, by increasing task-related behavioral and neural signatures. MSIT trials consisted of three numbers between 0-3, two of which had the same value (Fig. 1A). Participants had to identify, via button press, the identity of the number that was unique, not its position. Each trial contained one of two levels of cognitive interference or conflict. Low conflict or congruent (C) trials had the unique number in a position corresponding to its keyboard button, and flanking stimuli were always ‘0’, which is never a valid response. High conflict, or incongruent trials (I), had the position of the unique number different from the keyboard position, requiring execution of a non-intuitive visuo-motor mapping (Simon effect). On high conflict trials, the non-unique numbers were valid responses (flanker effect). To reduce the formation of response sets, we pseudo-randomized the trial sequence such that more than two trials in a row never shared the same interference level or correct response finger. This forces frequent strategy shifts and increases attention demands, which in turn increase the need to engage or deploy cognitive control. Each participant performed 1-3 sessions of MSIT. Each session consisted of multiple blocks of 32 or 64 trials, with brief rest periods in between blocks. These periods ranged from 0.9 to 57.6 minutes; the median break time was 5.25 minutes and 89.6% were under 10 minutes. During blocks, participants were instructed to keep their first through third fingers of their right hand on the response keys corresponding to the numbers 1-3. They were instructed to be as fast and as accurate as possible. Stimuli were presented for 1.75 seconds, with an inter-trial interval randomly jittered within 2-4 seconds. Stimuli were presented on a computer screen with either Presentation software (Neurobehavioral Systems) or Psychophysics Toolbox[68].

Electrophysiologic Recording

We recorded local field potentials (LFP) from a montage of 8-18 bilaterally implanted depth electrodes (Fig. 1C, left, and Extended Data Fig. 1). The decision to implant electrodes and the number, types, and location of the implantations were all determined on clinical grounds by a team of caregivers independent of this study. Depth electrodes (Ad-tech Medical, Racine, WI, USA, or PMT, Chanhassen, MN, USA) had diameters of 0.8– 1.0 mm and consisted of 8-16 platinum/iridium-contacts, each 1-2.4 mm long. Electrodes were localized by using a volumetric image coregistration procedure. Using Freesurfer scripts (http://surfer.nmr.mgh.harvard.edu), the preoperative T1-weighted MRI (showing the brain

anatomy) was aligned with a postoperative CT (showing electrode locations). Electrode coordinates were manually determined from the CT[69]. The electrodes were then mapped to standard cortical parcels or regions using an automatic, probabilistic labeling algorithm[70]. Intracranial recordings were made using a recording system with a sampling rate of 2 kHz (Neural Signal Processor, Blackrock Microsystems, US). At the time of acquisition, depth recordings were referenced to an EEG electrode placed on skin (either cervical vertebra 2 or Cz).

Open-Loop Capsule Stimulation

We delivered electrical stimulation to either the dorsal or ventral internal capsule and surrounding striatal nuclei (Fig. 1B, right). We stimulated only one site in each block. We compared dorsal and ventral stimulation because this portion of the internal capsule has a well-described dorso-ventral topography[30-32]. More ventral capsule fibers tend to originate in more ventral aspects of PFC, particularly orbitofrontal cortex (OFC). More dorsal fibers, in contrast, tend to originate in dorsolateral and dorsomedial PFC and dorsal cingulate. The latter structures are strongly implicated in cognitive control, particularly during conflict tasks[19-21,39,57]. Thus, we hypothesized that dorsal stimulation would be more effective than ventral in this specific task. We did not have a pre-specified hypothesis regarding left vs. right stimulation, but tested hemispheres separately because our past clinical studies of capsular stimulation suggested lateralized effects[71]. We did not test any other stimulation sites for their effects on MSIT performance, because we had a specific hypothesis regarding the internal capsule. Other sEEG electrodes were stimulated in separate experiments in these same participants, but those are reported elsewhere[64,72,73] and the outcomes of those experiments did not influence the present study.

We varied the order in which we stimulated the different sites (Supplementary Table 2). Stimulation experiments were performed towards the end of each participant's stay, when he/she was back on anti-epileptic medication, to reduce the risk of evoking a seizure. Such a session always started with 1 or 2 blocks of unstimulated trials and ended with an unstimulated block (Fig. 1D). In blocks with stimulation, it occurred on only 50% of the trials. The stimulated trials were chosen pseudo-randomly, with no more than 3 consecutive trials being stimulated. All participants received the same pseudo-random order of stimulated and unstimulated trials. The stimulation was a 600 ms long train of symmetric biphasic (charge balanced) 2-4 mA, 90 μs square pulses at a frequency of 130 Hz. Stimulation was delivered through a neighboring pair of contacts on a single depth electrode (bipolar), with the cathodal (negative) pulse given first on the more ventral contact. Stimulation was delivered by a Cerestim 96 (Blackrock Instruments), with parameters set manually by the experimenter and stimulation triggered by a separate PC that was either delivering or monitoring task or behavioral stimuli.

The stimulation frequency was chosen based on a previous study[18]; it is also the frequency most commonly used in clinical DBS for psychiatric disorders. In contrast to that study, here we were able to harmonize stimulation parameters between participants, because stimulation was not directly linked to medical treatment. All stimulation was delivered at the image onset to influence a decision-making process that begins with that onset. In all

participants, before task-linked stimulation, we first tested stimulation at 1, 2 and 4 mA, for 1 second, repeated 5 times with 5-10 seconds between each 130 Hz pulse train. We informed participants that stimulation was active and repeatedly asked them to describe any acute perceptions. We verified a lack of epileptiform after-discharges and ensured that the participants could not detect stimulation, e.g. through unusual sensations. If participants reported any sensation (e.g., tactile experiences in limbs or head), we limited task-linked stimulation to the next lowest intensity. In a secondary analysis, the recordings during these test stimulations were used to verify that dorsal and ventral stimulation sites had different patterns of cortical activation (see Extended Data Fig. 3).

Before and after all stimulation experiments, including closed-loop stimulation (see below), an experienced psychiatric interviewer (ASW) asked participants to describe their current emotional state. We similarly prompted participants to describe any subjective experiences at arbitrary times throughout each stimulation experiment. All such reports were videotaped and transcribed. During the behavioral task runs, participants were fully blind to whether stimulation was active, i.e. we did not inform them that we were beginning stimulated blocks.

Behavior Data Analysis

The primary behavior readout is the reaction time (RT), as all participants were over 95% accurate. First, we analyzed the effects of stimulation and task factors at the trial level using a generalized linear mixed effect model (GLME), as in our prior work[18]:

$$RT \sim Conflict + blockStim + blockNum + (1 \mid Participant)$$

This and all other GLMEs analyzing RT data used a log-normal distribution and identity link function. Fixed effects in the GLME were Conflict (a binary variable coding the trial type as being low (0) or high (1) conflict), stimulation site (blockStim), and block number (blockNum) to account for fatigue or practice effects. Stimulation (blockStim) was coded at the block level, i.e. whether the stimulation site in a given block was dorsal vs. ventral capsule or left vs. right, not whether stimulation was on vs. off on a given trial. Block-level coding was a more parsimonious fit to the data (Akaike Information Criterion = −449.3 for block-level coding, −359.7 for trial-level coding). Participant was a random effect. All categorical variables were automatically dummy-coded by MATLAB's "fitglme" function. We excluded trials with missing responses and with incorrect responses.

We further tested for a possible interaction between stimulation and the trial-to-trial conflict level, by fitting an alternate model with an interaction term:

$$RT \sim Conflict + blockStim + Conflict^{*}blockStim + blockNum + (1 \mid Participant)$$

We assessed this model against our primary GLME by comparing Akaike's Information Criterion (AIC), which decreases in models that are more parsimonious fits to the observed data.

To develop closed-loop control and neural decoding strategies, we needed to convert this block-level analysis to a trial-by-trial estimate of participants' RT. Because the task rapidly switches back and forth between trial types, however, RT on any given trial is influenced by changes in conflict or interference in addition to (putative) stimulation effects and random variability. We sought to separately measure these processes and their change in response to capsular stimulation. To achieve this, we applied a state space or latent variable modeling framework[74,75]. In this type of model, the variables of interest (here, the "true" baseline RT and conflict-induced slowing, without influence from stochastic RT variations) are not directly observable. They are assumed to influence an observable variable (the actually observed RT) through a functional scaling, with additive noise. A further advantage is that this class of models is Markovian – the estimate of the unobserved variables ("states") on any given trial depends only on the currently observed RT and the estimate of the states on the prior trial. This allows efficient, trial-by-trial computation that is well suited to real time monitoring and control. For exactly this reason, a specific class of state-space model, the Kalman filter, has long been used for neural decoding in motor brain-computer interface applications[37,38,76].

