Experimental Manipulation of the Orbitofrontal Cortex Impacts Short-Term Markers of Human Compulsive Behavior: A Theta Burst Stimulation Study

Supplemental Information

Supplemental Methods & Materials

Randomization and adequacy of blinding

The random allocation sequence was generated at the start of the study (without blocking) using a random number generator. A single research assistant, responsible for administering the assigned TBS intervention but not involved in screening/eligibility or outcome assessments, maintained the randomization file and assigned participants to one of the two treatment conditions sequentially according to the scheme, upon arrival at the participant's first baseline/sham TBS visit. Clinical interviewers completed a forced-choice guess (iTBS or cTBS) at the 1-week follow-up assessment visit. This measure confirmed that the clinical assessor who was responsible for administering the laboratory probe of compulsive behaviors and other clinical interview measures was unaware of the participant's treatment allocation (59.7% overall accuracy; Fisher's exact for treatment allocation * clinician's guess p=.20).

Participants

Participants were recruited from 09/2017 to 03/2020 via referrals from our OCD specialty treatment programs, social media and web advertisements, clinicaltrials.gov, and a large (>200,000 member) local research registry available to University of Pittsburgh investigators. Given the brief (~3week) span of procedures and consistent with contemporary fMRI and TMS standards, stable treatment regimens (behavioral and pharmacologic—excluding agents that reduce seizure threshold) were allowed, as we did not expect these to either induce the pathology of interest or interfere with brain modulation. The following eligibility criteria were applied:

Inclusionary Criteria:
Participants were:

1) Between the ages of 18 and 55 years

2) Scored >1SD above the mean of healthy controls on at least one self-report scale of CBs, according to published normative data in healthy controls. These scales included:
    a. four relevant CB subscales (washing, checking, ordering, mental neutralizing) taken from the Obsessive Compulsive Inventory—revised (OCI-R(1)), a well-validated self-report inventory with excellent subscale factor structure, subscale stability, and discriminant validity. The hoarding OCI-R subscale was not used to determine eligibility, as it failed to discriminate OCD patients from controls(1), and hoarding may have a distinct neurobiology from other CBs(2).
    b. the Massachusetts General Hospital Hairpulling Scale(3,4), a well-validated and widely used self-report scale of the severity of compulsive hairpulling/trichotillomania symptoms
    c. the Skin Picking Scale(5), a valid and reliable self-report scale for the assessment of severity in medical and psychiatric patients who endorse compulsive skin picking
    d. the Threat-Related Reassurance Seeking Scale(6), a validated measure that correlates with symptoms of OCD, social anxiety, and GAD.
3) To ensure CBs were clinically significant, participants must be rated by a trained clinical rater as having at least moderate severity on at least 2 out of the 5 compulsion subscale items in the Yale-Brown Obsessive Compulsive Scale (Y-BOCS-II)(7).
4) Agreed to videotaping of structured clinical interview
5) Reported that they would reside in the Pittsburgh area for at least 5 weeks

Exclusionary Criteria:

1) Failure to meet standard MRI inclusion criteria: those who endorsed claustrophobia, those who had cardiac pacemakers, neural pacemakers, surgical clips in the brain or blood vessels, surgically implanted metal plates, screws or pins, cochlear implants, implanted uterine devices, metal braces, or other metal objects in their body, especially in the eye. History of significant injury or surgery to the brain or spinal cord that would impair interpretation of results. Pregnancy as determined by urine pregnancy tests on females.
2) Medical contraindications for Transcranial Magnetic Stimulation (TMS):
    a. Presence of a neurologic disorder or medication therapy known to alter seizure threshold (e.g., stroke, aneurysm, brain surgery, structural brain lesion, brain injury, frequent/severe headaches)
    b. Recurrent seizures or epilepsy in participant or family history of hereditary epilepsy
    c. Pregnancy
    d. Metallic implants in body or other devices that may be affected by magnetic field
    e. Significant heart disease or cerebrovascular disease
    f. Medications with seizure threshold lowering potential, e.g., clomipramine, Monoamine Oxidase inhibitors (MAOi's), imipramine, clozapine
3) Acute suicidality or other psychiatric crises requiring treatment escalation
4) Changes made to treatment regimen within 4 weeks of baseline assessment
5) Reading level <6th grade, per self-report
6) Presence of bipolar, psychotic, autism spectrum, or substance use disorder
7) Presence of movement disorder or tics affecting manual responses
8) Inability to read text from 2 feet away (corrective lenses allowed)

Diagnostic interviews were conducted at the screening visit by trained raters utilizing the Mini International Neuropsychiatric Interview (MINI) and selected modules from the SCID-5 to assess compulsive behavior spectrum disorders and specific phobias not assessed by the MINI. Principal and secondary diagnoses of the sample, which were reviewed and determined by a licensed clinical psychologist (RBP), reflected a range of compulsive behavior disorders, with patients exhibiting moderate obsessive-compulsive disorder (Y-BOCS) symptoms on average, as illustrated in Table 1 (main text).

