nature biomedical engineering

PEER-REVIEW INFORMATION

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

Corresponding author: Alik Widge

Editorial note

This document includes relevant written communications between the manuscript's corresponding author and the editor and reviewers of the manuscript during peer review. It includes decision letters relaying any editorial points and peer-review reports, and the authors' replies to these (under 'Rebuttal' headings). The editorial decisions are signed by the manuscript's handling editor, yet the editorial team and ultimately the journal's Chief Editor share responsibility for all decisions.

Any relevant documents attached to the decision letters are referred to as **Appendix #**, and can be found appended to this document. Any information deemed confidential has been redacted or removed. Earlier versions of the manuscript are not published, yet the originally submitted version may be available as a preprint. Because of editorial edits and changes during peer review, the published title of the paper and the title mentioned in below correspondence may differ.

Correspondence

Fri 03 Apr 2020
Decision on Presubmission Enquiry nBME-20-0612-PE

Dear Prof Widge,

Thank you for submitting to *Nature Biomedical Engineering* your Presubmission Enquiry, "Closed loop enhancement and neural decoding of human cognitive control".

As you may know, we screen Presubmission Enquiries against our editorial criteria. These editorial judgements are based on considerations of fit to the journal's scope and, when enough information is provided, of the degree of advance, broad implications, and breadth and depth of the work.

The topic of the Presubmission Enquiry is within the remit of the journal, and we would like to invite you to submit a full manuscript so that we can carry out a full editorial assessment. From the information that you have provided, we would most likely be looking for demonstration of efficacy of closed-loop cognitive control and/or subjective wellbeing with a suitably powered number of participants, and in that regard using metrics of wellbeing and cognitive control that are suitable for the psychiatric disease under investigation.

I should also ask you to please fill in our reporting summary and policy checklist. (Please note that these forms are dynamic PDF files that can only be properly visualized and filled in by using Acrobat Reader.)

Both documents are aimed at ensuring good reporting standards and at easing the interpretation of results, and will be available to the reviewers. Should the manuscript be eventually published, the reporting summary will be attached to the published PDF of the paper and will also be available as supplementary information. More information is available on the editorial policies page.

When you are ready to submit the manuscript, please upload the revised manuscript files as well as the

reporting summary and policy checklist. Please also upload a brief cover letter that describes the main results of the work and places it into a broader context.

Best wishes,

Michelle

__
Dr Michelle Korda
Senior Editor, *Nature Biomedical Engineering*

Thu 18 Jun 2020
Decision on Article NBME-20-0612A

Dear Prof Widge,

Thank you again for submitting to *Nature Biomedical Engineering* your Article, "Closed loop enhancement and neural decoding of human cognitive control".

The manuscript has been seen by 3 experts, whose reports you will find at the end of this message, and please excuse the unusual delay in reaching a decision (one reviewer needed additional time to assess the work and provide a report / we had difficulties in securing reviewers).

You will see that although the reviewers have good words for aspects of the work, they raise serious technical concerns about the choice of the internal capsule/striatum for placement of the electrodes, and the testing paradigm for neuropsychiatric disorders and therefore they feel that the evidence to support the therapeutic claims is lacking. Having considered the reviewers' advice, we have reached the conclusion that the work is unlikely to provide the high technical quality and scientific significance that we look for in manuscripts that we consider for further external peer review.

We hope that you will find the referee reports helpful when revising the work.

Best wishes,

Michelle

__
Dr Michelle Korda
Senior Editor, *Nature Biomedical Engineering*

* Although we cannot offer to publish your manuscript in *Nature Biomedical Engineering*, the work may be appropriate for another journal published by Nature Research. Should you wish to explore suitable journals and transfer your manuscript to a journal of your choice, please use our manuscript transfer portal. This transfer link remains active until used. If you transfer your manuscript to Nature-branded journals or to the Communications journals, you will not have to re-supply manuscript metadata and files.

All Nature Research journals are editorially independent, and the decision to consider manuscripts are taken by their own editorial staff. For more information, please see our manuscript transfer FAQ page.

_____
Reviewer #1 (Report for the authors (Required)):

Basu et al. used epilepsy patients implanted with stereo-EEG electrodes to evaluate the neurophysiological basis of cognitive control and designed stimulation strategies to augment that control. This is a scientific study that leverages brain stimulation/recording opportunities provided by clinical practice. The long-term goal of the work is to develop “adaptive” DBS methods for modulating aberrant brain activity underlying neuropsychiatric disorders. The work presented in this study is a potentially useful step toward achieving that bigger goal. However, while not intending to minimize the clinical and technical difficulty associated with performing this study, the results are somewhat incremental, given several other recent publications from this group.

Specific Comments:

1) In my opinion the manuscript was written in a way that is too reliant on previous publications (from the authors, as well as others) to explain the point, purpose, and results of the study. My fear is that only a handful of experts in this hyper-specialized field would be able to follow this paper. Frankly, I consider myself one of those experts and I felt lost some of the time. An introduction figure may help provide background and context.

2) The anterior limb of the internal capsule is a complicated assortment of fiber pathways. Given that the clinical sEEG electrodes were not prospectively targeted to a specific pathway (or connectivity to a specific cortical territory) it is somewhat surprising that this study was able to identify a common stimulation area for enhanced cognitive control. However, this is an encouraging finding for the potential robustness of future DBS intervention attempting to employ the concepts of this study. Therefore, I suggest the authors highlight this “clinical limitation” of the study, and also provide some discussion (and preferably tractography results) on the neuroanatomical connections thought to be stimulated in the dorsal capsule of these patients.

3) However the possible clinical robustness point of comment 2, is also a pretty substantial scientific limitation of this study. Without a clear prospective hypothesis, this study comes across as a fishing expedition. Yes, you found some “random” places to stimulate/record that gave you an effect, but this was in a small number of subjects with lots of options provided by the sEEG electrodes. As such, it is unclear if the results would be reproducible if you were to try it with a prospective hypothesis on what you are trying to stimulate and where you are trying to record.

4) Clearly, cognitive control deficit is an important component of neuropsychiatric disease, but in my opinion the wording used in this manuscript opens itself up to the potential for lots of “unintended consequences” and “misinterpretation” by both clinicians and laymen from outside the field of neuromodulation. Is there really clinical justification for a permanently implanted DBS system for “treating cognitive control deficits” as a primary objective? How well would it need to perform to reach that justification? Even under the best intentioned circumstances these issues could easily be spun in a nefarious direction. Therefore, the discussion needs a dedicated paragraph about the ethical implications of the long-term goals of this project.

Reviewer #2 (Report for the authors (Required)):

The manuscript by Basu et al describes a study testing direct electrical stimulation in internal capsule/striatum in patients temporarily implanted with intracranial electrodes for epilepsy monitoring. Both open-loop and closed-loop stimulation were tested for effect on cognitive control (as assessed through reaction time in a task). The authors report improved cognitive control resulting from stimulation and changes in neural activity.