Here, we used the COMPASS toolbox[75] for MATLAB to fit a model that extracts a trial-level estimate of the RT independent of conflict effects. This model takes the form:

$$\log y_{\mathrm{RT},k} = x_{base,k} + \boldsymbol{I}_{conflict,k}\, x_{conflict,k} + \varepsilon_k \qquad \varepsilon_k \sim N(0, \sigma_\varepsilon^2) \tag{1}$$

where $y_{\mathrm{RT},k}$ is the RT on trial k, and the x_k are latent, unobserved variables that we have previously termed "cognitive states". The observation noise, ε_k would capture other non-structured processes that influence the trial-to-trial RT. Note that this model follows the same distribution or link assumptions as the static GLME above, namely a log-normal distribution. One of the differences between this modeling approach and the classic Kalman filter is that RT is markedly non-Gaussian. It can only have positive values, and is very skewed – RT distributions across a range of cognitive tasks have long right tails[75,77,78]. The log-normal assumption corrects for this. The latent variables were modelled as:

$$x_{base,k} = a_1\, x_{base,k-1} + v_{1,k} \qquad v_{1,k} \sim N(0, \sigma_{1,v}^2) \tag{2.1}$$

$$x_{conflict,k} = a_2\, x_{conflict,k-1} + v_{2,k} \qquad v_{2,k} \sim N(0, \sigma_{2,v}^2) \tag{2.2}$$

where, a_1 and a_2 define the decay of the state variables over time. $v_{1,k}$ and $v_{2,k}$ are mutually independent white noise processes with zero mean and variance $\sigma_{1,v}^2$ and $\sigma_{2,v}^2$, respectively. That is, we assumed that these two processes can vary entirely independently of one another (even though stimulation may influence both). Because both the state values on each trial and the parameters linking those values to the RT are unknown, this model has no closed-form solution. COMPASS estimates model parameters through an iterative expectation-maximization algorithm[75]. We verified convergence of this algorithm on all participants by inspection of the model likelihood plot. We also tested models that included both accuracy and RT as observable outputs; these yielded no significant improvement over RT-only models (see main text and Supplementary Fig. 5).

$x_{base,k}$ represents the expected RT in the absence of conflict or other external influencing factors, whereas $x_{conflict,k}$ represents the expected effect of conflict on the RT. $x_{conflict,k}$ is an indicator variable, such that $x_{conflict,k}$ only affects the expected RT on high-conflict trials. x_{base} can be thought of as encoding more general, overarching aspects of cognitive control, such as effortful attentional focus on task stimuli, maintenance of goals in working memory, and preparation to inhibit a prepotent response on incongruent or high-conflict trials. $x_{conflict}$, in that framework, represents the cognitive load of actually deploying the response inhibition in response to conflict. This is sometimes framed as a "reactive" model of cognitive control. We derived this two-variable model from both theory and prior work. We previously observed[18] (and confirmed in this work, see above) that capsular stimulation affects both high- and low-conflict trials. That is, it mainly alters the expected RT x_{base}. We considered, however, that reactive control in response to conflict might also be affected, and wanted to specifically test for that potential effect. Further, in prior work, we found that this two-state model was necessary to accurately model RT during this same task[74]. We verified that the two-state model was an appropriate fit to our data by comparing the RT residuals to those expected from a white-noise process. Unmodeled variance (e.g., a third cognitive process not captured in our model) would lead to structured residuals that violate white-noise expectations.

The state-space model assumes that cognitive states are slowly varying, i.e. they show a strong autocorrelation. (This was true in practice; the estimated values of a_1 and a_2 were close to 1 for all participants.) We thus cannot use the GLME to analyze stimulation-induced change in these latent variables (x_{base}, $x_{conflict}$) because they strongly violate the GLME's assumption that individual datapoints are independent. We instead used non-parametric permutation testing, which is well-established as a method for inferential statistics on autocorrelated time-series[79]. The stimulation labels of individual blocks were shuffled 1,000 times, with the shuffling nested within individual participants. This created a distribution of cognitive state values under the assumption of no difference between stimulation sites (or between stimulated and non-stimulated trials). From that distribution, we inferred the p-value of the actual state values under stimulation. For both the raw RT GLME and the cognitive state permutation tests, we compared up to 4 stimulation sites in each participant to baseline (no stimulation, NS1). Within each analysis, we corrected the p-values for these multiple comparisons using a false discovery rate (FDR) step-down procedure via MATLAB's "fdr" function.

Closed Loop Stimulation

In three participants, we used that same state space model to implement closed-loop control. First, for each participant, we estimated model parameters by running the expectation-maximization fitting algorithm on 1-3 days of prior MSIT performance without brain stimulation. We used all available data for each participant to fit these models. These parameters were then provided to a real-time engine (also based on COMPASS, specifically the real-time decoding function, compass_filtering) that estimated x_{base} and $x_{conflict}$ on each trial. This specifically leverages the Markovian nature of our model, in that estimating these states on each new trial is extremely fast once the model parameters are initially estimated.

We attempted to control x_{base}, which we considered to track the overall difficulty of sustaining attention and exerting cognitive control during conflict (more difficulty leading to longer RTs). In that framework, cognitive control enhancement would be reflected in a decrease in x_{base}. To achieve this, if the estimate on trial k was above a manually selected threshold, the system delivered electrical stimulation at the time of image or stimulus presentation on trial $k+1$. The threshold and stimulation effects were visible to the experimenter, but not to the participant, through a custom GUI running on a separate laptop. We set the threshold to attempt a decrease in x_{base} from its unstimulated value. The supervising experimenter (ASW) observed the participant's best performance from real-time state estimation during the initial non-stimulation block (64-128 NS1 trials). This process estimated both the mean (maximum likelihood value) of x_{base}, plus a confidence bound. We then set the threshold/target for x_{base} to be below that lower confidence bound, i.e. to be outside the range of possible values that were consistent with the baseline or unstimulated performance. There was no quantitative rule for the distance between the estimated state bound and the control target; it was selected *ad hoc* to be outside our subjective estimate of overall process variability. Stimulation parameters and hardware were identical to the setup for open-loop stimulation. Because of limited experimental time, we could do either open-loop or closed loop stimulation experiments with a particular participant, but not both. We attempted closed-loop control using the stimulation site that had the largest behavioral effect during open-loop experiments in the 6 previous participants. When that site was unavailable (not implanted in a given participant), we used the next best choice that was available.

For analysis of the closed loop stimulation results, we re-ran the complete state-space estimation offline over the whole dataset, rather than using the less-accurate state values estimated in real time. A key difference is that the offline estimation contains a forward (filtering) and backward (smoothing) pass, allowing future data to influence each trial's estimate non-causally. By considering more information, this offline estimate more accurately reflects the "true" cognitive process and its change in response to stimulation. To directly compare closed-loop and open-loop stimulation, we normalized the state values between these two runs such that that the unstimulated blocks in both paradigms had a mean value of 1. That is, both open-loop and closed-loop results were expressed as change vs. the unstimulated condition on the same day.

Neural Data Analysis – Preprocessing

Local field potentials (LFP) were analysed using custom analysis code in MATLAB (Mathworks) based on FieldTrip (http://www.fieldtriptoolbox.org). To reduce the influence of volume conduction[80], LFPs were bipolar re-referenced by subtracting those recorded at consecutive electrode contacts on the same electrode shank. LFP was recorded from electrode pairs spanning 16 brain regions: prefrontal, cingulate, orbitofrontal, temporal, and insular cortices, amygdala, hippocampus, nucleus accumbens, and caudate (Extended Data Fig. 1). All LFP data were decimated to 1000 Hz and de-meaned relative to the entire recording. 60 Hz line noise and its harmonics up to 200 Hz were removed by estimating noise signals through narrow bandpass filtering, then subtracting those filtered signals from the original raw signal. We removed pathological channels with interictal epileptiform

discharges (IEDs). We detected such channels with a previously-reported algorithm that adaptively models distributions of signal envelopes to discriminate IEDs from normal LFP[81]. We then used a Morlet wavelet decomposition to estimate power in 6 frequency bands (4-8, 8-15, 15-30, 30-55, 65-110, and 135-200 Hz). This decomposition used a 1 Hz frequency resolution and 10 ms timestep, with the default parameters of 7 cycles per wavelet and a Morlet/Gaussian width of 3 standard deviation. We fractionated the high gamma (65-200 Hz) band into lower and upper bands to bypass the stimulation frequency at 130 Hz and a 60Hz harmonic at 120 Hz. All time-frequency decomposition was performed after individual trials or epochs were cut from the continuous recording, so that artifacts on stimulation trials could not influence power calculations on non-stimulation trials. Each trial was cut with a buffer of 2 seconds on each side to mitigate edge effects in the wavelet transform; data from these buffers were discarded before analysis.