TMS ramp-up and tolerability

Each TBS session began with a ramp-up block of 600 pulses (delivered in an iTBS or cTBS pattern, depending on the participant's allocation), during which the stimulator amplitude was gradually increased by the experimenter from 0% Maximum Stimulator Output (MSO) to either the target amplitude (active TBS day: 110% Resting Motor Threshold; sham TBS day: 20% MSO), or the maximum amplitude tolerated by the participant. As in clinical TMS procedures, the participant guided the pace and stopping point of the ramp-up at all times via continuous two-way communication with the experimenter (who stood directly beside the participant and monitored for both verbal and visual cues of discomfort), and the participant could request to stop increasing, to decrease the amplitude, or to discontinue stimulation altogether, at any time. No participant requested fully discontinuing the ramp-up or active stimulation blocks prior to the completion of the full dose of 600 pulses. Overall, of 69 participants who attempted the active stimulation day, n=59 (86%) tolerated the target amplitude of 110% RMT, reaching and retaining this amplitude within the ramp-up block, and completed the active block at the full target amplitude; n=4 (5.8%) tolerated ≥85% RMT within the ramp-up block and completed the active block at ≥85% RMT (86-104% RMT); and n=6 (8.7%) tolerated the stimulation only at <85% RMT and thus were considered not to have received an efficacious

amplitude of TBS. A subjective pain rating, representing the maximum pain reached during stimulation, was taken by the experimenter immediately following the completion of the active TBS block using a 1-10 scale, where 10 represented the point at which pain would have become intolerable. Pain ratings did not differ as a function of treatment allocation (cTBS: mean=7.12, SD=1.3; iTBS: mean=7.6, SD=2.0; t_{65} = -1.16, p=.25).

TMS neuronavigation and dosing

Navigator software was used to allow for stereotaxic registration of the participant's brain with the TMS coil, improving anatomical accuracy and minimizing variability across subjects(8,9). For baseline/sham sessions (prior to MPRAGE acquisition), a standardized MNI template was used for neuronavigation; for the active TBS session, the participant's own MPRAGE was used to improve anatomical accuracy. The location of the left OFC/FPC target was determined in 3D based on established anatomical landmark-based protocols(10). The experimenter began by utilizing the EEG International 10-20 system, which accounts for variability in participant skull size and is consistently used in clinical TMS applications when neuronavigation is not feasible, to identify location Fp1. The coil was initially placed in this location, and oriented tangential to the scalp, to observe the resulting focal point of stimulation as projected onto the structural brain image within the neuronavigation system. A digital marker was set at the estimated location of the focal point of stimulation on the structural image, and these target coordinates were then checked to verify they fell within the target OFC/FPC area (left Brodmann area 10). If the target region was not well-approximated by Fp1, the coil was then moved progressively downwards and/or laterally until the focal point in the neuronavigation system displayed a focal area within Brodmann area 10. Consistent with the *a priori* OFC/FPC target, the idealized goal of the neuronavigation procedure was to target the stimulation at Brodmann area 10 within as close proximity as possible to Brodmann's area 11 (as shown in **Figure 1A**, main text); however, the practical ability to place the neuronavigational focal point

within more ventral areas was constrained by the physical/spatial properties of the TMS coil and the participant's head shape (e.g., distance between forehead/brow and bridge of nose). Thus, this process was necessarily idiographic and therefore resulted in slightly differing scalp positions and focal point MNI coordinates for each participant, but the focal point was uniformly just above the left eye, near or overlaying the brow, and variability in focal point coordinates exhibited a fairly limited impact on findings in *post hoc* sensitivity analyses (please see details in "Sensitivity and exploratory analyses of clinical variables and procedural variables: Neuronavigational target coordinates" below). **Figure S1** below reflects the focal position and resulting scalp distribution of electrical field stimulation onto an example head.

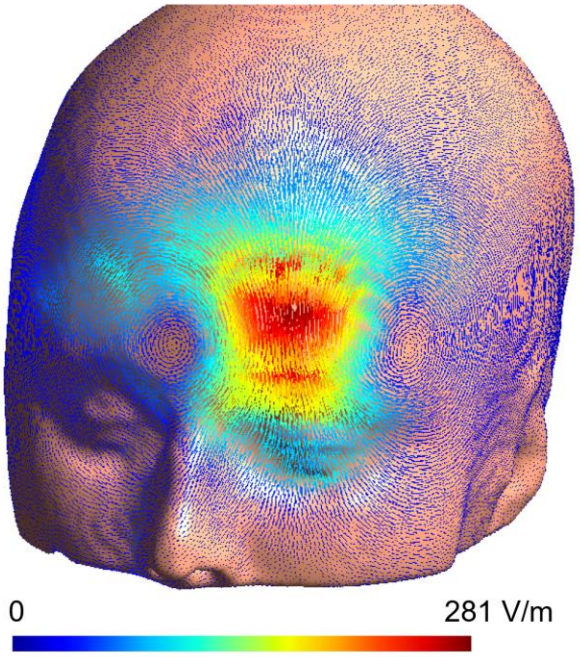


Figure S1. Computer simulation of the scalp distribution of the electric field induced by TBS over the left orbitofrontal/frontopolar cortex. The TMS coil placement approximated the targeting paradigm in the study. The TMS pulse amplitude was 100% of the maximum stimulator output. The simulation was carried out with SimNIBS 2 [https://simnibs.github.io/simnibs/build/html/index.html; (11)] using the default head model, tissue conductivity values, and MCF-B65 coil model.

Dosing was informed by meta-analyses of iTBS and cTBS protocols delivered over the motor cortex(12), suggesting reliable increase (iTBS) and decrease (cTBS) in motor evoked

potentials for 50-60min, with large effect sizes peaking 10-15min post-TBS. Preliminary findings suggest effects of similar TBS protocols extend to dorsolateral and medial PFC/OFC/FPC regions(13-17). We applied these standard doses (600 pulses in the full amplitude block) in each of the standardized iTBS and cTBS patterns.