Critical details are lacking which make interpretation of the results difficult. The authors should provide additional details about the methods, patient traits, modeling parameters, threshold selection, etc. Importantly, in multiple places the authors overstate their findings and make claims that are not supported by their data. Some claims even border on deceptive and should be properly qualified.

• Statements in abstract and introduction about participant self-reports should either be removed or defended in the results section.
• Line 64-65: ‘with evidence of clinical utility’ is too strong – no clinically-validated metric was administered or tested. There is potential for clinical utility, but future studies are needed.
• While overall percentage of trials which were incorrect was low, provide statistical comparison that this percentage did not differ across stimulation conditions (specifically comparing open-loop with closed-loop stimulation, and closed-loop stimulation with no stimulation).
• In stimulated sessions, how was carry-over effect of stimulation between blocks avoided? How long were ‘brief rest periods’?
• Figure 2: Why do NS2 trials and stimulated trials have such different theta power responses? If stimulation is causing the change in NS2 trials, it should also be reflected in the stimulation trials (even if a ‘next trial’

effect, because there are consecutive trials of stimulation). How is stimulation artifact handled in stimulated trials?
• Figure S3 A and C: Cannot clearly see superimposed plot because of line thickness and/or number of trials plotted.
• Line 141-142: Elaborate on how threshold was determined
• In multiple places, closed-loop stimulation is stated to be more effective than open-loop stimulation. The subtleties of this are not stated clearly enough – CL stimulation was not different from OL stimulation in changing reaction times on conflict trials (presumably the trials in which cognitive control is challenged the most).
• Describe how we should interpret significance in xconflict (vs none) when there is not significance in conflict-related reaction times during OL stim vs no stim (Figures 3D vs S4B)
• What controls were conducted to ensure that changes attributed to closed-loop stimulation are not just regression to the mean?
• Line 183: value + confidence bound greater than 100%?
• Line 215-216: What available neural implant (full device) can simultaneously record from 6 brain regions and calculate 10 spectral features?
• Line 216: 'could easily be used in a non-structured setting' is an overstatement given the results provided
• Provide information on how often closed-loop stimulation was triggered; how did this compare to open-loop stimulation delivered to 50% of trials?
• Provide a table with patient traits – any comorbid neuropsychological diagnoses? Disorders of attention, anxiety, etc? Medication status? Location of seizure foci?
• Behavior data analysis: Justification for block level coding of stimulation (blockStim) needed beyond it provided a 'more parsimonious fit to the data'. At the very least show the testing of the fit for block coding and trial-by-trial coding.
• Provide wavelet parameters
• Was wavelet conducted on the continuous time-series data? If yes, what prevented stimulation artifact from bleeding into neighboring time periods or even trials?
• Line 657: Does this result stand if you take the channel that had the highest theta during NS1 trials?
• In stimulated sessions, is there a difference in reaction time and/or theta power between the first NS1 block (before the stimulation blocks) and the NS1 block at the end?
• What controls are included to determine whether changes were specific to cognitive control (vs general activation, mood effects etc). Perhaps including error rate in your model in addition to RT would be beneficial.
• More information is needed about how the state-space model was learned including training and validation.
• It appears that the neural decoding model was built on the training set, but that you then applied overfitting techniques (feature reduction) to the test set data. Variables were dropped if there was not a significant change in the RMSE. The model should be developed fully on the training set (could consider applying penalization to the coefficient vectors). More information on the significance of this model is also needed.

Reviewer #3 (Report for the authors (Required)):

The author studied a cognitive control task employing intracranial electrode epilepsy monitoring and electrical stimulation in the internal capsule/striatum. They compared the effects of closed-loop vs. open-loop stimulation on this particular cognitive control task. The study was carefully performed. However, I am concerned about the interpretation of the results.

Major criticism:
The manuscript implies that there is just one type of cognitive control, and an intervention that improves this "generic" cognitive control will have therapeutic value. I am surprised about several statements. Is there really just one type of cognitive control, just one target to fix all diseases related to the same generic impairment of the same generic cognitive control mechanism?

My concern is that the authors overclaim and oversell their results. The results provided in their manuscript will likely be interesting for cognitive neuroscientists. However, the therapeutic value remains to be shown (and not just claimed). Wouldn't it be possible that some patients might improve with respect to a particular, 'artificial' cognitive control impairment, while not having any clinically relevant improvement?! The authors should demonstrate that the stimulation protocol they use actually reduces symptoms. In my view, the results should be adequately presented in a more cognitive neuroscience-oriented framework without therapeutic claims. The scope of Nature Biomedical Engineering is not adequate for these results.
Relevant studies in that field, e.g., Wu et al., Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. PNAS 2018, should be discussed.

Thu 05 Jan 2021
Decision on Article NBME-20-0612B-Z

Dear Prof Widge,

Thank you for your revised manuscript, "Closed loop enhancement and neural decoding of human cognitive control", which has been seen by the original reviewers and one more expert, and we apologise for the delay in coming to a decision on the manuscript--we had difficulty finding the extra Reviewer. In their reports, which you will find at the end of this message, you will see that the reviewers acknowledge the improvements to the work and are essentially happy with the manuscript. We would therefore like to progress with an accept in principle provided that you address the points of Reviewer #3 and the minor textual points of Reviewer #4. In particular, we would expect that a revised version of the manuscript provides clarification and attenuation of the claims with regard to the clinical benefits of the technology with a discussion of the limitations of the current study as suggested by Reviewer #3.

As before, when you are ready to resubmit your manuscript, please upload the revised files, a point-by-point rebuttal to the comments from all reviewers, the (revised, if needed) reporting summary, and a cover letter that explains the main improvements included in the revision and responds to any points highlighted in this decision.

As a reminder, please follow the following recommendations:

* Clearly highlight any amendments to the text and figures to help the reviewers and editors find and understand the changes (yet keep in mind that excessive marking can hinder readability).

* If you and your co-authors disagree with a criticism, provide the arguments to the reviewer (optionally, indicate the relevant points in the cover letter).

* If a criticism or suggestion is not addressed, please indicate so in the rebuttal to the reviewer comments and explain the reason(s).

* Consider including responses to any criticisms raised by more than one reviewer at the beginning of the rebuttal, in a section addressed to all reviewers.

* The rebuttal should include the reviewer comments in point-by-point format (please note that we provide all reviewers will the reports as they appear at the end of this message).

* Provide the rebuttal to the reviewer comments and the cover letter as separate files.

We hope that you will be able to resubmit the manuscript within 12 weeks from the receipt of this message. If this is the case, you will be protected against potential scooping. Otherwise, we will be happy to consider a revised manuscript as long as the significance of the work is not compromised by work published elsewhere or accepted for publication at *Nature Biomedical Engineering*. Because of the COVID-19 pandemic, should you be unable to carry out experimental work in the near future we advise that you reply to this message with a revision plan in the form of a preliminary point-by-point rebuttal to the comments from all reviewers that also includes a response to any points highlighted in this decision. We should then be able to provide you with additional feedback.

