Supplemental Results

Behavioral results detailed

Overall, participants made errors on 2.3% ± 0.4 of the Go trials and 3.2% ± 0.7 of the NoGo trials, but on 11.6% ± 2.1 of the Conflict trials (ANOVA, within subjects repeated measures, $F(2,33) = 21.35$, $p < 0.001$). The more difficult Conflict trials showed significantly higher error rates than the easier Go ($t(33) = 4.82$, $p < 0.05$, post hoc t-test) and the NoGo trials ($t(33) = 4.62$, $p < 0.05$, post hoc t-test). The Go and the NoGo trials showed no significant difference in error rates ($t(33) = 1.59$, $p > 0.05$, post hoc t-test). Correct Conflict trials also had significantly longer response times than correct Go trials across experimental sessions (963.8 ± 17.1 ms versus 831.5 ± 17.2 ms; $t(33) = 15.18$, $p < 0.001$, paired t-test). The difference in response times between Conflict and Go trials was robust at the individual session level. In 32 of the 34 experimental sessions, response times were significantly slower during the Conflict condition ($p < 0.05$, within session unpaired t-test).

Notably, participants adjusted and slowed their behavioral responses following correct trials that involved Conflict and following trials that in which the participant committed an error. Response times during correct Go trials that followed a correct Conflict trial were 139.0 ± 11.8 ms slower than response times during correct Go trials that followed a previous correct Go trial (940.0 ± 18.5 ms versus 800.9 ± 17.4 ms; $t(33) = 11.75$, $p < 0.001$, paired t-test). These differences were significant at the within session level in the majority of individual experimental sessions (26 of 34 experimental session, $p < 0.05$, within session unpaired t-test). Moreover, across sessions, participants committed more errors during Go trials following correct Conflict trials than during Go trials following previous correct Go trials (6.7% ± 1.0 versus 1.3% ± 0.4; $t(33) = 5.23$, $p < 0.001$, paired t-test). Similar to the across-trial adjustments that occurred following the more difficult Conflict trials, response times were also significantly longer during correct Go trials that followed errors than during correct Go trials that followed correct trials (871.8 ± 24.6 ms versus 820.6 ± 17.0 ms; $t(14)=2.66$, $p < 0.05$, paired t-test).

Within-contact analysis of mPFC theta and beta power demonstrates focality

Our primary analyses focused on the average changes in mPFC theta and beta power when all five bipolar contacts in each recording strip were averaged together prior to averaging across sessions (Supplementary Fig. S2). We were also interested in whether the observed changes represented a focal process or a diffuse cortical phenomenon. We therefore examined changes in spectral power within each electrode contact for each recording session. In 26 out of the 32 sessions with mPFC recordings, at least one electrode contact exhibited a significant increase in theta power during the task ($p < 0.05$, permutation test). This effect was not present in all electrode contacts, however. Indeed, within these 26 sessions, we found that only 3.04 ± 0.26 of the five mPFC electrode contacts ($60.7 \pm 5.1\%$) showed a significant increase in theta power relative to baseline ($p < 0.05$, permutation test; Supplementary Fig. S3). The differences between trial types was also focal. Out of the 32 sessions, 19 sessions demonstrated significantly higher theta power during the Conflict or NoGo trials relative to the Go trials. Out of these 19 sessions, however, only 2.30 ± 0.25 of the five contacts ($46.0 \pm 4.9\%$) showed a significant difference at the within-contact level. Similarly, in the beta band, 25 of the 32 sessions showed a significant post-response increase in beta power, but only 3.44 ± 0.26 of the five contacts ($68.9 \pm 5.1\%$) showed a significant increase at the within-contact level. Finally, 14 of the 32 sessions showed a significant post-response difference between Go and AntiGo trials, but only 2.4 ± 0.3 of the five contacts ($47.5 \pm 6.7\%$) showed a significant difference at the within-contact level.