TMS electric field distribution

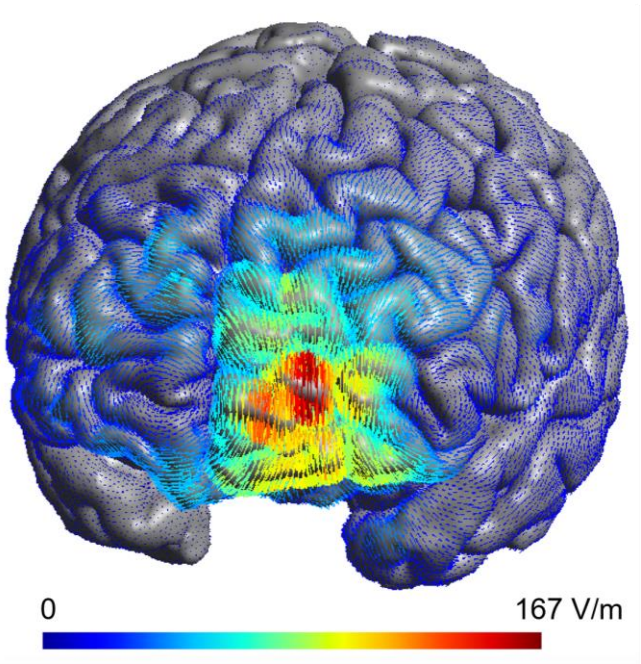


Figure S2. Computer simulation of the electric field distribution induced by TBS in the left orbitofrontal cortex/frontopolar cortex (OFC/FPC). The TMS coil placement approximated the targeting paradigm in the study. The TMS pulse amplitude was 100% of the maximum stimulator output. (Since the electric field strength scales linearly with the pulse amplitude, the shown distributions can be scaled proportionally to other pulse amplitudes.) The simulation was carried out with SimNIBS 2 [https://simnibs.github.io/simnibs/build/html/index.html; (11)] using the default head model, tissue conductivity values, and MCF-B65 coil model.

Habit override task

In one previously described paradigm used in prior OCD research (18-21), patients and controls were overtrained to acquire a habit (pressing levers with each foot) in response to cues

(colored rectangles) in order to avoid an aversive outcome (electrical shocks via arm bands on each wrist, which delivered safe (~1mA) electrical shocks that were idiographically titrated at the start of each visit such that they were rated as moderately unpleasant, but not painful). Following habit acquisition, one of the two shock bands was then 'devalued' by being explicitly disconnected from the subject's wrist, in full view. OCD patients continued to perform the overlearned avoidance response (pressing the foot lever) to a larger degree than controls, in spite of equivalent explicit knowledge that shocks would no longer be received.

We used this paradigm as the starting point to develop a novel, brief, fully automated form of training in 'habit override.' During the acute TBS modulation window, a habit override task was administered, modeled after previous OCD research(19) (**Fig. S3A**). At each of the two visits (baseline/sham and active TBS), a habit acquisition block was delivered just prior to TBS, followed by TBS (sham or active), followed by the habit override block. This allowed OFC/FPC modulation to occur strictly in the context of practice in <u>overriding</u> an existing habit via goal-directed behavior (a highly clinically relevant skill, akin to the goals of exposure and response prevention). To reduce practice effects over repeated administrations, alternate forms (2 non-overlapping stimulus sets) were developed, and the valued/devalued electrode sides (R vs. L) was counterbalanced across visits.

In <u>habit acquisition</u> (delivered prior to TBS), participants were instructed that their goal was to avoid receiving shocks to the left and right foot by pressing appropriate buttons whenever a conditioned cue appears (three cues for each side). To promote "stamping in" of simple stimulus-response habits, cognitive load was kept low by training habits in blocks—2 distinct cue->response pairs in each block (e.g., pairs of colors, fruits, etc.). Across 480 trials (3sec/trial; 24min total), participants overlearned these avoidance behaviors. In the subsequent <u>habit override</u> task (240 trials, 3sec/trial; 12min total), which was administered immediately (15-30min) following TBS (i.e., during the window when neuromodulatory effects were expected present), the same pairs of cues were presented in blocks. Within each block, 1 of the 2

overlearned habits was ‘devalued’. A written and verbal instruction was provided informing participants which one of the 2 electrodes had been physically disconnected (in full view) and stating they should attempt to resist the relevant ‘devalued’ avoidance response (i.e., override that habit), while continuing the remaining ‘valued’ habit.

Participants’ explicit knowledge of cue contingencies was rated at the close of each block using a 1-5 Likert-like scale (1=cue definitely did not predict a shock; 5=cue definitely did predict a shock). At the close of the entire task, participants rated the strength of their urges to respond to devalued cues on a 1-5 Likert-like scale.

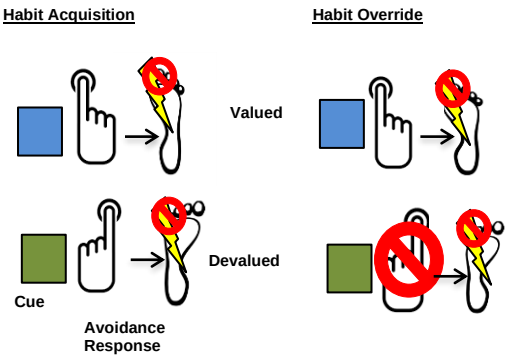
We expected participants would achieve the correct explicit knowledge of cue contingencies but might nevertheless experience some uncertainty as to whether shocks might be received following devalued cues. Thus, the task was intended to provide practice in learning to override a habit while simultaneously tolerating uncertainty, which further increases the clinical relevance of the goal-directed skill being acquired, as the ability to tolerate uncertainty is a key goal in gold-standard behavioral treatment(22) (exposure and response prevention). Likewise, based on prior findings in OCD(18,19), we expected participants might persist in the ‘devalued’ habit in spite of explicit knowledge that such responses are unnecessary.