We look forward to receive a further revised version of the work. Please do not hesitate to contact me should you have any questions.

Best wishes,

Michelle

Dr Michelle Korda
Senior Editor, *Nature Biomedical Engineering*

Reviewer #1 (Report for the authors (Required)):

the authors have addressed my primary concerns

Reviewer #2 (Report for the authors (Required)):

This revision is greatly improved. They have addressed my previous concerns and moderated their claims. They have added quite a bit of methodological and analysis detail.

Reviewer #3 (Report for the authors (Required)):

Re my point concerning “generic” cognitive control” and the authors’ response:
The manuscript implies that cognitive control is reasonably associated with a physiological process that can be controlled through closed-loop stimulation. Cognitive control is a broad concept which goes significantly beyond the specific experiment. The authors do not motivate why this specific experimental setup can be extrapolated to all sorts of processes where cognitive control is or might be relevant. I am concerned that (in the current form of the manuscript) the authors overclaim and oversell their results.
+++
Re “In fact, this is one of the great controversies of the RDoC model – whether the constructs it measures can be clinically useful.”
I am not sure whether this type of (“non-medical”) intervention is in the scope of Nature Biomedical Engineering.
+++
Re “We disagree with the implication that this means the present work is less valuable because it describes one useful intervention target. This is still the first demonstration that closed loop control is feasible in this cognitive domain, and through this target, in humans. The move to successful closed loop in humans was not possible in the referenced PNAS work, and thus is a significant advance.”
However, the Halpern study clearly demonstrates clinical benefit.

Reviewer #4 (Report for the authors (Required)):

This is a very interesting, though fairly complex, paper that explores the use of invasive neurostimulation to improve cognitive control in a conflict task. As is true of most human invasive brain recordings studies; the study is “opportunistic”; taking advantage of patients with externalized intracranial leads implanted for clinical purpose of epilepsy monitoring. There is a significant technical innovation in that the authors developed a model-based method for tracking cognitive control on a trial-by-trial basis; and used that to implement closed loop stimulation; eg stimulating when the prior trial showed a cognitive control deficit. Closed loop control was superior to open loop stimulation. Effects of stimulation were on the “baseline” reaction time (for nonconflict trials) and did not specifically improve performance (but didn’t detract from performance) in conflict trials nor in the state variable that modelled/tracked high conflict trials.
Theta band activity distinguished conflict/no conflict trials and was reduced by stimulation, consistent with the role of theta in cognitive control as assessed in prior, noninvasive eeg studies. However, the authors also did

an “encoding model” approach to predict the neural signal from cognitive performance and found a fairly wide range of frequencies were utilized in the encoding model, with theta having a less prominent role in the encoding. Potential reasons are discussed appropriately.

The overall importance of the work lies in the fact that Deficits in Cognitive control are widely felt to be important across a number psychiatric disorders, and neuromodulation based interventions for these are widely sought. The authors address the likely need to target aberrant brain activity in a temporally precise manner (closed loop); rather than focusing on spatial precision as most trials of neuromodulation have done. It is technically a careful study. This paper does provide a technical and conceptual framework for a sophisticated neurostimulation paradigm to improve cognitive control. So it is both a technical and conceptual advance.

I did not review the initial submission, but was asked to comment on the revised version and on response to reviewers. The authors have addressed most of the reasonable criticisms with substantial new analyses, and clarified several points that made the first submission a difficult read. Given that this is already a revised manuscript, I am not emphasizing suggested revisions, but I ask the authors to check on two minor points that may represent errors or typos (I do not find any major technical problems).

1)Figure 4d lacks any significance bars; is the effect of ventral stim meant to be shown as significant here?
2)Check line 262 – authors use reference 1 for a statement about existing DBS devices but reference 1 (“Mental Disorders Top The List Of The Most Costly Conditions In The United States: $201 Billion:”, may not relate to the properties of advanced brain sensing devices?

Mon 05 Jul 2021
Decision on Article NBME-20-0612B-Z

Dear Prof Widge,

Thank you for your patience in waiting for the feedback on your revised manuscript, "Closed loop enhancement and neural decoding of human cognitive control". As noted earlier in e-mail correspondence by my colleague Michelle Korda, having consulted with Reviewers #3 and #4 (whose comments you will find at the end of this message), I am pleased to write that we shall be happy to publish the manuscript in *Nature Biomedical Engineering*, provided that the points specified in the attached instructions file are addressed.

When you are ready to submit the final version of your manuscript, please upload the files specified in the instructions file.

For primary research originally submitted after December 1, 2019, we encourage authors to take up transparent peer review. If you are eligible and opt in to transparent peer review, we will publish, as a single supplementary file, all the reviewer comments for all the versions of the manuscript, your rebuttal letters, and the editorial decision letters. **If you opt in to transparent peer review, in the attached file please tick the box 'I wish to participate in transparent peer review'; if you prefer not to, please tick 'I do NOT wish to participate in transparent peer review'.** In the interest of confidentiality, we allow redactions to the rebuttal letters and to the reviewer comments. If you are concerned about the release of confidential data, please indicate what specific information you would like to have removed; we cannot incorporate redactions for any other reasons.
More information on transparent peer review is available.

Please do not hesitate to contact me should you have any questions.

Best wishes,

Pep

__
Pep Pàmies
Chief Editor, *Nature Biomedical Engineering*

_____
Reviewer #3 (Report for the authors (Required)):

Re 3. "Using an animal variant of this paradigm (an extradimensional set-shift), we have largely replicated/reverse-translated these behavioral effects. That is a very different task, and yet the same stimulation appears to produce comparable change.

Point (3) is still unpublished and beyond the scope of this human-focused article (we are working to complete the experiments and submit it separately), but the others needed to be motivated/explained as R3 said. We have added a further paragraph to the Discussion covering points (1) and (2)."
The manuscript should be self-consitent and not require support by unpublished data.

I agree that "substantial challenges remain before these results can be directly applied in the clinic".

Reviewer #4 (Report for the authors (Required)):

I had only 2 minor concerns in the last round, which i believed the authors have addressed

Rebuttal 1

Reviewer #1

1) In my opinion the manuscript was written in a way that is too reliant on previous publications (from the authors, as well as others) to explain the point, purpose, and results of the study. My fear is that only a handful of experts in this hyper-specialized field would be able to follow this paper. Frankly, I consider myself one of those experts and I felt lost some of the time. An introduction figure may help provide background and context.

We agree that in trying to keep the Introduction brief, we spent too little time overviewing the literature of cognitive control, its linkage to cortico-striatal circuitry, and the role of model-based analyses. We have substantially revised and expanded the Introduction and Methods to explicitly describe what the known barriers to our technology were, and why each step in our solution was needed. These revisions have in particular focused on clarifying the modeling approach without requiring/expecting the reader to have read our specific prior work or the state-space literature. We did retain references to our prior work where it was necessary to highlight that a given analysis was hypothesis-driven, pre-specified, or designed to replicate prior work (see point below).