Neural Data Analysis – Mid-Frontal Theta Power

Exercise of cognitive control is associated with higher theta (4-8 Hz) power in a fronto-cingulate network[18,20,28,29,49]. We have specifically reported that stimulation in the internal capsule increases that task-evoked theta[18]. As an initial manipulation check, we sought to replicate these prior results. We analyzed an epoch of 0.1-1.4 seconds after image onset, which covered the decision-making period up to the median RT. We focused this analysis on non-phase-locked oscillations, which are more strongly correlated with cognitive conflict effects than are their phase-locked counterparts[25,82]. From the target epoch, we subtracted the time-domain evoked response (ERP), which contains all phase-locked activity. We calculated this ERP separately for high- and low-conflict trials, and subtracted the appropriate ERP from each trial's time-domain data. We then transformed the time-domain to a time-frequency representation as above. We averaged power in our analysis epoch within the theta band. For visualization, we normalized this power as a log ratio relative to a baseline period of 0.5 seconds preceding image onset. For analysis, this log transformation is built into the GLM (see below).

To verify that higher conflict evoked higher frontal theta, we analyzed the blocks without stimulation. This avoids confounding effects of stimulation and conflict. For each participant, we pre-selected pre-frontal cortical (PFC) channels that had a significant increase over baseline in task-evoked theta (t-test with threshold of $p<0.05$ uncorrected). For this initial pre-screening step, to avoid a circular analysis, we did not split trials into high/low conflict. Rather, we identified channels that showed a theta-band response in general to performing MSIT. In this reduced set of channels, we then divided the trials into low and high conflict, then computed the non-phase-locked theta power, as noted above. We combined all pre-selected channels in each PFC region, and for each region we fit the GLME: Theta ~ Conflict + (1|Participant), where Conflict is a binary variable coding the trial type as being low (0) or high (1) conflict. This and all other GLMEs analyzing LFP power data used a log-normal distribution and identity link function. We false discovery rate (FDR) corrected the resulting p-values for testing of multiple PFC regions.

We then tested whether open loop capsular stimulation caused a significant increase in theta in the unstimulated trials within a stimulation block (NS2) compared to those in the

unstimulated blocks (NS1; see Fig. 1C-D). To accurately assess stimulation effects, we discarded stimulation trials that we presumed to be substantially contaminated by artifact. We then compared two types of non-stimulated trials (Fig. 1D). NS1 trials were from blocks in which no brain stimulation was given on any trial. NS2 trials were from blocks with stimulation, but were the pseudo randomly-selected 50% of trials that did not receive stimulation. These NS2 trials were artifact-free, but still showed the behavioral effect of stimulation, and thus should also show physiologic changes related to that behavior change. We therefore tested whether the normalized theta power in NS2 trials was significantly greater than that in NS1, again using a GLME: Theta ~ blockStim + (1|Participant)

For this model, we chose one PFC channel for each participant that had the highest theta during NS2 trials (regardless of conflict level or stimulation site, again to avoid circular analysis). P-values were again FDR corrected to control for testing of multiple stimulation sites against non-stimulation.

Neural Data Analysis – Decoding

To effectively treat psychiatric disease, closed-loop stimulation will ultimately need to be applied outside of a structured task-based paradigm, i.e. as patients go about their daily lives. That, in turn, requires the ability to detect lapses in cognitive control and conflict responses directly from brain activity. We thus developed a neural decoder for our cognitive state variables. We applied a neural encoding-decoding analysis with automatic feature selection, as in [74]. The decoded variables were x_{base} and $x_{conflict}$ from the model in equation (1). The neural features used for decoding were the LFP power, in the above-mentioned frequency bands, averaged over a 2 second interval starting at the MSIT image onset. We broadened the analysis beyond the theta band because prior literature suggests that while theta is strongly associated with cognitive control, other frequencies also carry substantial information about task performance[27]. The 2 second epoch was chosen to include both the response and post-response processing. This wider window produced smoother features with less trial to trial variance, improving decoder stability. Here, we averaged only across a 2 second time interval (200 samples) to get power features per-trial. Similar to the theta analysis, the encoder model fitting considered only NS1 and NS2 trials, to prevent the influence of stimulation artifact. We focused on LFP spectral power (rather than other potential behavioral covariates such as connectivity or coherence) because power can be efficiently computed within currently available implantable neural devices[41,83,84]. Successful decoding of task performance from power alone could pave the way for use of these closed-loop controllers in clinical settings.

Decoding analyses were performed with out-of-sample validation, using both stimulated and unstimulated MSIT datasets. For each participant's data, 20-66% of the total trials were used to fit an encoding model (training set). These consisted of NS1 trials in unstimulated datasets and both NS1 and NS2 trials in the stimulated experimental datasets. The training trials were selected from contiguous blocks of trials that, collectively, covered the full range of the states during an experiment. The encoding model that we used is a linear model of the form $Y_k \sim 1+\beta x_k$, where Y_k is a neural feature and x_k is one of the cognitive states on the k-th trial. We considered a feature to be a candidate for decoding if the modified F-statistic[74]

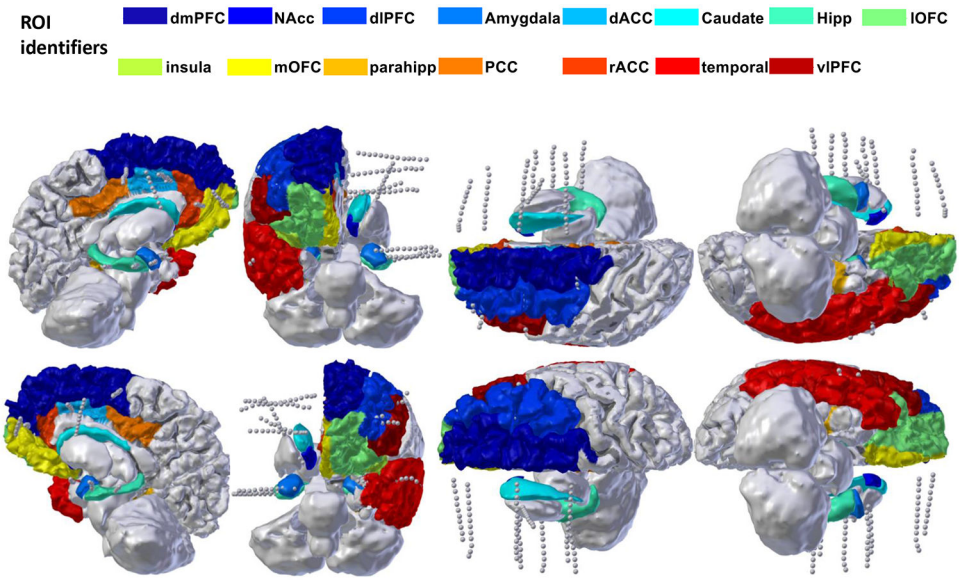
of the corresponding model corresponded to $p < 0.01$ (uncorrected). This procedure selected a set of candidate neural features that potentially independently encoded each cognitive state. The exact number of training trials for each dataset was determined as the minimum required to have a non-zero number of features selected by the encoding procedure.

Next, to reduce overfitting, we pruned the selected set of neural features. The validation set for this pruning used 21-50% of the dataset for each participant, and had 25-50% overlap with the training set. We allowed this overlap to ensure that the validation set was sufficiently large for model pruning without leaving too few trials available for the subsequent test set (see below). On this validation set, we estimated the posterior distribution of the cognitive state from the neural data, through a Bayesian filtering process similar to the one used to estimate these same latent variables from RT[74]. We calculated the root mean square error (RMSE) between the neurally decoded state and the "true" (estimated from behavior) cognitive state in this held-out test set[74]. We then sequentially dropped the feature whose removal led to the most improvement in RMSE. The final decoder was then the set of features that survived this dropping step, i.e. where dropping any further feature would increase RMSE on the validation set.

Finally, we reported the decoder's performance on a fully held out test set, namely the 18-46% of remaining trials from each participant that had not been used for either decoder training or feature pruning. The exact number of trials used for each participant's decoder fitting, pruning, and testing is given in Supplementary Table 6. A key challenge in performing these analyses was that the latent cognitive states (x_{base} and $x_{conflict}$) are themselves multivariate Gaussian estimates. The estimate's value can depend on the starting point of the expectation-maximization process used to fit the state-space model. To control for this, we re-ran the behavioral estimation for each participant 1,000 times with different random seeds, producing 1,000 estimates of the underlying trajectory. (Further details are given in [74]). We then evaluated the neural decoder's performance based on whether its point estimate of the decoded state was within the confidence interval derived from these multiple trajectories.