Each of these expectations was confirmed. The task was free of practice effects, with no significant changes in explicit knowledge or response rates to valued, devalued, and safe cues from the baseline to post-TBS visits (*p’*s>.13). Explicit knowledge of cue contingencies was successfully retained from habit acquisition (delivered prior to TBS) to the subsequent habit override block (delivered immediately following TBS), as evidenced by significant differences in explicit ratings across the safe, devalued, and valued cues ($F_{2,130}$=221.9,p<.001), though a moderate degree of uncertainty regarding shock delivery was nevertheless evident for the devalued cue, as intended and consistent with prior research in both healthy controls and patients (**Fig. S3B)**. Participants also made intermittent behavioral responses to devalued cues, while responding accurately to valued cues and withholding responses to safe cues

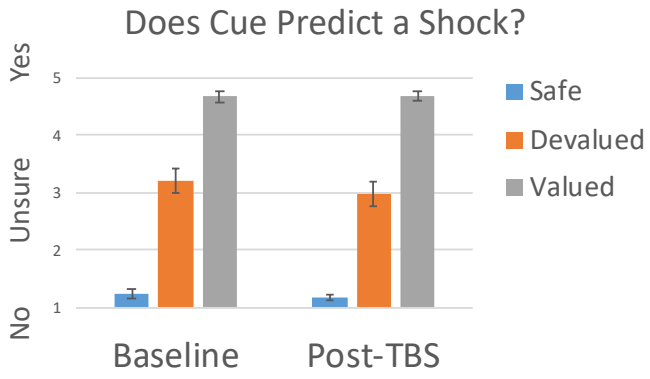
($F_{2,130}$=196.2,p<.001,**Fig. S3C**), and reported moderate urges to respond to devalued cues (mean=2.9 on a 1-5 scale), suggesting that the task provided salient practice in overriding a potent lab-acquired habit, which patients felt ‘compelled’ to complete.

Figure S3. A) Schematic of habit task components (habit acquisition, which was delivered just prior to sham and active TBS sessions, and habit override, which was delivered immediately following sham and active TBS sessions). A third set of cues (not shown) were "safe" cues that never predicted a shock to either side.
B) Explicit and C) implicit (behavioral) responses during the habit override task among CB patients. Devalued cues, which were relevant only to a now-disconnected electrode, were rated as moderately uncertain to predict a shock, and were moderately likely to result in habitual/over-learned behavioral button presses (though such presses were no longer needed to avoid a shock). No practice effects were evident across the two repeated sessions. According to pairwise post hoc comparisons, devalued cues differed from both safe and valued cues ($p<.001$), at both sessions, for both explicit and implicit responses.

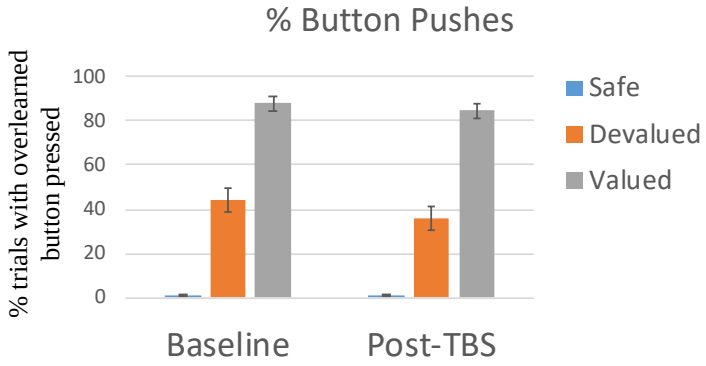
A) Schematic of Task



B)



C)



fMRI acquisition and processing

Pseudocontinous Arterial Spin Labeling (pcASL) sequence: The total duration (4min) and number of acquisitions (25 labeled/unlabeled pairs) in the pcASL sequence were consistent with established conventions in the neuroimaging field(23-25). Although 25 pairs represents a slightly decreased number of acquisitions relative to many widely used sequences (e.g., 30-60 pairs), and could potentially have reduced SNR and/or power in the present study, previously published work suggests that reproducible findings can be obtained with as few as 6 labeled/controlled pairs of images in <1min of pcASL acquisition time (23). Standard preprocessing steps were conducted in SPM12, supplemented by the ASL toolbox (Asltbx): (1) motion correction, applied separately to the labeled/tagged images and control/untagged images; (2) coregistration of pcASL time series data to the ASL M0 map; (3) spatial smoothing (8-mm full width half maximum); (4) coregistration of the MPRAGE structural image to the mean ASL map (derived from the motion correction step); (5) generation of an eroded white matter mask using SPM segmentation outputs; (6) calculation of Cerebral Blood Flow (CBF) and perfusion maps from the time series of labeled and control images via the ASL toolbox; (7) coregistration of the perfusion/CBF maps to the structural MPRAGE image; and (8) normalization to the ICBM template in SPM. The resulting single-subject CBF maps at each of the two time points were used in AFNI to calculate difference maps (active-TBS – sham/baseline) for each participant. Difference maps were compared across the iTBS and cTBS groups in AFNI via unpaired 3dttest++ commands.

To reduce multiple comparisons for tests of prefrontal target engagement, a broad prefrontal/anterior mask was defined by all voxels within the MNI template with y-coordinate≥22.

A voxel-wise search was then performed within this mask to identify prefrontal clusters differentially modulated by TBS condition (voxel-wise p≤.005, map-wise p<.05 as determined by AFNI's 3dClustSim command). 3dClustSim was applied using AFNI's spatial autocorrelation function ("mixed ACF" approach), which has been shown to provide accurate type I error control for voxel-wise *p*<.005 under other relevant conditions of the present study (26). Following recommended procedures and considerations (27), a voxel-wise threshold of *p*<.005 was selected to optimize the balance between Type I and Type II error rates under the sample size constraints of the current study.