We have added the introduction/flowchart panel as suggested, in Figure 1A.

2) The anterior limb of the internal capsule is a complicated assortment of fiber pathways. Given that the clinical sEEG electrodes were not prospectively targeted to a specific pathway (or connectivity to a specific cortical territory) it is somewhat surprising that this study was able to identify a common stimulation area for enhanced cognitive control. However, this is an encouraging finding for the potential robustness of future DBS intervention attempting to employ the concepts of this study. Therefore, I suggest the authors highlight this "clinical limitation" of the study, and also provide some discussion (and preferably tractography results) on the neuroanatomical connections thought to be stimulated in the dorsal capsule of these patients.

This is a critical point on which we agree with R1. We do not think it is at all surprising that we found common effects across patients. The internal capsule is a mix of many fiber bundles, but the very extensive human/macaque work of Susanne Haber (and some of our own tractography work, e.g. Makris et al. Brain Imaging & Behavior 2016) shows that these bundles obey topographic rules. Specifically, there is a strong tendency for cortico-thalamic fibers from more dorsal and lateral PFC to run dorsally within the capsule, and those from more ventral/medial PFC to run ventrally within the capsule. Thus, just as one can reliably expect that a certain zone of M1 is more likely to be linked to hand vs. leg function, one can similarly expect, even without patient-specific imaging, that a given part of the capsule is likely to contain fibers originating in a specific sub-zone of PFC. This is argued to be the reason why deep brain stimulation is possible and potentially effective within this white matter target (see, e.g., the just-released summary in https://doi.org/10.1016/j.biopsych.2020.06.031).

This was the reasoning behind our choice to explore dorsal vs. ventral stimulation – we hypothesized that these two sites would capture sufficiently different subparts of cortico-thalamic or cortico-striatal circuitry to have different behavioral effects. (This may be akin to the idea of a specific cortical territory that R1 mentions.)

Relevant to this work, cognitive control (and especially the type we measured here, cognitive conflict task performance) is linked to more dorsal PFC, especially DLPFC and cingulate. Those fibers run more dorsally in the capsule in most individuals. As such, we expected that dorsal stimulation would have larger behavioral effects. We agree that not discussing this topography and the resulting predictions was a major omission/failure, and we apologize for it. We have added material to the Introduction, Discussion, and Methods to overview this literature and place our results into context.

We do not have patient-specific tractography, as there was no clinical indication for such scans in these participants. We have placed our results in tractographic context through the literature just mentioned. As further evidence that this topography/tractography applies in our participants, we now include data on cortical evoked responses to capsule stimulation. Figure S3 shows that, as would be expected, more dorsal stimulation produced larger activations (evoked potentials) in DLPFC and anterior cingulate.

3) However the possible clinical robustness point of comment 2, is also a pretty substantial scientific limitation of this study. Without a clear prospective hypothesis, this study comes across as a fishing expedition. Yes, you found some "random" places to stimulate/record that gave you an effect, but this was in a small number of subjects with lots of options provided by the sEEG electrodes. As such, it is unclear if the results would be reproducible if you were to try it with a prospective hypothesis on what you are trying to stimulate and where you are trying to record.

We agree fully that it is important to have clear prospective hypotheses in small-N, rare population studies such as these. To that point, we wish to politely but firmly state that this specific critique is factually incorrect. This study was conducted under a prospective hypothesis, and the chosen stimulation sites were not in any way random. In the Methods section of the originally submitted manuscript, within the first paragraphs, we very clearly stated this:

> "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[17]. 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 Figure 2 were specified to replicate the prior study and demonstrate that its effects were robust to changes in study population and stimulation details."

That is: we did not attempt MSIT-linked stimulation in any other sites that are not reported here. We did so in other tasks and experiments conducted with the same participants, e.g. https://www.biorxiv.org/content/10.1101/825893v1 , but those were separate experiments, conducted by different lead investigators, with different hypotheses.

Put another way, we deliberately did not explore the "lots of options", because that would be problematic in exactly the way that R1 states. Similarly, our recording analyses (described starting at section 8 of the methods) were explicitly designed to replicate the findings of a prior study, and in that paragraph we describe multiple

ways that our analyses directly align with that prior work. (As under critique 1, we agree that this does cause the paper to make substantial reference back to other works, but we believe that emphasizing the degree of rigor here is important enough to justify that style flaw.) We thus do not believe this critique is valid/fair.

R1 is correct, however, that we failed to explain why we tested a dorsal-ventral split. This relates back to the prior point – we tested this because prior anatomic studies of the capsule predict a difference in the circuitry modulated even by this small shift. We had this model in our minds, but did not write it into the paper, and this was a major oversight for which we apologize. In the Introduction, Methods, and Discussion, we now make this model and expectation explicitly clear.

4) Clearly, cognitive control deficit is an important component of neuropsychiatric disease, but in my opinion the wording used in this manuscript opens itself up to the potential for lots of "unintended consequences" and "misinterpretation" by both clinicians and laymen from outside the field of neuromodulation. Is there really clinical justification for a permanently implanted DBS system for "treating cognitive control deficits" as a primary objective? How well would it need to perform to reach that justification? Even under the best intentioned circumstances these issues could easily be spun in a nefarious direction. Therefore, the discussion needs a dedicated paragraph about the ethical implications of the long-term goals of this project.

There are two points here, and both are valid. First, that we need to explain more about what this clinical system would look like and why we think it would be reasonable. This is to our minds the same critique raised by R3. Second, that because there are significant neuroethical implications, we need to formally describe these and acknowledge the potential societal harms from misuse of the technology. We have added a paragraph to the Discussion to cover these points.

We do believe that there could be substantial value in a system that directly addresses cognitive deficits, with cognitive control as one potential working model of that idea. First, we have suggested elsewhere (in our 2019 Nature Communications paper) that remediation of such deficits is actually a mechanism of action when DBS of these same regions is used to treat depression or OCD. That suggestion is supported by a recent study of Tyagi et al. (Biological Psychiatry 2019), who also found improvements in set-shifting (another form of cognitive control) during DBS for OCD, using the limbic STN as an entry point. The Haber paper we referenced above, plus another recent work from Andreas Horn and colleagues (https://www.nature.com/articles/s41467-020-16734-3), argue that this common effect derives from the fact that both targets can access the same pathways. Put another way: we believe clinicians are already using DBS (and VNS, and TMS) to improve cognitive control, but are not explicitly naming that as what we are doing. Second, and closely related, there is a strong sense from thought leaders in clinical and research psychiatry that a cross-diagnostic focus on cognition and decision-making is the clearest way to break through decades of stalled clinical progress. This idea is embodied in NIMH's Research Domain Criteria (https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/index.shtml), and is sufficiently supported that it is woven throughout the newly released NIMH Strategic Plan (https://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml). The idea of targeting decision-making is also the heart of the emerging concept of "computational psychiatry". This is covered in Wang & Krystal Neuron 2014, or the book of the same title

by Redish & Gordon (2016). Third, there is sufficient clinical justification for cognitive-targeted neuromodulation for there to be a company based entirely around the idea of an implant to boost memory (https://niatherapeutics.com/). That company is successfully raising money and building its first prototype (https://medium.com/@dan_84414/nia-therapeutics-raises-1-5m-to-finalize-design-of-device-for-restoring-memory-50fe8fcde20a). We emphasize that **these factors, collectively, are why this paper is timely and of broad interest**. A great deal of money and labor is flowing into this idea of improving decision processes; we believe this is the first evidence that those ideas can be turned into a workable technology with potential clinical applications.