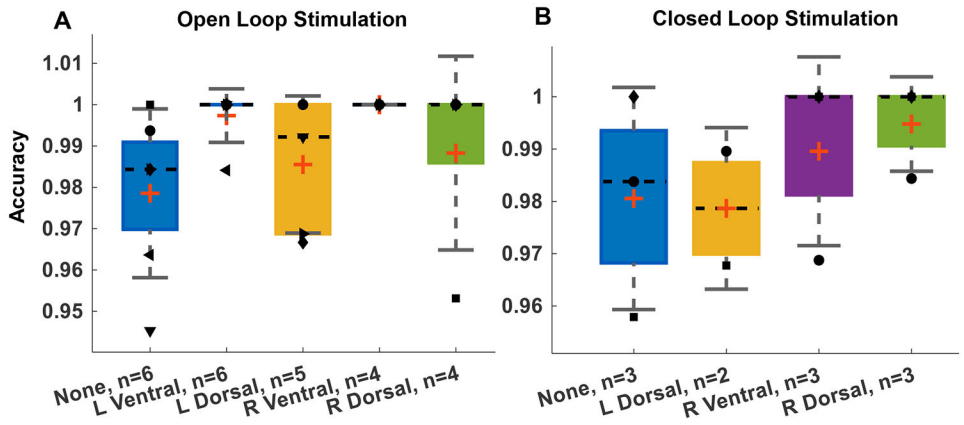
We fit this encoding-decoding model separately to data from unstimulated sessions (consisting of only NS1 trials) as well as to stimulated sessions (both NS1 and NS2 trials), to determine how the encoding structure was altered by electrical stimulation. We did not include stimulated trials in this analysis, because there is a prominent stimulation artifact that makes these trials easily discriminable. In cases of stimulation-behavior correlation, behavior could be trivially decoded simply by detecting the artefact.

Extended Data



ED Fig. 1. iEEG Recording montage.

Example recording montage from a single participant, with cortical parcellation overlaid. Electrode shanks represented by the grey dotted lines access a broad network covering multiple prefrontal structures, superficial and mesial temporal lobe, and striatum/internal capsule.



ED Fig. 2. Accuracy results.

Accuracy during different stimulation experiments, for A) open-loop and B) closed-loop capsular stimulation. Boxes show the mean and confidence intervals for accuracy with stimulation at each site. Colors indicate stimulation sites as in the main text. The p-value above each bar represents a binomial exact test of accuracy compared to the non-stimulated baseline condition, with Benjamini-Hochberg false discovery rate correction. All accuracies are above 95%, with accuracy during stimulated blocks being very slightly higher in most cases. No results exceed chance significance.

We did not have open- and closed-loop data from the same participants. To compare the CL and OL conditions, we therefore compared their accuracies across participants with a Fisher exact test for each of the three stimulation sites (L Dorsal, R Ventral, R Dorsal) that were used in both conditions.

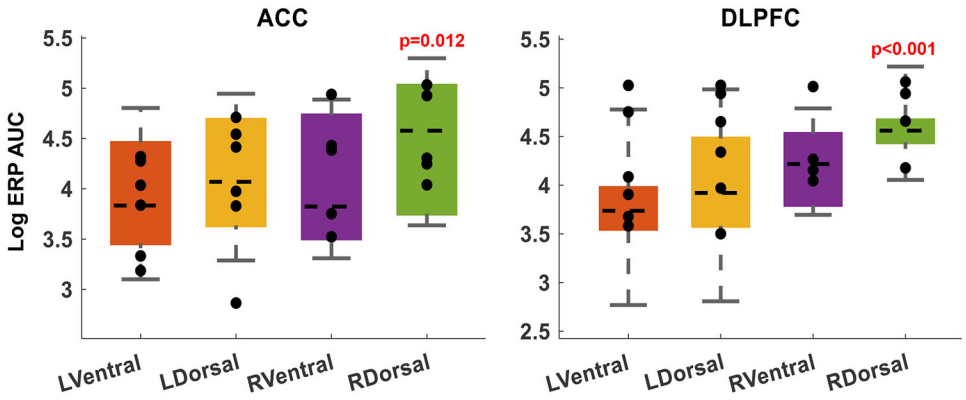
L Dorsal: p=0.645

R Ventral: p=0.440

R Dorsal: p=0.655

These provide no evidence for a difference between OL and CL conditions.

These results do not support a change in accuracy with any stimulation type. That is, the observed decrease in reaction times is a true performance improvement, not a shift along a speed-accuracy tradeoff. We were unable to analyse accuracy in the GLME framework because the differences between stimulation sites are so small as to make the models non-identifiable in all cases.

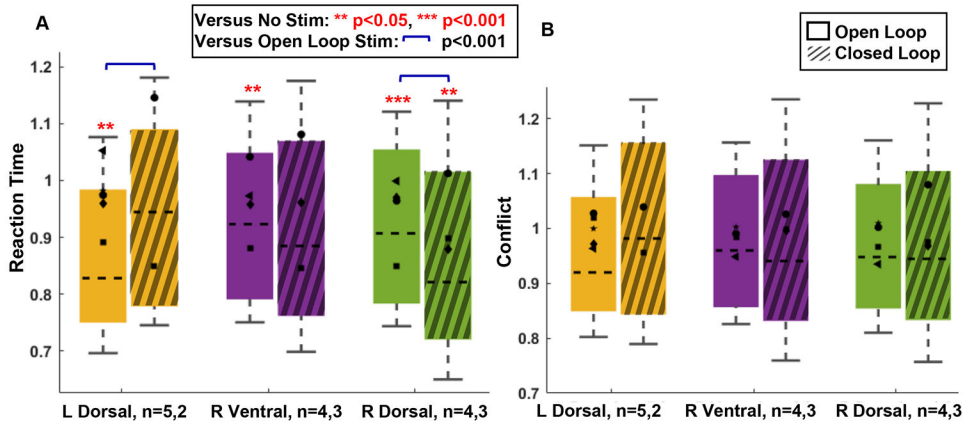


ED Fig. 3. Cortical response to internal capsule stimulation.

Topographic structure of the internal capsule yields differential cortical effects from stimulation at different capsular sites. Before task-linked stimulation, we performed safety/perceptibility testing, where we repeatedly stimulated each potential site with brief 130 Hz pulse trains (see Methods). Each of those trains created an evoked response potential (ERP) in various cortical regions. For each participant, we collected all sEEG channels that were localized to grey matter of DLPFC or ACC. We then quantified the post-train ERP as the sum of the area under its polyphasic curve (AUC). We limited this analysis to channels ipsilateral to the site of stimulation. Each marker represents the mean log(AUC) in one participant. Boxes show the mean and confidence intervals for the ERP AUC from stimulation at each site.

The stimulation sites that were more effective behaviorally produced the largest ERPs in these cognitive-control-associated regions, with right dorsal stimulation having the largest effects. (p-values represent t-test on the regression coefficients of a log-normal GLM, i.e. the same analysis used in main text Fig. 2). In the left hemisphere, dorsal stimulation produced larger responses than ventral stimulation, but this did not reach statistical significance given the small number of trials (5 test trains per participant).

These results are consistent with the known topography of the internal capsule, where fibers that connect DLPFC and ACC to thalamus run in the dorsal-most part of the anterior limb, i.e. in close proximity to our chosen dorsal electrodes.



ED Fig. 4. Open-loop and closed-loop effects in manifest data.

Effect of open-loop and closed-loop capsular stimulation on A) reaction time (RT) and B) Conflict related RT. Conflict related RT is calculated as the residual reaction time after subtracting the mean reaction time of the congruent trials in the same block, i.e. it has an expected value of 0 ms on non-conflict trials. We consider it as the closest raw/manifest data analogue of $x_{conflict}$. We note, however, that both of these manifest RT variables include the Gaussian noise that is removed by the state-space filtering that produces x_{base} and $x_{conflict}$. As such, the data in this figure are by definition noisier, and the analysis has lower statistical power. This leads to smaller effects in the open-loop results compared to main text Fig. 3. Closed-loop stimulation of the right dorsal internal capsule (our most effective open-loop intervention) was more effective than its open loop counterpart at reducing raw RT (the counterpart of x_{base}). Consistent with the specificity illustrated in main text Fig. 4C, there was no advantage for closed-loop stimulation on the conflict-specific RT (the counterpart of $x_{conflict}$). All formatting and graphical elements follow the conventions of main text Fig. 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge technical assistance with data collection from Afsana Afzal, Gavin Belok, Kara Farnes, Julia Felicione, Rachel Franklin, Anna Gilmour, Aishwarya Gosai, Mark Moran, Madeleine Robertson, Christopher Salthouse, Deborah Vallejo-Lopez, and Samuel Zorowitz. We also thank the research participants, without whose generous help none of this would have been possible. This work was supported by grants from the Defense Advanced Research Projects Agency (DARPA) under Cooperative Agreement Number W911NF-14-2-0045 issued by the Army Research Organization (ARO) contracting office in support of DARPA's SUBNETS Program, the National Institutes of Health (UH3NS100548, R01MH111917, R01MH086400, R01DA026297, R01EY017658, K24NS088568), Ellison Foundation, Tiny Blue Dot Foundation, MGH Executive Council on Research, OneMind Institute, and the MnDRIVE and Medical Discovery Team-Addictions initiatives at the University of Minnesota. The views, opinions, and findings expressed are those of the authors. They should not be interpreted as representing the official views or policies of the Department of Defense, Department of Health & Human Services, any other branch of the U.S. Government, or any other funding entity.

Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. Pre-processed and anonymized neural and behavioural data

are available through Zenodo at https://zenodo.org/record/5083120#.YOhvWehKiUk and https://zenodo.org/record/5085197#.YOhtouhKiUk.

References

1. Roehrig C Mental Disorders Top the List Of The Most Costly Conditions In The United States: $201 Billion. Health Aff. Proj. Hope 35, 1130–1135 (2016).
2. Gordon JA On being a circuit psychiatrist. Nat. Neurosci 19, 1385–1386 (2016). [PubMed: 27786177]
3. Insel TR Disruptive insights in psychiatry: transforming a clinical discipline. J. Clin. Invest 119, 700–705 (2009). [PubMed: 19339761]
4. Mayberg HS Targeted electrode-based modulation of neural circuits for depression. J. Clin. Invest 119, 717–725 (2009). [PubMed: 19339763]
5. Graat I, Figee M & Denys D The application of deep brain stimulation in the treatment of psychiatric disorders. Int. Rev. Psychiatry Abingdon Engl 29, 178–190 (2017).
6. Sullivan CRP, Olsen S & Widge AS Deep brain stimulation for psychiatric disorders: From focal brain targets to cognitive networks. NeuroImage 225, 117515 (2021). [PubMed: 33137473]
7. Scangos KW & Ross DA What We've Got Here Is Failure to Communicate: Improving Interventional Psychiatry With Closed-Loop Stimulation. Biol. Psychiatry 84, e55–e57 (2018). [PubMed: 30261977]
8. Widge AS, Malone DA & Dougherty DD Closing the Loop on Deep Brain Stimulation for Treatment-Resistant Depression. Front. Neurosci 12, (2018).
9. Widge AS & Miller EK Targeting Cognition and Networks Through Neural Oscillations: Next-Generation Clinical Brain Stimulation. JAMA Psychiatry 76, 671–672 (2019). [PubMed: 31116372]
10. Cuthbert BN & Insel TR Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11, 126 (2013). [PubMed: 23672542]
11. Kirkby LA et al. An Amygdala-Hippocampus Subnetwork that Encodes Variation in Human Mood. Cell 175, 1688–1700.e14 (2018). [PubMed: 30415834]
12. Veerakumar A et al. Field potential 1/f activity in the subcallosal cingulate region as a candidate signal for monitoring deep brain stimulation for treatment-resistant depression. J. Neurophysiol 122, 1023–1035 (2019). [PubMed: 31314668]
13. Widge AS et al. Treating refractory mental illness with closed-loop brain stimulation: Progress towards a patient-specific transdiagnostic approach. Exp. Neurol 287, 461–472 (2017). [PubMed: 27485972]
14. Badre D Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn. Sci 12, 193–200 (2008). [PubMed: 18403252]
15. Grahek I, Shenhav A, Musslick S, Krebs RM & Koster EHW Motivation and cognitive control in depression. Neurosci. Biobehav. Rev 102, 371–381 (2019). [PubMed: 31047891]
16. Kouneiher F, Charron S & Koechlin E Motivation and cognitive control in the human prefrontal cortex. Nat. Neurosci 12, 939–945 (2009). [PubMed: 19503087]
17. Solomon M et al. The neural substrates of cognitive control deficits in autism spectrum disorders. Neuropsychologia 47, 2515–2526 (2009). [PubMed: 19410583]
18. Widge AS et al. Deep brain stimulation of the internal capsule enhances human cognitive control and prefrontal cortex function. Nat. Commun 10, 1–11 (2019). [PubMed: 30602773]
19. Widge AS, Heilbronner SR & Hayden BY Prefrontal cortex and cognitive control: new insights from human electrophysiology. F1000Research 8, (2019).
20. Cavanagh JF & Frank MJ Frontal theta as a mechanism for cognitive control. Trends Cogn. Sci 18, 414–421 (2014). [PubMed: 24835663]
21. Sharpe MJ et al. An Integrated Model of Action Selection: Distinct Modes of Cortical Control of Striatal Decision Making. Annu. Rev. Psychol 70, 53–76 (2019). [PubMed: 30260745]
22. Bari A & Robbins TW Inhibition and impulsivity: Behavioral and neural basis of response control. Prog. Neurobiol 108, 44–79 (2013). [PubMed: 23856628]

23. Burguière E, Monteiro P, Mallet L, Feng G & Graybiel AM Striatal circuits, habits, and implications for obsessive–compulsive disorder. Curr. Opin. Neurobiol 30, 59–65 (2015). [PubMed: 25241072]
24. Cavanagh JF & Frank MJ Frontal theta as a mechanism for cognitive control. Trends Cogn. Sci 18, 414–421 (2014). [PubMed: 24835663]
25. Cohen MX Midfrontal theta tracks action monitoring over multiple interactive time scales. NeuroImage 141, 262–272 (2016). [PubMed: 27475291]
26. Ryman SG et al. Impaired Midline Theta Power and Connectivity During Proactive Cognitive Control in Schizophrenia. Biol. Psychiatry 84, 675–683 (2018). [PubMed: 29921417]
27. Provenza NR et al. Decoding task engagement from distributed network electrophysiology in humans. J. Neural Eng 16, 056015 (2019). [PubMed: 31419211]
28. Smith EH et al. Widespread temporal coding of cognitive control in the human prefrontal cortex. Nat. Neurosci 22, 1883–1891 (2019). [PubMed: 31570859]
29. Voytek B et al. Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. Nat. Neurosci 18, 1318–1324 (2015). [PubMed: 26214371]
30. Haber SN et al. Circuits, networks, and neuropsychiatric disease: transitioning from anatomy to imaging. Biol. Psychiatry 87, 318–327 (2020). [PubMed: 31870495]
31. Haber SN Corticostriatal circuitry. Dialogues Clin. Neurosci 18, 7–21 (2016). [PubMed: 27069376]
32. Makris N et al. Variability and anatomical specificity of the orbitofrontothalamic fibers of passage in the ventral capsule/ventral striatum (VC/VS): precision care for patient-specific tractography-guided targeting of deep brain stimulation (DBS) in obsessive compulsive disorder (OCD). Brain Imaging Behav. 10, 1054–1067 (2016). [PubMed: 26518214]
33. Li N et al. A unified connectomic target for deep brain stimulation in obsessive-compulsive disorder. Nat. Commun 11, 3364 (2020). [PubMed: 32620886]
34. Dubreuil-Vall L, Chau P, Ruffini G, Widge AS & Camprodon JA tDCS to the left DLPFC modulates cognitive and physiological correlates of executive function in a state-dependent manner. Brain Stimulat. 12, 1456–1463 (2019).
35. Tyagi H et al. A Randomized Trial Directly Comparing Ventral Capsule and Anteromedial Subthalamic Nucleus Stimulation in Obsessive-Compulsive Disorder: Clinical and Imaging Evidence for Dissociable Effects. Biol. Psychiatry 85, 726–734 (2019). [PubMed: 30853111]
36. Sheth SA et al. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature 488, 218–221 (2012). [PubMed: 22722841]
37. Wodlinger B et al. Ten-dimensional anthropomorphic arm control in a human brain–machine interface: difficulties, solutions, and limitations. J. Neural Eng 12, 016011 (2015). [PubMed: 25514320]
38. Hochberg LR et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature 485, 372–375 (2012). [PubMed: 22596161]
39. Shenhav A, Cohen JD & Botvinick MM Dorsal anterior cingulate cortex and the value of control. Nat. Neurosci 19, 1286–1291 (2016). [PubMed: 27669989]
40. Gillan CM, Kosinski M, Whelan R, Phelps EA & Daw ND Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife 5, e11305 (2016). [PubMed: 26928075]
41. Bourget D et al. An implantable, rechargeable neuromodulation research tool using a distributed interface and algorithm architecture. in 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER) 61–65 (2015). doi:10.1109/NER.2015.7146560.
42. Bach DR, Hoffmann M, Finke C, Hurlemann R & Ploner CJ Disentangling Hippocampal and Amygdala Contribution to Human Anxiety-Like Behavior. J. Neurosci 39, 8517–8526 (2019). [PubMed: 31501296]
43. Mégevand P et al. The Hippocampus and Amygdala Are Integrators of Neocortical Influence: A CorticoCortical Evoked Potential Study. Brain Connect. 7, 648–660 (2017). [PubMed: 28978234]
44. Cavanagh JF & Shackman AJ Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. J. Physiol.-Paris 109, 3–15 (2015). [PubMed: 24787485]