Multiband-multiecho fMRI resting state BOLD sequence: The total duration (7min) and number of acquisitions (480 TRs) in the BOLD sequence were consistent with established conventions in the neuroimaging field(28). A custom Matlab script was first applied to reconstruct raw timeseries data from the three echo times through a weighted summation of multi-echo signals, thereby combining the three images to optimize for different T^*_2 across the brain, as described previously (29). This approach has been found to substantially reduce susceptibility artifact and increases BOLD signal-to-noise by up to 80% in OFC and other ventral regions(35-37). Standard preprocessing steps were then applied using Analysis of Functional Neuroimaging (AFNI) via the afni_proc.py program, including: despiking, slice time correction, spatial distortion correction (using pairs of AP/PA spin echo field maps collected just before the BOLD acquisition), motion correction using 6 parameters and their 1^{st} derivatives, linear and quadratic detrending to correct drift, cross-registration of functional data to a high-resolution structural scan acquired in the same fMRI session (axial MPRAGE: TR=2400; TE=2.22; 208 slices; flip angle=8°; 0.8mm isotropic voxels), normalization/warping to the Montreal Neurological Institute Colin-27 brain data set, spatial smoothing to 6-mm full width half maximum, scaling each voxel to a mean of 100, and removal of physiological/hardware artifacts via AFNI's fast-anaticor algorithm.

AFNI's 3dRSFC tool was then used to apply a standard bandpass filter (0.01 to 0.1 Hz) and to quantify fractional amplitude of low-frequency fluctuations (fALFF), a voxel-wise index of the absolute level of resting activation following TBS. For each participant, fALFF was calculated at every voxel. Whole-brain maps consisting of Winsorized mean fALFF within anatomical *a priori* target regions of left Brodmann's Area 47 and 11 (defined within AFNI's TT_Daemon atlas) were extracted for each individual (using AFNI's 3dmaskave function) and compared in external statistical software (SPSS) across the two TBS conditions with unpaired t-tests.

A validity check confirmed that fALFF values in both anatomical ROIs were unrelated to any motion parameter (*p*'s>.56), suggesting the fALFF resting state metric was unconfounded by motion artifacts.

Analytic Strategy

Overarching per-protocol strategy. On the basis of the preliminary nature, scope, and goals of this first-of-its-kind study in humans, and in an effort to appropriately balance Type I and Type II error risk within the context of the current design and sample size, we elected to conduct per-protocol analyses. Per-protocol analyses present important benefits in the present context and were deemed the most appropriate option given the stage of the present work. Namely, per-protocol analyses provide the best method to evaluate the impact of the study manipulations under optimized conditions, i.e., when participants receive the study procedures that were intended. Thus, while intent-to-treat (ITT) analyses test the effect of "assigning a treatment," per-protocol analyses test the effect of "using (i.e., tolerating) a treatment" (30); the latter question is more germane to the goals of the present experimental study. Although ITT analyses reduce the risk of selection bias, which is essential when the goal is to draw definitive/confirmatory conclusions regarding intervention efficacy, they also can dramatically

reduce power particularly when 1) the randomized sample is relatively small and 2) the participants who deviate from the protocol do so in highly impactful ways (e.g., when a single-session TBS manipulation is not delivered at a dose capable of truly modulating brain activity).

The CONSORT diagram in the main text (Figure 2) illustrates that allocated/randomized participants who were not included in our analyses (n=16) were almost entirely comprised of (1) those who did not return for the active stimulation day, nor for subsequent assessments, and thus received no intervention and contributed no data after baseline (n=9) and (2) those who did not tolerate TBS at an adequate amplitude to have a cortical effect (n=6). Thus, as noted above, inclusion of such participants in analysis was expected to adversely impact power to detect the true effect of the single-session intervention (when delivered as intended) in this first-of-its-kind report. However, we verified (main text Table 1; Figure 2 legend) that the participants left out of per-protocol analyses do not differ systematically at baseline from those included in analyses [on neural and behavioral (CB probe) outcome measures, nor on any demographic or clinical features reported in Table 1, reducing the risk of selection bias in our sample.

Across analyses, to be comprehensive, we report both uncorrected *p*-values and *p*-values corrected for multiple comparisons, after applying either 1) map-wise cluster thresholding (pcASL analysis) or 2) False Discovery Rate (FDR) correction for the total number of tests within a given domain/aim (three behavioral outcomes—detailed further below).

fMRI indices of target engagement. For fMRI indices, our primary index of target engagement—pcASL—was acquired both at baseline and post-active TBS. A voxel-wise analysis was used to identify any prefrontal regions where the degree of change in CBF across the two visits was moderated by TBS condition. Voxel-wise difference scores were first computed as difference maps (post-active TBS - baseline/sham-TBS) for each individual and compared in AFNI with voxel-wise unpaired t-tests across the two groups (see further details

above in “fMRI acquisition and processing”). Mean CBF from the identified functional ROI exhibiting differential change as a function of group (averaged across all voxels significant at $p<.005$) was then extracted for each individual at each timepoint (baseline and post-TBS) and compared across the two timepoints in *post hoc* comparisons performed with SPSS software, using paired t-tests applied to each group separately.

We observed *post hoc* that the pcASL index suffered routinely from more severe signal dropout within the OFC target region relative to our custom multiband-multiecho BOLD resting state sequence. A secondary index of target engagement—fALFF (which was measured only at post-TBS)—was therefore used to probe for additional target engagement within the more orbital anatomical areas of the OFC target (BAs and 11). To reduce multiple comparisons and focus hypothesis tests strictly on anatomical ventral OFC regions selected *a priori*, fALFF was quantified in the anatomically defined left BA 47 and left BA 11 target regions and compared across the two groups with unpaired t-tests performed in SPSS.