To the ethical points, we have previously worked with neuroethicists to elicit patient perspectives on this type of closed-loop therapy, and have shown that there is interest and acceptance from the people who would be most affected (Goering et al. AJOB Neuroscience 2017; Klein et al. Brain-Computer Interfaces 2016). As such, there is a clear case for potential societal benefit through relief of otherwise intractable disease. There is a larger question of whether our results can be a basis for cognitive enhancement of otherwise neurotypical persons. If they could, this would raise significant ethical questions about whether it would ever be acceptable for a surgeon and team to provide such enhancement, and if it were acceptable, how access to such a technology could be provided in a socially just/equitable fashion. We agree that we needed to explicitly acknowledge and discuss those potential negative aspects/harms from our technology, and we have done so.

On the other hand, we suggest that this concern of R1's directly rebuts his/her prior statement that this work may be incremental. If our demonstration of closed-loop augmentation raises this level of concern for misinterpretation and misuse, that indicates that it is (A) a major technical advance not previously demonstrated, (B) not trivial to perform, and (C) of substantial interest to a broad technical and broader societal audience.

Reviewer #2
Critical details are lacking which make interpretation of the results difficult. The authors should provide additional details about the methods, patient traits, modeling parameters, threshold selection, etc. Importantly, in multiple places the authors overstate their findings and make claims that are not supported by their data. Some claims even border on deceptive and should be properly qualified.

Broadly, we agree that these additional details would strengthen the paper, and that some claims in the original manuscript were overly strong. We have made numerous revisions, detailed below.

• Statements in abstract and introduction about participant self-reports should either be removed or defended in the results section.

We agree that these could be removed from the abstract and introduction without substantially altering the paper, and have done so. We think they are sufficiently interesting to retain in the Results, in that they are actual unexpected statements made by participants that may help understand the underlying value of our work. In the course of revising Results and Discussion in response to concerns above, we have also clarified that these are from a small subset of the overall study, i.e. not necessarily reflective of the majority of participants.

• Line 64-65: 'with evidence of clinical utility' is too strong – no clinically-validated metric was administered or tested. There is potential for clinical utility, but future studies are needed.

Agreed. We have markedly softened this language, and there is a large paragraph in the Discussion providing an overview of the technical gaps remaining between this study and clinical use.

• While overall percentage of trials which were incorrect was low, provide statistical comparison that this percentage did not differ across stimulation conditions (specifically comparing open-loop with closed-loop stimulation, and closed-loop stimulation with no stimulation).

We have edited Figure S2 and its caption to show this statistical comparison, including the specific comparisons R2 requests. No condition differs from the non-stimulated trials in terms of accuracy.

• In stimulated sessions, how was carry-over effect of stimulation between blocks avoided? How long were 'brief rest periods'?

In the Methods we now report these rest periods; median of 5.25 minutes and almost 90% of them under 10 minutes. The range extends to 57.6 minutes for some rare cases where experiments were interrupted by multiple events (e.g., participant wanting to eat lunch, then to use toilet, followed by a family or clinical team meeting). After these long interruptions, we re-initiated the experiment with a fully non-stimulated block before returning to stimulation.

We agree that these short breaks do raise a risk of carry-over; they were necessary due to the compressed experimental schedule created by a limited window of clinical opportunity. To address this, Figure S9 now plots the available examples of two types of potential carry-over: from effective stimulation (dorsal capsule) to unstimulated blocks, and from the most effective stimulation to statistically ineffective stimulation (transitions from right-sided dorsal to left-sided ventral stimulation). In the first case, there is no carry-over; the state x_{base} immediately begins to increase. In the second, there likely is, as it seems to take >10 trials for x_{base} to move visibly upward. The more important point, though, is that any carry-over would actually reduce the statistical significance of our results, i.e. that it cannot produce false or spurious positives. Since the core claim in Figure 2 is a difference between the median RTs of block types relative to the non-stimulated condition, carry-over would blur/obscure that difference. In practice, this was likely mitigated by our varying the order of block types between participants.

• Figure 2: Why do NS2 trials and stimulated trials have such different theta power responses? If stimulation is causing the change in NS2 trials, it should also be reflected in the stimulation trials (even if a 'next trial' effect, because there are consecutive trials of stimulation). How is stimulation artifact handled in stimulated trials?

We focus these analyses on sites far from the stimulation, where the volume-conducted artifact is non-zero, but small enough that we did not see amplifier saturation We agree that it is curious that we do not see the theta increase until the following NS2 trial. Our working theory is that the acute burst of high-frequency stimulation disrupts rhythmic activity, a disruption that has often been argued to be a therapeutic mechanism of deep brain stimulation. It is not physiologically clear why this would lead to a theta increase on the following trial. One possibility is that a temporary disruption could then phase-reset/entrain the cortical microcircuits that generate PFC theta, making them more synchronous with each other. This would be reflected in higher apparent power (less destructive interference). The challenge with that model is that it does not clearly explain other channels (discussed below) where theta decreases during stimulated blocks. This suggests a more complex shift in network configurations is the actual mechanism. We are actively exploring these deeper mechanistic questions in ongoing work.

Practically, because these stimulated trials are not relevant to the overall point of the paper (we never analyze them further or use them to support any of our overall claims), we have removed the referenced element from Figure 2.

• Figure S3 A and C: Cannot clearly see superimposed plot because of line thickness and/or number of trials plotted.
We have adjusted the Figure (which is also renumbered to S6) and have added an inset for clarity.

• Line 141-142: Elaborate on how threshold was determined

We believe this is a request for elaboration on the threshold for closed-loop control? We have expanded that discussion in the Methods, particularly of the subjective nature of this threshold. One experimenter (ASW) set the threshold manually, such that driving the mean state to that threshold would drive the overall RT outside its observed bounds during non-stimulation experiments. We have noted this in the Discussion as an inherent limitation of working with epilepsy participants.

• In multiple places, closed-loop stimulation is stated to be more effective than open-loop stimulation. The subtleties of this are not stated clearly enough – CL stimulation was not different from OL stimulation in changing reaction times on conflict trials (presumably the trials in which cognitive control is challenged the most).