45. Gibson WS et al. The Impact of Mirth-Inducing Ventral Striatal Deep Brain Stimulation on Functional and Effective Connectivity. Cereb. Cortex N. Y. N 1991 27, 2183–2194 (2017).
46. Okun MS et al. Deep brain stimulation in the internal capsule and nucleus accumbens region: responses observed during active and sham programming. J. Neurol. Neurosurg. Psychiatry 78, 310–314 (2007). [PubMed: 17012341]
47. Zelmann R et al. CLoSES: A platform for closed-loop intracranial stimulation in humans. NeuroImage 223, 117314 (2020). [PubMed: 32882382]
48. Bush G & Shin LM The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. Nat. Protoc 1, 308–313 (2006). [PubMed: 17406250]
49. Smith EH et al. Frequency-dependent representation of reinforcement-related information in the human medial and lateral prefrontal cortex. J. Neurosci 35, 15827–15836 (2015). [PubMed: 26631465]
50. McTeague LM et al. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. Am. J. Psychiatry 174, 676–685 (2017). [PubMed: 28320224]
51. Computational Psychiatry. (The MIT Press, 2016).
52. National Institute of Mental Health Strategic Plan for Research. 48 https://www.nimh.nih.gov/strategicplan (2020).
53. Wu H et al. Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. Proc. Natl. Acad. Sci 115, 192–197 (2018). [PubMed: 29255043]
54. Wu H et al. Brain-responsive neurostimulation for loss of control eating: early feasibility study. Neurosurgery nyaa300 (2020) doi:10.1093/neuros/nyaa300.
55. Martin DM, McClintock SM, Forster JJ, Lo TY & Loo CK Cognitive enhancing effects of rTMS administered to the prefrontal cortex in patients with depression: A systematic review and meta-analysis of individual task effects. Depress. Anxiety 34, 1029–1039 (2017). [PubMed: 28543994]
56. Grisanzio KA et al. Transdiagnostic symptom clusters and associations with brain, behavior, and daily function in mood, anxiety, and trauma disorders. JAMA Psychiatry 75, 201–209 (2017).
57. Inzlicht M, Shenhav A & Olivola CY The effort paradox: effort is both costly and valued. Trends Cogn. Sci 22, 337–349 (2018). [PubMed: 29477776]
58. Klein E et al. Brain-computer interface-based control of closed-loop brain stimulation: attitudes and ethical considerations. Brain-Comput. Interfaces 3, 140–148 (2016).
59. Cabrera LY et al. Authentic self and last resort: international perceptions of psychiatric neurosurgery. Cult. Med. Psychiatry (2020) doi:10.1007/s11013-020-09679-1.
60. Goering S, Klein E, Dougherty DD & Widge AS Staying in the loop: relational agency and identity in next-generation DBS for psychiatry. AJOB Neurosci. 8, 59–70 (2017).
61. Conrad EC, Humphries S & Chatterjee A Attitudes toward cognitive enhancement: the role of metaphor and context. AJOB Neurosci. 10, 35–47 (2019). [PubMed: 31070552]
62. Bick SK et al. Caudate stimulation enhances learning. Brain J. Neurol 142, 2930–2937 (2019).
63. Prerau MJ et al. Characterizing learning by simultaneous analysis of continuous and binary measures of performance. J. Neurophysiol 102, 3060–3072 (2009). [PubMed: 19692505]
64. Paulk AC et al. Bidirectional modulation of human emotional conflict resolution using intracranial stimulation. bioRxiv 825893 (2019) doi:10.1101/825893.
65. Sani OG et al. Mood variations decoded from multi-site intracranial human brain activity. Nat. Biotechnol 36, 954–961 (2018). [PubMed: 30199076]
66. Faul F, Erdfelder E, Buchner A & Lang A-G Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods 41, 1149–1160 (2009). [PubMed: 19897823]
67. González-Villar AJ & Carrillo-de-la-Peña MT Brain electrical activity signatures during performance of the Multisource Interference Task. Psychophysiology 54, 874–881 (2017). [PubMed: 28220517]
68. Kleiner M, Brainard D & Pelli D What's new in Psychtoolbox-3. Perception 36, 1 (2007).

69. Dykstra AR et al. Individualized localization and cortical surface-based registration of intracranial electrodes. NeuroImage 59, 3563–3570 (2012). [PubMed: 22155045]
70. LaPlante R et al. The interactive electrode localization utility: software for automatic sorting and labeling of intracranial subdural electrodes. Int. J. Comput. Assist. Radiol. Surg 12, 1829–37 (2017). [PubMed: 27915398]
71. Widge AS et al. Predictors of hypomania during ventral capsule/ventral striatum deep brain stimulation. J. Neuropsychiatry Clin. Neurosci 28, 38–44 (2015). [PubMed: 26404172]
72. Basu I et al. A neural mass model to predict electrical stimulation evoked responses in human and non-human primate brain. J. Neural Eng 15, 066012 (2018). [PubMed: 30211694]
73. Basu I et al. Consistent linear and non-linear responses to invasive electrical brain stimulation across individuals and primate species with implanted electrodes. Brain Stimulat. 12, 877–892 (2019).
74. Yousefi A et al. Decoding Hidden Cognitive States From Behavior and Physiology Using a Bayesian Approach. Neural Comput 31, 1751–1788 (2019). [PubMed: 31335292]
75. Yousefi A et al. COMPASS: An Open-Source, General-Purpose Software Toolkit for Computational Psychiatry. Front. Neurosci 12, 957 (2019). [PubMed: 30686965]
76. Vaskov AK et al. Cortical Decoding of Individual Finger Group Motions Using ReFIT Kalman Filter. Front. Neurosci 12, 751 (2018). [PubMed: 30455621]
77. Ratcliff R & McKoon G The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Comput. 20, 873–922 (2008). [PubMed: 18085991]
78. Palmer EM, Horowitz TS, Torralba A & Wolfe JM What are the shapes of response time distributions in visual search? J. Exp. Psychol. Hum. Percept. Perform 37, 58–71 (2011). [PubMed: 21090905]
79. Oostenveld R, Fries P, Maris E & Schoffelen J-M FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience https://www.hindawi.com/journals/cin/2011/156869/ (2011) doi:10.1155/2011/156869.
80. Bastos AM & Schoffelen J-M A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. Front. Syst. Neurosci 9, 175 (2015). [PubMed: 26778976]
81. Janca R et al. Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. Brain Topogr. 28, 172–183 (2015). [PubMed: 24970691]
82. Cohen MX & Donner TH Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. J. Neurophysiol 110, 2752–2763 (2013). [PubMed: 24068756]
83. Skarpaas TL, Jarosiewicz B & Morrell MJ Brain-responsive neurostimulation for epilepsy (RNS® System). Epilepsy Res. 153, 68–70 (2019). [PubMed: 30850259]
84. Stanslaski S et al. Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation. IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc 20, 410–421 (2012).

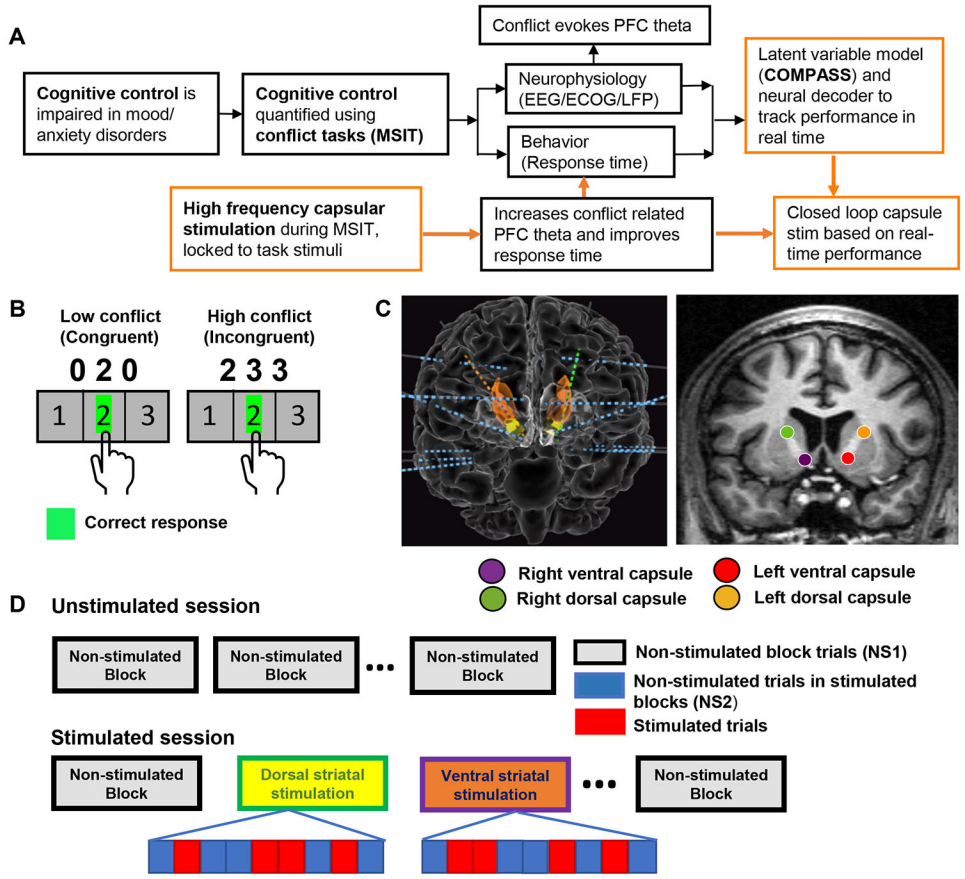


Fig. 1 |. Experimental paradigms.