Laboratory assessments of idiographic compulsive behaviors. For analyses of laboratory CBs, the primary focus was to identify group*time interactions exhibited across the two primary timepoints: baseline/sham-TBS and acute post-TBS sessions. Longitudinal analyses were conducted with Hierarchical Linear Modeling (HLM) software (31). Linear mixed models, with subject as a random effect and group (iTBS vs. cTBS) as a fixed factor, were used to test for an effect of TBS condition on the slope of each of three outcomes (urges, effort, and time spent engaged in CBs; one separate HLM model per outcome) over the two timepoints. Time was coded such that the intercept reflected values of the outcome variable at the baseline assessment. There was no missing data in these analyses. Follow-up paired t-tests were used to test for significant changes from baseline within each group separately (iTBS and cTBS).

To assess the relationship between change in PFC activation levels and change in behaviors, pre-to-post-TBS difference scores were computed for CBF (in the functional ROI identified in voxel-wise search, described above) and for each of the three behavioral indices

from the laboratory CB probe. Difference scores were correlated across CBF and each of the three laboratory CB indices using Pearson's *r*.

To probe the durability of TBS effects on laboratory CBs at the 1-week follow-up, baseline and 1-week values on each laboratory CB index were compared across the two timepoints using paired t-tests applied to each group separately. Given modest sample sizes, this within-groups approach for exploratory analyses maximized power to detect maintenance of effects within either group in isolation. In these exploratory analyses, we specifically hypothesized that any iatrogenic deficits induced by one of the two TBS arms would be short-lived and would not persist at follow-up, given that a single session of TBS would be expected, at worst, to interfere with the acquisition of a novel skill (habit override), but have only a transient impact on brain function; on the other hand, we hypothesized that beneficial effects might be maintained in one of the conditions if a clinically relevant, generalizable skill (i.e., the ability to override habits) was acquired through our synergistic biobehavioral approach.

Supplemental Analyses

ANCOVA sensitivity analyses of the laboratory probe of compulsive behaviors at +90min

When comparing iTBS vs. cTBS at +90min using an ANCOVA to covary baseline values, both of the significant group differences reported in the main text were upheld (urge strength: $F_{1,59}$=5.9, p=.018; time spent in CBs: $F_{1,59}$=5.71, p=.020).

Sensitivity and exploratory analyses of clinical variables (YBOCS, diagnostic subgroups) and procedural variables

We did not anticipate an effect of our experimental procedures on clinical outcome measures, which are designed to capture relatively longer time frames of symptomatology that pertain to often deeply entrenched patterns of behavior. In the context of our single-session procedures and short overall study duration, we did not think this was a reasonable expectation. We expected that additional design features which are standard for clinical TMS interventions, e.g. multi-session repeated administrations, would be necessary in order to observe such effects. Nevertheless, we probed for any such impact of our procedures on clinical measures in a series of sensitivity and exploratory analyses, as detailed below.

Main effects on clinical outcomes. In a repeated-measures ANOVA comparing baseline and 1-week values on the YBOCS (compulsions subscale, obsessions subscale), there was no significant differential effect of our single-session iTBS vs. cTBS procedures on clinical symptom severity (compulsions subscale: $F_{1,59}$=0.59, p=.45; obsessions subscale: $F_{1,59}$=0.007, p=.932). We anticipated that our brief single-session procedure would likely be ineffective on clinical outcome measures designed to measure more enduring, overall symptom severity in real-world settings.

Moderating effects of clinical variables. In sensitivity analyses, we probed for potential moderating effects of 1) initial YBOCS-compulsions subscale severity, 2) initial YBOCS-obsessions subscale severity, and 3) principal diagnosis (OCD vs. any other compulsive spectrum disorder), to assess factors that might have interacted with group assignment (iTBS vs. cTBS) in predicting acute neuroimaging (pcASL) and/or behavioral (urge strength, time spent in CBs, effort needed to resist) outcomes at +90min. Notably, these analyses revealed a significant 3-way interaction between time (baseline vs. TBS+90min), initial symptom severity on the YBOCS compulsions subscale, and group (iTBS vs. cTBS) at the neural level of analysis on mPFC CBF ($F_{1,54}$=8.55, p=.005), as well as on behavioral CB probe outcomes for time spent

in CBs ($F_{1,58}$=4.28, p=.043), and a similar non-significant trend for urge strength ($F_{1,58}$=2.87, p=.095). The nature of the interaction in all three cases was that patients with higher initial symptom severity showed a larger decrease in the outcome following cTBS, whereas no relationship existed between baseline severity and behavioral outcomes in the iTBS group. One possible interpretation of this pattern is that the cTBS procedures, which exhibited a beneficial effect at the group level in the current sample, might be most effective in higher-severity CB patients.

Similar 3-way interactions were found when assessing baseline severity of obsessions (per YBOCS-obsessions subscale) as a moderator of TBS condition's impact on both mPFC CBF ($F_{1,54}$=7.53, p=.008) and a non-significant trend was also present for urge strength ($F_{1,58}$=3.01, p=.088). These findings suggest the observed, exploratory pattern described above may have been driven by overall obsessive-compulsive symptom severity, rather than compulsion severity *per se*—particularly with regard to potential moderating effects on neural activity changes.

No significant interaction effects were observed for diagnostic subgroup on either neural or behavioral outcomes (p's>.51), though power for detecting such effects is likely constrained by the subgroup sample sizes within each treatment arm.