R2 is correct that we saw less change on Conflict trials -- but that was expected. In our formulation, and in the paper we sought to replicate/extend (Widge et al. Nature Communications 2019), the effect was <u>not</u> expected to be greater in conflict trials. We specifically expected the effects to load onto x_{base}, and designed this two-state model in part to detect that same effect. We have expanded the Methods, particularly in section 5, to make this expectation clear. In that regard, the effect <u>was</u> greater in x_{base} during closed loop control (Figure 4B), hence we believe the statement is justified.

The subtlety, however, is that we saw a larger effect from CL stimulation, in part because the controller gave more stimulations than were given in the OL block. This motivated the analysis of Figure 4D, where we examined whether we saw a larger change per unit energy (i.e., per delivered stimulation train). As noted in the original manuscript, we did observe this greater efficiency, but it did not reach the pre-determined statistical significance. (We do not feel it would be appropriate to describe this as a "trend".) We have added sentences to the Results, at line 181, to clarify this further.

• Describe how we should interpret significance in xconflict (vs none) when there is not significance in conflict-related reaction times during OL stim vs no stim (Figures 3D vs S4B)

Essentially, $x_{conflict}$ in Figure 3D can be thought of as a "denoised" version of the raw RT shown in Figure S4B (which is now Figure S10). That is, we presume that the RT on a conflict trial is composed of the raw task RT (x_{base}), the specific slowing in response to conflict ($x_{conflict}$), and noise processes unrelated to the decision (e.g., variability in the time to perceive the stimulus or execute the motor command, what some modelers call the "nondecision time"). The points analyzed in Figure S10 include that noise. As such, they have a lower SNR compared to the points in Figure 3D, where the noise is (at least, in theory) removed in the process of extracting $x_{conflict}$. (Figure S6 shows that this <u>is</u> noise, i.e. that it is a spectrally white Gaussian process.) It is not at all surprising that adding the noise back destroys statistical significance – it is by definition a reduction in statistical power. We included this analysis to address concerns that have come up in others' attempts at model-based closed-loop stimulation (e.g., Ezzyat et al. Current Biology 2017) and in informal discussion of this work at conferences. It has been argued that showing change in a model/prediction/derived variable, without also showing manifest/raw data, is deceptive or may create an impression of significant change even without behavior change. This figure shows that we <u>do</u> successfully improve raw RT, and that use of the state-space models merely clarifies which components of conflict decision-making/cognitive control are being altered.

We have added text to the caption of Figure S10 to clarify this point.

• What controls were conducted to ensure that changes attributed to closed-loop stimulation are not just regression to the mean?

Mathematically, regression to the mean cannot explain the results in Figure 4B, and the concept may not be applicable. This is a block-level analysis. That is, it represents the overall distribution of RTs during closed-loop blocks -- it does <u>not</u> reflect the RTs of trials immediately after delivery of individual closed-loop stimulation events. This is why we do not think the concept of regression applies. If we stimulated on a particularly slow trial, but the participant then returned to a stable mean (with no effect of stimulation), then the mean performance across blocks would not differ. Further, the performance would be identical to that of our open-loop experiments. Figure 4B specifically shows that performance during closed-loop was significantly faster than open-loop, as well as faster than baseline.

We further note that we were able to force RTs (and their smoothed version, x_{base}) outside the range that they "naturally" occupy. This relates to the point above about the threshold -- it was selected specifically to permit such a demonstration.

As a final point, we performed a stimulation-triggered averaging analysis on the behavior (Figure S11). Specifically, for the 3 patients who performed CL experiments, we identified points in their prior (unstimulated) days of behavior at which the CL controller would have stimulated, had it been active. We then averaged the tracked behavioral variable (x_{base}) for the trials immediately following these "virtual detection events", and compared them to x_{base} after actual detections during the closed-loop experiment. These curves are dramatically different, being almost flat after "virtual" detections but dropping during the actual CL experiment. If all we were observing was regression to the mean, then the "virtual" curve would also sharply decrease.

• Line 183: value + confidence bound greater than 100%?

This is an inherent limitation of the use of the standard deviation as a measure of spread/confidence interval when the mean is near the edge of a truncated range. We have added a clarifying statement.

• Line 215-216: What available neural implant (full device) can simultaneously record from 6 brain regions and calculate 10 spectral features?

This is possible with the Medtronic RC+S, and with Cortech's Brain Interchange System (which exists and is in trials, but is not marketed). We have added a reference to the former; the latter is not described in a publication to the best of our knowledge. It is our belief that this may also be possible with research features of the Percept PC based on Medtronic communications, but that is not verified.

As a more esoteric point, this is a subject specific decoder, i.e. not every patient would need all 6 brain regions active at once. There are two different clinical trials underway of approaches to constructing such reduced decoders, where they begin by a large SEEG-style implant, identify a few electrodes with strong decoding capability, then convert only those sites to chronic implant. One is the PRESIDIO trial at UCSF (https://clinicaltrials.gov/ct2/show/NCT04004169), the other is an NIH UH3 between Baylor and UCLA (https://clinicaltrials.gov/ct2/show/NCT03437928). Both have already enrolled patients.

• Line 216: 'could easily be used in a non-structured setting' is an overstatement given the results provided

We have removed this sentence.

• Provide information on how often closed-loop stimulation was triggered; how did this compare to open-loop stimulation delivered to 50% of trials?

This was a major omission; it was reported in an early draft but lost during edits to reduce word count. As noted above, this is over 50% and we have added it back in the text. The fact that the controller stimulated in over 50% of trials led to the efficiency analysis of Figure 4D (discussed more above).

• Provide a table with patient traits – any comorbid neuropsychological diagnoses? Disorders of attention, anxiety, etc? Medication status? Location of seizure foci?

Now provided as Supplementary Table S1.

• Behavior data analysis: Justification for block level coding of stimulation (blockStim) needed beyond it provided a 'more parsimonious fit to the data'. At the very least show the testing of the fit for block coding and trial-by-trial coding.

AIC numbers are now included in the Methods at the specified point (-449.3 for the blockStim coding vs. -359.7 for trial-by-trial coding, i.e. a fairly large difference).

• Provide wavelet parameters

These are now in the methods. As noted, we will also provide the analysis code directly via GitHub. This was an entirely standard FieldTrip-based decomposition using the default parameters of ft_freqanalysis.

• Was wavelet conducted on the continuous time-series data? If yes, what prevented stimulation artifact from bleeding into neighboring time periods or even trials?

No; it was performed on the trial-structured data. The edges of those windows were discarded to reduce edge effects. We have clarified this in the Methods.

Even if we had done the wavelet decomposition on continuous data, it is mathematically impossible for the artifact to bleed into the NS2 trials that are our primary data. Consider the 5-8 Hz theta signal that we treated as our primary physiologic outcome. A wavelet covering 7 cycles at the slowest frequency (5 Hz) is only supported over 1400 ms. The majority of the support is of course in a much smaller window around the center of the Morlet, i.e. over a few hundred ms. Each trial was at least 3.75 seconds long (counting the ITI before the next trial). As such, there is at least one complete wavelet window with absolutely no stimulation artifact present before the next NS2 trial begins and is analyzed.