A) Overview of cognitive conflict/control processing and the current study. Conflict, which requires high control, evokes PFC theta oscillations and slows decision (response) times. Orange boxes highlight this paper's work. We developed intermittent stimulation approaches that improve these physiologic and behavioral correlates of control. We also developed a state-space model and neural decoder to track these quantities in real time. We then merged these for closed-loop enhancement of cognitive control, based on state-space behavior tracking. **B)** Schematic of the Multi Source Interference Task (MSIT), in which participants must inhibit pre-potent responses on 50% of trials. **C)** Schematic of a typical montage of depth electrodes (left) with the caudate nucleus and nucleus accumbens colored as orange and yellow respectively and stimulation targets (right). **D)** Trial structure during experimental sessions. In an unstimulated experimental session (top), blocks of trials were separated by brief rest periods. During an open-loop stimulation experiment (bottom), some blocks of MSIT trials had no stimulation at all (NS1), while others had stimulation only on a randomly-selected 50% of trials. Un-stimulated trials in these blocks are designated NS2.

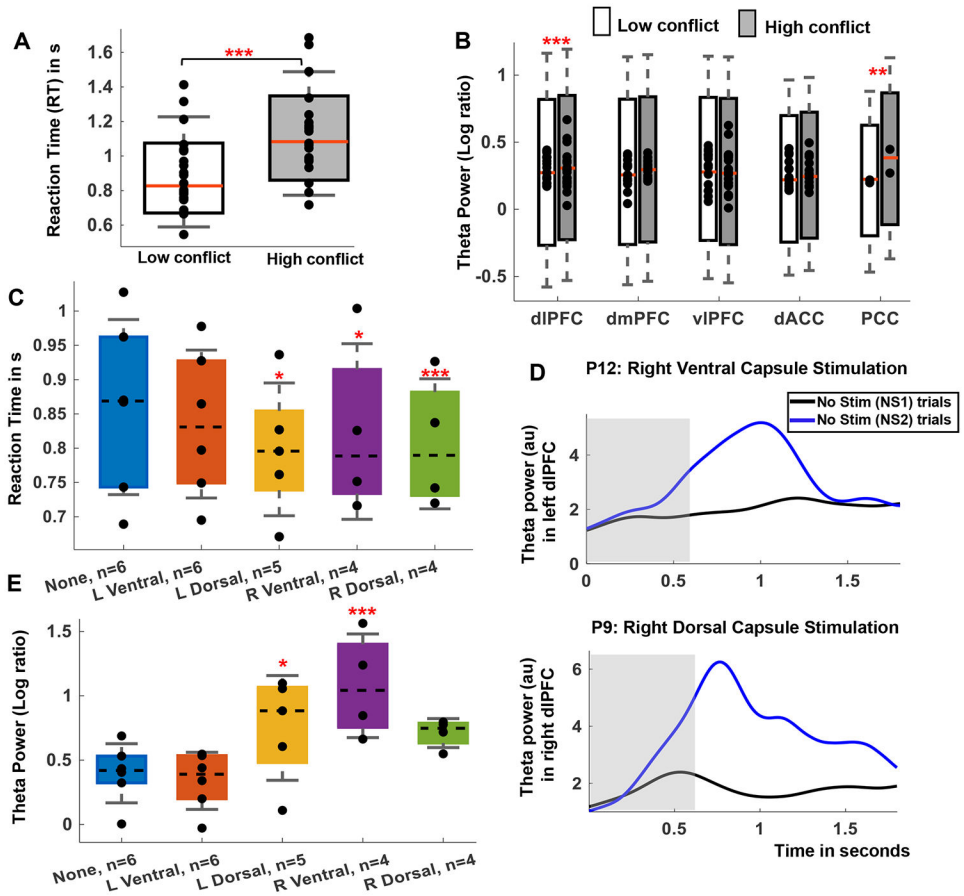


Fig. 2 | Effect of conflict and open-loop capsular stimulation on cognitive control:
A) Reaction time (RT) during low-conflict (white background) and high-conflict (grey background) unstimulated trials (NS1) from 21 participants show significant difference ($p<0.001 = 7.8400e-124$, calculated at the single-trial level across 21 participants). Markers show the mean for each condition for individual participants. **B)** Theta power during low-conflict (white background) and high-conflict NS1 (grey background) trials in frontal regions. Power is expressed as the log ratio relative to pre-trial background. As in prior studies, conflict slowed responses and evoked significantly higher theta in dlPFC ($p<0.001$) and PCC ($p=0.006$), p-values calculated at the single trial level. Markers show the mean for each condition for individual participants. **C)** RT during open loop stimulation during MSIT. In a–c, the dots represent the individual participants, the bars show the median (red line, or black dashed line) and the bar maxima/minima correspond to 75th and 25th percentile respectively; error bars show the standard error of the mean. Stimulation improved task performance, as reflected in significant lower RTs with L Dorsal ($p=0.015$), R Ventral ($p=0.031$) and R Dorsal ($p<0.001$) capsular stimulation. Colors correspond to stimulation sites in Fig 1C, and inferential testing is performed against the unstimulated condition. **D)** Example theta power traces from PFC channels in 2 participants (P9 and P12), with right ventral (top) and right dorsal (bottom) capsular stimulation. NS1 and NS2 trials are as in Fig 1D. The stimulation was from 0-0.6 seconds (grey window), although note that the curves show only data from non-stimulated trials. **E)** Task-evoked theta power (log ratio vs.

pre-trial baseline), across PFC channels, in NS1 trials (None) compared to NS2 trials (all other conditions). L Dorsal and R Ventral capsular stimulation significantly increased theta power ($p=0.043$, $p<0.001$ respectively), as in prior reports[18]. Number of subjects (n) for each stimulation type is specified on the X-axis. In all panels, p-values (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$) are reported after correcting for multiple comparisons using a false discovery rate. dlPFC, dorsolateral PFC; dmPFC, dorsomedial PFC; vlPFC, ventrolateral PFC; dACC, dorsal anterior cingulate cortex; PCC, posterior cingulate cortex.

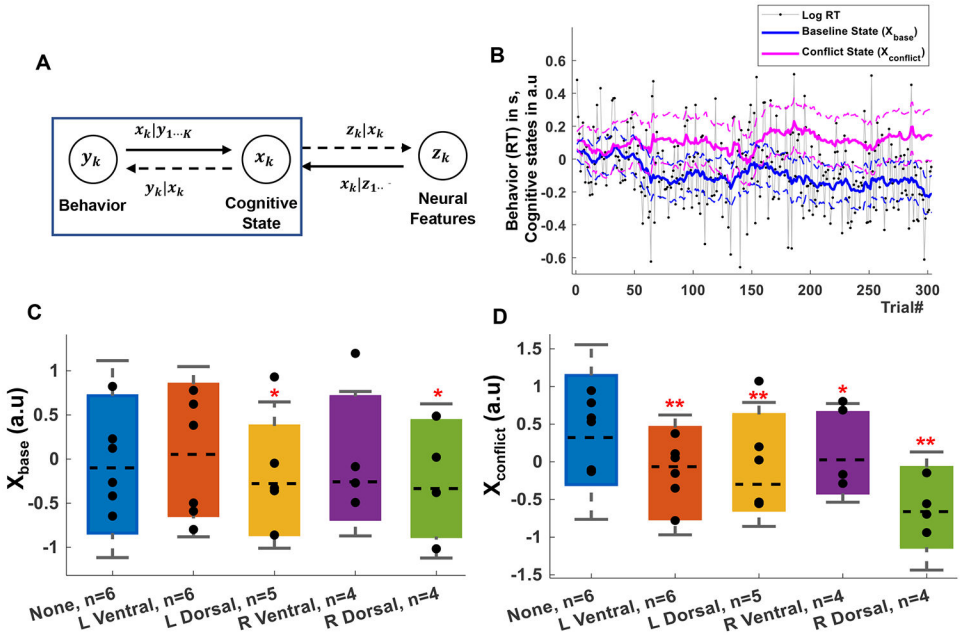


Fig. 3 |. Effect of open-loop capsule stimulation on cognitive control:

A) Schematic of the modeling framework, where behavior and neurophysiology are linked through a low-dimensional latent state space. Here, we focus on inferring latent states from behavior (blue box). **B)** Example of a participant's raw behavior (RT) and its decomposition into *xbase* and *xconflict*. **C)** Effect of open-loop, randomly interleaved stimulation on *xbase* (expected RT). Only dorsal stimulation significantly improved this task component (p=0.026). **D)** Effect of the same stimulation on *xconflict* (expected RT to high conflict trials). Stimulation in all internal capsule sites altered this aspect of cognitive control (R Ventral: p=0,02, rest: p<0.001). Number of subjects (n) for each stimulation type is specified on the X-axis. In C, D, *p*-values (**p* < 0.05; ***p* < 0.001) are reported after correcting for multiple comparisons using a false discovery rate. Statistical inference is through non-parametric permutations due to the highly autocorrelated state variables (see Methods). Markers represent individual participants, bars show the median (dashed line) and the bar maxima/minima correspond to 75th and 25th percentile respectively, error bars show standard error of the mean.