Neuronavigational target coordinates. The neuronavigation system's estimate of the focal point of stimulation was recorded at the active TBS session for each participant as a set of MNI coordinates. Variability in the x-, y- and z MNI coordinates were correlated with neural and behavioral outcome measures by calculating correlations separately within each group (cTBS, iTBS), as the directionality of any impact of target coordinates on outcomes would be expected to be opposite in the two groups. These analyses did not reveal a wide-spread systematic impact of focal point coordinates and outcomes. There were no relationships observed between the MNI coordinate, in any plane, to the pcASL values observed in the mPFC (i.e. the functional

ROI in Figure 3B, main text) post-TBS (either absolute values post-TBS, or change from pre-to-post-TBS), in either group. With regard to the fALFF index post-TBS and the three behavioral CB outcomes, there were also no significant relationships between MNI coordinates in the **x-** and **z-** planes and any outcome in either group. However, for the **y-**coordinate, two significant patterns were observed, both specific to the cTBS group: smaller Y coordinate values (i.e., relatively posterior focal coordinates, which, for our frontal target, equates to greater cortical depth of the focal point of stimulation) were associated with (1) lower left OFC fALFF values, in the left BA11 (r=.46, p=.009) and (marginally) in the left BA47 (r=.34, p=.055) and (2) greater degree of improvement (from baseline to post-TBS) in terms of time spent in CBs during the laboratory probe (r=.39, p=.02). No significant relationships were observed in the iTBS group for any coordinate or variable.

These exploratory findings could tentatively suggest that, for cTBS, greater cortical depth of the stimulation target (i.e., smaller Y-coordinate value) promoted greater generalization of stimulation into ventral OFC areas, which in turn promoted more robust behavioral improvements. However, these findings must be considered in light of their exploratory nature, small within-group sample sizes, lack of multiple comparisons correction, and lack of robust generalization across all neural and behavioral outcome measures.

Timing of study procedures. To probe whether dynamic changes in neural effects of TBS may have influenced findings in the current study, we examined whether neural outcomes (mPFC CBF and OFC fALFF post-TBS) were correlated with the time elapsed between the end of the active TBS session and the start of fMRI data collection. Correlations were performed separately within each treatment group, as time elapsed would be expected to have opposing directionality of effects based on whether iTBS or cTBS was delivered. Time elapsed between the end of TBS and the onset of the scan was not related to any neural outcome measure in either group (p's>.15)

Relationships between pcASL and fALFF target engagement indices

Across participants, there was a correlation between pre-to-post-TBS change in CBF (the primary neural target engagement index) and post-TBS fALFF within the functional vmPFC ROI identified in the primary pcASL analysis (r=.27, p=.046), as well as a correlation between CBF in this functional vmPFC ROI and fALFF in the more ventral OFC regions included in our secondary analyses in the main text (averaging fALFF values for BAs 47 & 11; r=.31, p=.02).

Laboratory probe of compulsive behaviors

Under the supervision of a licensed psychologist with extensive experience in behavioral treatments for OCD (RBP), we idiographically designed, collaboratively with the participant, a laboratory task involving a triggering object, image, or scenario that corresponded to CBs typical to the participant. In support of the laboratory task's validity and clinical relevance, the laboratory triggered indices of CB severity were correlated with clinician-rated compulsive behavior symptom severity, as indexed by the gold standard Yale-Brown Obsessive-Compulsive Scale (Y-BOCS; (7)) compulsions subscale (Table S1). With one exception, these relationships were evident both for the screening/baseline assessments, and again at the 1-week follow-up, when the Y-BOCS interview was repeated, and suggest that the laboratory task captured clinically relevant symptomatology, while simultaneously allowing for capture of immediate shifts in vulnerability to primary compulsions (e.g., at TBS +90min) that cannot be effectively captured with standard clinical instruments which are designed to measure enduring, overall symptom severity across a clinically relevant window of time (e.g., 1 week or more).

Table S1: Correlation coefficients

	Urge	Effort	Time spent in CBs
YBOCS-C: baseline	.41***	.28*	-.01
YBOCS-C: 1-week f.u.	.39**	.26*	.22^{+}

+p<.10, *p<.05, **p<.01, ***p<.001

All findings reported in the main text were obtained during a 5min CB probe task that was catered ideographically to the participant's primary compulsion, as determined during the screening visit via Y-BOCS and MINI diagnostic interviews. A second, exploratory 5min CB trigger/probe task was completed immediately following the first one, which was identical to the first CB probe task, except that it was designed to capture a secondary compulsive behavior target within a distinct symptom domain. Thus, the second CB probe was not necessarily directly related to the participant's chief complaint and had more variable clinical relevance depending on the participant's clinical presentation. To reduce multiple comparisons and maximize the clinical relevance of the laboratory probe data, this exploratory data was not included in any of the present reported analyses.