The artifact does bleed into the remainder of the stimulated trial in this analysis strategy, which was another reason to focus our analysis entirely on NS2 trials that we know are artifact-free.

• Line 657: Does this result stand if you take the channel that had the highest theta during NS1 trials?

The result is, as hinted, sensitive to which channels are picked. Some channels show an increase from NS1 to NS2, whereas others show a decrease. We agree this is worth illustrating, and have added Figure S5. There are two key points from that figure:

1) In the three stimulation conditions that were behaviorally effective (R Dorsal, L Dorsal, R Ventral), more channels show a theta increase than show a decrease, specifically in PFC areas affiliated with cognitive control: DLPFC, VLPFC, and to a lesser degree, DMPFC and ACC. DLPFC and VLPFC were also the sites of the largest theta increase in our 2019 EEG x DBS paper, i.e. these results are concordant with our prior study.
2) This same pattern is absent in the stimulation condition (L Ventral) that was behaviorally ineffective, supporting our broad framework (and prior finding) that theta power is a key physiologic correlate of cognitive control.

• In stimulated sessions, is there a difference in reaction time and/or theta power between the first NS1 block (before the stimulation blocks) and the NS1 block at the end?

This analysis is now in Figure S4 in the revised supplement. There is a difference, in the exact opposite of what would be expected if stimulation effects were explainable by either regression to a mean or practice effects. That is, RT is significantly higher and theta significantly lower in the final NS1 block compared to the first -- the opposite of the effect we report in the intervening stimulation blocks. We suspect that this is a fatigue effect, and discuss that in the caption of the above-mentioned figure.

Thank you for suggesting this analysis; we believe it adds to the robustness/believability of our claims.

• What controls are included to determine whether changes were specific to cognitive control (vs general activation, mood effects etc). Perhaps including error rate in your model in addition to RT would be beneficial.

We have verified that adding error rate (accuracy) will not change the results. As noted by the non-significant results in Figure S2 (also discussed above), there are simply too few errors, in any block, to meaningfully detect differences between block types. To illustrate this, we re-fit the state space model in all subjects, but in the data likelihood step, we also made error rates contingent on x_{base} and $x_{conflict}$ (in the core model, they are not included). The state variables inferred under this model are nearly identical to those inferred without using error information (median r=0.99 between x_{base} values inferred under the two different assumptions). We illustrate this additional analysis in Figure S8.

We have added a sentence to the Discussion noting that our results cannot be explained fully by mood, because most of our participants did not note changes in mood. Motivational factors, or general increases in the perceived efficacy/value of control, are a possible explanation for our results. We are currently working on

more advanced task designs that can dissect these explanations. (That is a very non-trivial problem, to the point that it is the topic of a major R01-level grant currently under review.) We also mention this in the Discussion, where we discuss alternate models of cognitive control.

• More information is needed about how the state-space model was learned including training and validation.

In terms of model structure, this was pre-selected based on our hypotheses about stimulation effects and our understanding of this specific task. We describe this more fully in the Methods, in the modeling segment of Methods section 5. We have added a description of the underlying EM process and how we monitored convergence, and have included plots of that convergence in Figure S7. These are in addition to the model structure validation plots included in Figure S6. We do not believe the concept of training/test set validation applies here, as the behavioral state-space model is not intended to be predictive. We address validation of the predictive model (the decoder) in the response to R2's next critique.

• It appears that the neural decoding model was built on the training set, but that you then applied overfitting techniques (feature reduction) to the test set data. Variables were dropped if there was not a significant change in the RMSE. The model should be developed fully on the training set (could consider applying penalization to the coefficient vectors). More information on the significance of this model is also needed.

This is a reasonable stipulation. We have implemented it within the limits of this dataset, namely the relatively small amount of data available for each participant due to the unique setting/time limits. We achieved this by a train-validate-test split. The initial encoder model is trained on one set of trials (up to 66% of the original dataset). The remaining trials were divided into 2 subsets for validation and testing, each of which used part of the remaining data. (To ensure sufficient data for feature reduction, the validation dataset overlapped the training dataset; see Methods for details.) We emphasize that for all participants, the testing dataset was completely disjoint from the training and feature reduction datasets. We agree that this is a more appropriate practice, and as such, we report it in place of the original Figure 5. In support of the robustness of our original findings, this revised train-test splitting procedure altered which frequency bands were heavily weighted in the decoder, but did not noticeably change decoding performance (e.g., 84.6 +/- 11% overlap between behaviorally and neurally decoded x_{base} in the original manuscript, 84.02 +/- 15.8% in the revised procedure). Similarly, our findings that stimulation seems to shift encoding from PFC towards basal ganglia remain unchanged.

Regarding "significance of this model", we believe that R2 is asking whether the decoding/prediction can be considered to be significantly better than chance? We reported an approach to this verification (and showed that the variable selection procedure does select true neural correlates of behavior) in the original paper describing this decoding approach (Yousefi, Basu, et al. Neural Computation 2019). Briefly, this involves re-fitting the decoder with random/nonsense behavior trajectories and demonstrating that after the feature pruning procedure, few to no neural variables are considered correlates of the noise. We report this analysis in Figure S13. For all

participants, the number of features selected (determined to be behaviorally correlated) on the true dataset was outside the highest density interval of the distribution of feature selection on shuffled/random data.

Reviewer #3
Major criticism:
The manuscript implies that there is just one type of cognitive control, and an intervention that improves this "generic" cognitive control will have therapeutic value. I am surprised about several statements. Is there really just one type of cognitive control, just one target to fix all diseases related to the same generic impairment of the same generic cognitive control mechanism?

If we understand this concern correctly, it is a desire to acknowledge that cognitive control can be measured more than one way, and may have component subprocesses (reviewed very well in, e.g., Inzlicht et al. Trends in Cognitive Sciences 2018). That is a very reasonable request. We used a standard measurement of cognitive control (see https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/cognitive-control.shtml and in particular the subconstruct of "performance monitoring"), in part because we have previously shown this metric to be sensitive to internal capsule stimulation and sought to replicate/extend those results. (See response to R1 above.) There are certainly other tasks that also get at this construct, and we have not shown what will happen if we stimulate during those other tasks. Similarly, there remain many open questions about how to invert this problem and identify patients who have strongly impaired cognitive control (by any measure) as a core source of their distress. We have added two paragraphs to the Discussion that address these points.

That said, we cannot locate any point in the original manuscript where we make the claim that R3 suggests, namely that there is "just one target to fix all diseases". In fact, we specifically cited our own results producing similar changes in a Flanker task using DLPFC stimulation (Dubreuil-Vall et al. 2019). We have further added citation to the work of Tyagi et al., who report changes in a somewhat different metric of cognitive control using STN stimulation, and to other work mentioned below.