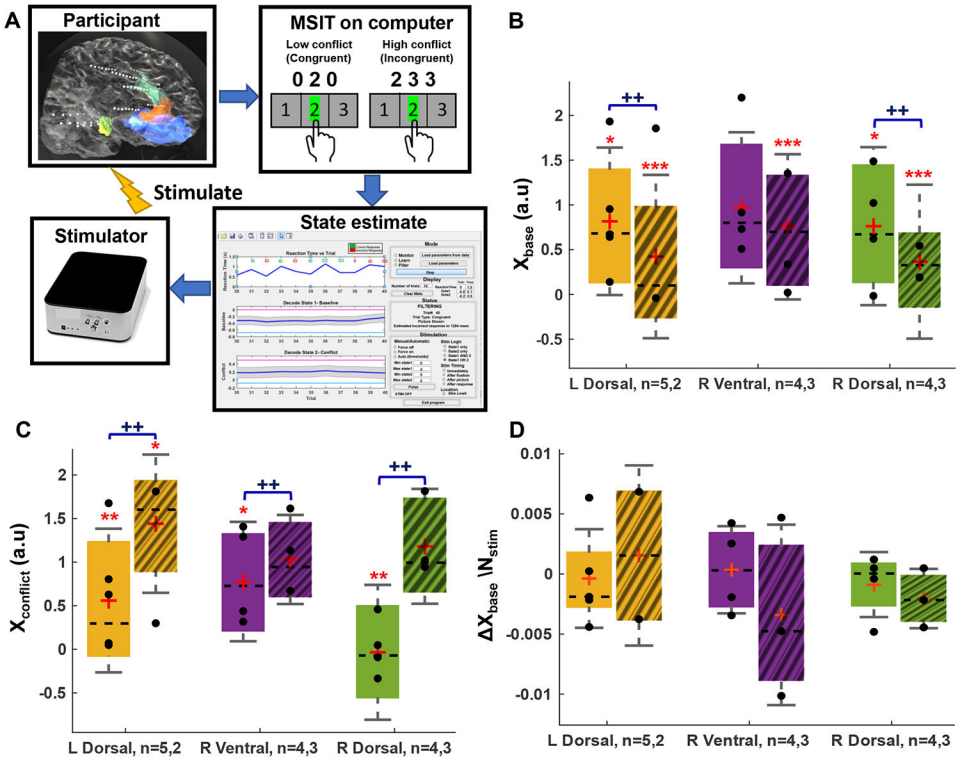


Fig. 4 |. Closed-loop internal capsule stimulation efficiently enhances cognitive control.

A) Schematic of closed loop paradigm. Briefly, the participant with 100+ electrodes recording iEEG (top left) performed MSIT (top right) on a desktop computer. A real time state estimator based in MATLAB (bottom right) calculated x_{base} and $x_{conflict}$ after each trial. This estimator included a threshold-based controller that then triggered stimulation from a neurostimulator (bottom left) on the next trial if the state was above a pre-determined threshold. The real time state estimates were displayed on a MATLAB based GUI. **B)** Effect of open-loop stimulation (solid bars) vs. closed-loop stimulation (dashed bars) on x_{base}. At dorsal sites within the capsule, closed-loop stimulation was more effective at reducing x_{base} (improving task performance, ***p<0.001). **C)** Effect of open- vs. closed-loop stimulation on $x_{conflict}$. Simulation conditioned on x_{base} did not reduce $x_{conflict}$, and in fact significantly increased it at multiple sites (***p<0.001). **D)** Comparison of open- vs. closed-loop effects on x_{base} (Δx_{base}) (with 0 representing no change from NS1) divided by the number of stimulated trials (N_{stim}). A negative value indicates a decrease (desired) in x_{base} caused by a specific stimulation on a block level. State values in B, C are normalized so that unstimulated blocks have a mean state value of 1 for each participant for both experiments, permitting comparison across participants. Significance is determined by a permutation test given the highly autocorrelated data. *p*-values (versus no stimulation: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; versus open loop stimulation: ++ p<0.001) are reported after correcting for multiple comparisons using a false discovery rate. Number of subjects (n) for each experiment is specified on the X-axis for open- and closed-loop participants. Markers represent individual participants, bars show the median (dashed line) and the bar maxima/minima correspond to 75th and 25th percentile respectively, error bars show standard error of the mean. Red crosses represent the mean values.

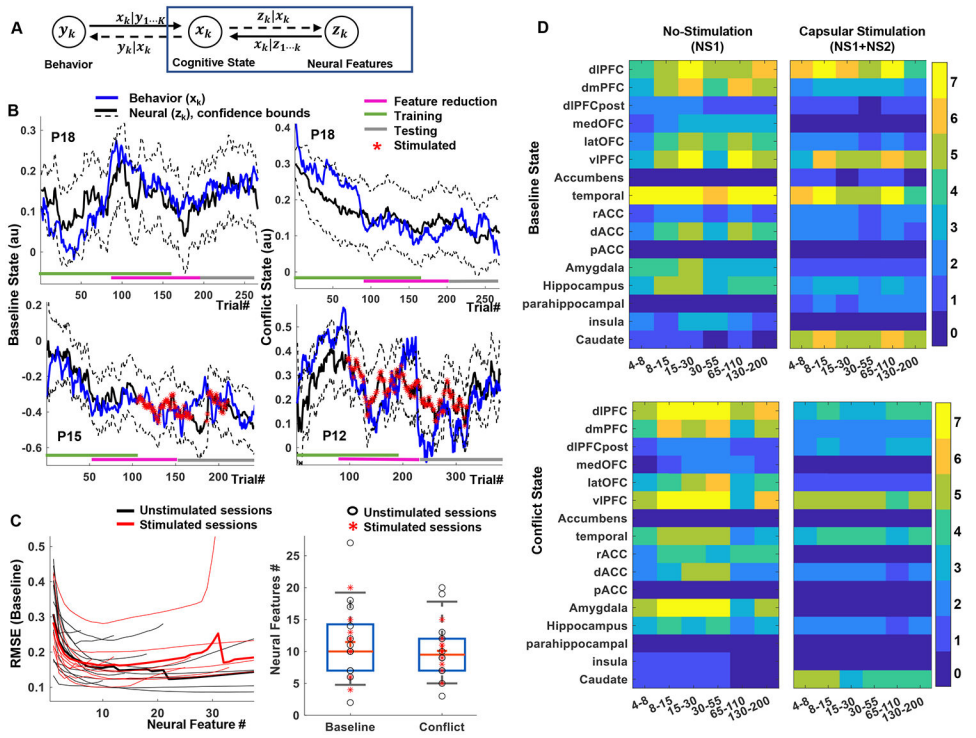


Fig. 5 |. Neural decoding of cognitive states.

A) Schematic of the encoding-decoding framework, which uses the same state variables as Fig 3 (Equation 2). Here, we identify linear dependencies between neural features (LFP power) and the latent cognitive states (blue box). **B)** Examples of baseline (left) and conflict (right) states as estimated from behavior and from neural decoding in subjects P18, P15 and P12. Colored bars at the bottom of the plot indicate the data segments used for training, feature pruning (validation), and testing. There is strong agreement throughout the task run, including on data not used to train the decoding model. Decoding quality is similar during unstimulated (top) and stimulated (bottom) experiments. **C)** Optimum neural feature set determined using a feature dropping analysis (left). Thin lines represent individual participants (n=21), thick lines the mean. The solid circles indicate the number of features that minimizes root mean squared error (RMSE) for x_{base}, for each participant, on a held-out validation set. This number of features did not differ between x_{base} and $x_{conflict}$, or between stimulated and unstimulated blocks (right). Markers represent individual participants, bars show the median (dashed line) and the bar maxima/minima correspond to 75th and 25th percentile respectively, error bars show standard error of the mean. **D)** Number of participants encoding x_{base} (top) and $x_{conflict}$ (bottom) in different brain regions and frequency bands during non-stimulated (left, NS1) and intermittent capsular stimulation (right, NS1+NS2) blocks. Capsular stimulation shifts encoding from cortical regions to subcortical, particularly caudate.