Supplemental References

1. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, Salkovskis PM: The Obsessive-Complusive Inventory: Development and validation of a short version. Psychol Assess 2002;14(4):485-95
2. Saxena S: Recent advances in compulsive hoarding. Curr Psychiatry Rep 2008;10(4):297-303
3. Keuthen N, O'Sullivan R, Ricciardi J, Shera D, Savage C, Borgmann A, Jenike M, Baaer L: The Massachusetts General Hospital (MGH) Hairpulling Scale: 1. Development and Factor Analyses. Psychother Psychosom 1995;64:141-5
4. Keuthen NJ, Flessner CA, Woods DW, Franklin ME, Stein DJ, Cashin SE, Trichotillomania Learning Center Scientific Advisory B: Factor analysis of the Massachusetts General Hospital Hairpulling Scale. J Psychosom Res 2007;62(6):707-9
5. Keuthen NJ, Wilhelm S, Deckersbach T, Engelhard IM, Forker AE, Baer L, Jenike MA: The Skin Picking Scale: scale construction and psychometric analyses. J Psychosom Res 2001;50(6):337-41
6. Cougle JR, Fitch KE, Fincham FD, Riccardi CJ, Keough ME, Timpano KR: Excessive reassurance seeking and anxiety pathology: Tests of incremental associations and directionality. J Anxiety Disord 2012;26(1):117-25
7. Storch EA, Rasmussen SA, Price LH, Larson MJ, Murphy TK, Goodman WK: Development and psychometric evaluation of the Yale-Brown Obsessive-Compulsive Scale--Second Edition. Psychol Assess 2010;22(2):223-32
8. Julkunen P, Säisänen L, Danner N, Niskanen E, Hukkanen T, Mervaala E, Könönen M: Comparison of navigated and non-navigated transcranial magnetic stimulation for motor cortex mapping, motor threshold and motor evoked potentials. Neuroimage 2009;44(3):790-5
9. Ruohonen J, Karhu J: Navigated transcranial magnetic stimulation. Clin Neurophysiol 2010;40(1):7-17
10. Crespo-Facorro B, Kim JJ, Andreasen NC, O'Leary DS, Wiser AK, Bailey JM, Harris G, Magnotta VA: Human frontal cortex: an MRI-based parcellation method. Neuroimage 1999;10(5):500-19
11. Windhoff M, Opitz A, Thielscher A: Electric field calculations in brain stimulation based on finite elements: an optimized processing pipeline for the generation and usage of accurate individual head models. Hum Brain Mapp 2013;34(4):923-35
12. Wischnewski M, Schutter DJLG: Efficacy and Time Course of Theta Burst Stimulation in Healthy Humans. Brain Stimul 2015;8(4):685-92
13. Grossheinrich N, Rau A, Pogarell O, Hennig-Fast K, Reinl M, Karch S, Dieler A, Leicht G, Mulert C, Sterr A, Padberg F: Theta Burst Stimulation of the Prefrontal Cortex: Safety and Impact on Cognition, Mood, and Resting Electroencephalogram. Biol Psychiatry 2009;65(9):778-84
14. Prause N, Siegle G, Deblieck C, Wu A, Iacoboni M: EEG to primary rewards: predictive utility and malleability by brain stimulation. Submitted
15. Chistyakov AV, Rubicsek O, Kaplan B, Zaaroor M, Klein E: Safety, tolerability and preliminary evidence for antidepressant efficacy of theta-burst transcranial magnetic stimulation in patients with major depression. Int J Neuropsychopharmacol 2010;13(3):387-93
16. Mastropasqua C, Bozzali M, Ponzo V, Giulietti G, Caltagirone C, Cercignani M, Koch G: Network based statistical analysis detects changes induced by continuous theta-burst stimulation on brain activity at rest. Front Psychiatry 2014;5:97
17. Hanlon CA, Dowdle LT, Austelle CW, DeVries W, Mithoefer O, Badran BW, George MS: What goes up, can come down: Novel brain stimulation paradigms may attenuate craving and craving-related neural circuitry in substance dependent individuals. Brain Res 2015;1628:199-209
18. Gillan CM, Morein-Zamir S, Urcelay GP, Sule A, Voon V, Apergis-Schoute AM, Fineberg NA, Sahakian BJ, Robbins TW: Enhanced Avoidance Habits in Obsessive-Compulsive Disorder. Biol Psychiatry 2014;75(8):631-8

19. Gillan CM, Apergis-Schoute AM, Morein-Zamir S, Urcelay GP, Sule A, Fineberg NA, Sahakian BJ, Robbins TW: Functional Neuroimaging of Avoidance Habits in Obsessive-Compulsive Disorder. Am J Psychiatry 2015;172(3):284-93
20. Gillan CM, Kalanthroff E, Evans M, Weingarden HM, Jacoby RJ, Gershkovich M, Snorrason I, Campeas R, Cervoni C, Crimarco NC, Sokol Y, Garnaat SL, McLaughlin NCR, Phelps EA, Pinto A, Boisseau CL, Wilhelm S, Daw ND, Simpson HB: Comparison of the Association Between Goal-Directed Planning and Self-reported Compulsivity vs Obsessive-Compulsive Disorder Diagnosis. JAMA Psychiatry 2019:1-10
21. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND: Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife 2016;5
22. Belloch A, Cabedo E, Carrio C, Fernandez-Alvarez H, Garcia F, Larsson C: Group versus individual cognitive treatment for Obsessive-Compulsive Disorder: Changes in non-OCD symptoms and cognitions at post-treatment and one-year follow-up. Psychiat Res 2011;187(1-2):174-9
23. Fernandez-Seara MA, Edlow BL, Hoang A, Wang J, Feinberg DA, Detre JA: Minimizing acquisition time of arterial spin labeling at 3T. Magn Reson Med 2008;59(6):1467-71
24. Kilroy E, Apostolova L, Liu C, Yan L, Ringman J, Wang DJJ: Reliability of two-dimensional and three-dimensional pseudo-continuous arterial spin labeling perfusion MRI in elderly populations: Comparison with 15o-water positron emission tomography. J Magn Reson Imaging 2014;39(4):931-9
25. Chen JJ, Jann K, Wang DJJ: Characterizing Resting-State Brain Function Using Arterial Spin Labeling. Brain Connect 2015;5(9):527-42
26. Cox RW, Reynolds RC, Taylor PA: AFNI and clustering: false positive rates redux. BioRXiv 2016;doi: http://dx.doi.org/10.1101/065862
27. Carter CS, Lesh TA, Barch DM: Thresholds, Power, and Sample Sizes in Clinical Neuroimaging. Biol Psychiatry Cogn Neurosci Neuroimaging 2016;1(2):99-100
28. Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL: Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J Neurophysiol 2010;103(1):297-321
29. Kim T, Zhao T, Bae KT: Enhancement of functional MRI signal at high-susceptibility regions of brain using simultaneous multiecho multithin-slice summation imaging technique. J Magn Reson Imaging 2016;44(2):478-85
30. Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Jager KJ: Intention to treat and per protocol analysis in clinical trials. Nephrology (Carlton) 2020;25(7):513-7
31. Raudenbush SW, Bryk AS: Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd ed. Thousand Oaks, CA, Sage Publications, Inc., 2002