My concern is that the authors overclaim and oversell their results. The results provided in their manuscript will likely be interesting for cognitive neuroscientists. However, the therapeutic value remains to be shown (and not just claimed). Wouldn't it be possible that some patients might improve with respect to a particular, 'artificial' cognitive control impairment, while not having any clinically relevant improvement?! The authors should demonstrate that the stimulation protocol they use actually reduces symptoms. In my view, the results should be adequately presented in a more cognitive neuroscience-oriented framework without therapeutic claims.

We firmly agree that it is important to highlight that, although a few participants made interesting statements, there is a wide gap (entailing formal clinical trials) between these present results and a demonstrated clinical therapy. In fact, this is one of the great controversies of the RDoC model – whether the constructs it measures can be clinically useful. The paragraphs noted above that were added to the Discussion include this point, as

well as R1's points about neuro-ethical implications. Throughout the paper, we have also altered language to emphasize that we improved performance on one specific cognitive control task.

At the same time, we do believe that highlights the value of this work. One of the problems with linking the RDoC constructs to clinical outcomes is the lack of specific ways to manipulate or target those constructs. Approaches like ours are a potential way to target specific cognitive systems, which can then become tools to answer the question "is it clinically useful to target this system?" R3 actually makes this exact point in his/her next comment:

Relevant studies in that field, e.g., Wu et al., Closing the loop on impulsivity via nucleus accumbens delta-band activity in mice and man. PNAS 2018, should be discussed.

It should; thank you for catching this and that work is now cited.

We note that this citation in many ways makes/supports our point about the potential value of targeting cross-diagnostic concepts such as cognitive control. In the referenced paper, Casey Halpern and colleagues directly targeted the construct of impulsivity, which they describe as "one of the most pervasive and disabling features common to many disorders of the brain". In that paper alone, they argue that this delta-band marker and single stimulation target have potential to treat "eating disorders, and even obesity and addiction". In presentations, Dr. Halpern has further stated that he believes a similar approach can be applied to treat obsessive compulsive disorder through stimulation of the same target. In other words -- the work R3 asks us to cite applies the same idea/framework that he/she describes above as "just one target to fix all diseases related to the same generic impairment". In some sense, this is necessarily part of any paper that attempts to discover neurostimulation approaches that precisely target cognitive domains.

Practically, we, R3, and the authors of that PNAS paper are probably all in complete agreement that there is no single master target for psychiatric disorders, or even for any specific cognitive construct. We have included that point in the expanded Discussion also.

We disagree with the implication that this means the present work is less valuable because it describes one useful intervention target. This is still the first demonstration that closed loop control is feasible in this cognitive domain, and through this target, in humans. The move to successful closed loop in humans was not possible in the referenced PNAS work, and thus is a significant advance.

Rebuttal 2

For clarity and brevity, we have removed comments that are mainly complimentary or brief (although we are still very grateful for them).
We have accepted the tracked changes from the prior version and tracked the new changes in this revision.

Reviewer #1/2 had no additional comments.

Reviewer #3

Re my point concerning "generic" cognitive control" and the authors' response:
The manuscript implies that cognitive control is reasonably associated with a physiological process that can be controlled through closed-loop stimulation. Cognitive control is a broad concept which goes significantly beyond the specific experiment. The authors do not motivate why this specific experimental setup can be extrapolated to all sorts of processes where cognitive control is or might be relevant. I am concerned that (in the current form of the manuscript) the authors overclaim and oversell their results.

The point about cognitive control being broad is very fair, as is the point that we need to explain why we think this one specific task has implications for what might happen with other paradigms (Stroop, Flanker, AX-CPT, etc.) that also are often used to measure the same construct.
We do believe that is true, for multiple reasons:

1. As noted, this work is derived from a prior study, where we used a variant of the same task (with emotional distractors) and a very different stimulation paradigm, but observed essentially the same results during open-loop stimulation. That is a modest generalization, but still evidence supporting the claim.
2. There is a fairly rich literature showing activation of dACC/dlPFC in fMRI, and theta oscillations over mid-frontal cortex (and/or originating from PFC in invasive studies) with multiple cognitive control tasks. That is, regardless of the paradigm, these tasks have a convergent behavioral effect (slowing of RT on harder trials) and a convergent physiologic signature. Our results are consistent with that signature, which adds confidence about generalizability.
3. Using an animal variant of this paradigm (an extradimensional set-shift), we have largely replicated/reverse-translated these behavioral effects. That is a very different task, and yet the same stimulation appears to produce comparable change.

Point (3) is still unpublished and beyond the scope of this human-focused article (we are working to complete the experiments and submit it separately), but the others needed to be motivated/explained as R3 said. We have added a further paragraph to the Discussion covering points (1) and (2).

Re "In fact, this is one of the great controversies of the RDoC model – whether the constructs it measures can be clinically useful."
I am not sure whether this type of ("non-medical") intervention is in the scope of Nature Biomedical Engineering.

We respectfully argue (and we believe the Editors agree, given the acceptance in principle) that clinically important engineering advances often need to be demonstrated in a more controlled, non-medical setting first, and that a technological advance in that context is important even if it has not yet advanced to a clinical trial.
We agree, as noted in our prior response to R3, that we need to be very clear about the gap between this current result and a clinical/medical effect in a psychiatric disease. The article specifically includes the statement "substantial challenges remain before these results can be directly applied in

the clinic", prominently in the concluding paragraph. We have also added a cautionary point in the new Discussion paragraph.

Re "We disagree with the implication that this means the present work is less valuable because it describes one useful intervention target. This is still the first demonstration that closed loop control is feasible in this cognitive domain, and through this target, in humans. The move to successful closed loop in humans was not possible in the referenced PNAS work, and thus is a significant advance." However, the Halpern study clearly demonstrates clinical benefit.

To the best of our knowledge, a report of clinical benefit is not contained in the article referenced (Wu et al *PNAS* 2018), where closed-loop stimulation was only performed in mice. We know of unpublished results from Dr. Halpern's team that are moving towards clinical benefit in loss of control eating, but as far as we can tell (including a literature search conducted in January 2021), that work remains in development and not part of the citeable peer-reviewed literature.

There is, however, a published report of the protocol for a human clinical trial based on that 2018 paper, and we have cited that report (and briefly expanded the Discussion) to emphasize that the Halpern result has indeed progressed to clinical use.

Reviewer #4:

1)Figure 4d lacks any significance bars; is the effect of ventral stim meant to be shown as significant here?

The lack of significance bars/stars is correct. These effects were numerically in the desired direction, but did not reach pre-specified significance thresholds. Lines 184 onward discuss this point.

2)Check line 262 – authors use reference 1 for a statement about existing DBS devices but reference 1 ("Mental Disorders Top The List Of The Most Costly Conditions In The United States: $201 Billion:", may not relate to the properties of advanced brain sensing devices?

This was indeed a reference manager glitch (failure to update a link) and is fixed. Thank you.