Trial type related differences in individual STN spiking recordings

On average, the task related MUA changes showed higher activity for the Conflict trials, followed by the Go trials, and then the NoGo trials (Fig. 2). Supplementary Fig. S4A shows five microelectrode recordings that showed this firing rate pattern at the individual level. However, not all MUA recordings demonstrated this pattern of trial type related differences. Supplementary Fig. S4B shows two recordings that showed higher firing for the NoGo trials than the Go trials, and Supplementary Fig. S4C shows two recordings that actually showed a decrease in firing during the task that was most pronounced during the NoGo trials. Notably, we also observed some MUA recordings that demonstrated changes in activity that depended on whether the subjects answered correctly, and some additional recordings that demonstrated changes that depended on the level of conflict on the previous trial. When we re-extracted the spiking using

traditional spike threshold detecting methods (Supplementary Fig. S5), we observed similar results as those seen when we calculated each microelectrode's MUA. The threshold detection method, however, produced noisier measurements of task related activity and required us to arbitrarily chose a different threshold for each recording.

Trial type related differences in spectral power across all frequencies

Supplementary Fig. S6 shows the mean changes in STN spectral power across all sessions plotted separately for the Go, Conflict, and NoGo trials. The Go trials showed very similar patterns as the 'All trials' plot shown in Fig. 1d, consistent with the fact that the majority of trials were Go trials. The STN showed higher theta power for the Conflict trials relative to the Go trials, and these differences occurred late in the task, around the time of the response. Notably, the NoGo trials showed significantly lower theta power relative to the Go trials. The NoGo trials also showed an earlier return to baseline of the task related beta power decrease.

Supplementary Fig. S7 shows the mean changes in mPFC spectral power across all sessions plotted separately for the Go, Conflict, and NoGo trials. The mPFC showed a pre-response increase in 2-5 Hz theta power and a post-response increase in 8-30 Hz beta power. Relative to the Go trials, the NoGo trials showed a significantly higher pre-response increase in theta power ($p < 0.05$, permutation test). The Conflict trials also showed significantly higher pre-response theta power, but they further showed a significantly attenuated post-response beta power increase ($p < 0.05$, permutation test).

Correlations between single trial power and response time

We analyzed the correlation between response time and mPFC theta power, STN theta power, and STN beta power in each trial. This revealed a significant, albeit weak correlation with response time for all three variables (Supplementary Fig. S9; see Supplemental Materials and Methods). In all correct trials that involved a response (Go and Conflict trials), the mean correlation coefficients (R) across sessions were equal to 0.04 ± 0.01 ($t(31) = 3.61$, $p < 0.01$, one-sample t-test) for mPFC theta power, 0.04 ± 0.02 ($t(28) = 2.07$, $p < 0.05$, one-sample t-test)

for STN theta power, and 0.06 ± 0.02 ($t(28) = 3.17$, $p < 0.01$, one-sample t-test) for STN beta power. We found no significant difference in correlation coefficients between correct Go and correct Conflict trials ($p > 0.05$, paired t-test). As a control, we also analyzed the correlation between response time and the power levels during the baseline period. This analysis revealed no significant correlations between response time and baseline levels of mPFC theta power, STN theta power, or STN beta power (Supplementary Fig. S9, $p > 0.05$, one-sample t-tests). Together these data suggest that oscillatory power may be related to individual trial response time. However, we did not find, a significant relation between the increases in mPFC-STN theta coherence and response times across sessions (mean correlation coefficient (R)= -0.01 ± 0.01, $t(27) = -2.03$, $p > 0.05$, one-sample t-test).

Trial type related differences in phase-MUA coupling

Recent work has suggested that one of the ways in which STN beta oscillations may affect basal ganglia output is through the beta phase entrainment of spiking activity. We quantified this effect by analyzing the consistency of the relationship between macro electrode LFP phase and micro electrode MUA amplitude (phase-MUA coupling (PMUAC), see Supplementary Materials and Methods). The left panel of Supplementary Fig. S10A, shows the mean PMUAC value (normalized to a surrogate distribution, see Supplementary Materials and methods) across all STN MUA recordings. This analysis revealed which frequencies (between 2-107 Hz) significantly modulated MUA amplitude throughout the different phases of the oscillatory cycle. The MUA amplitude was significantly modulated by the ongoing phase in the beta and theta frequency ranges. The middle panel of Supplementary Fig. S10A further shows that the beta band PMUAC significantly decreased when subjects were engaged with the task. When we analyzed the beta PMUAC for the Go, Conflict, and NoGo trials separately, we observed significantly lower PMUAC during the Go trials relative to the NoGo trials. These differences were only significant 1000 ms after each trial began, which was later than the mean response time for the Go trials. Thus, we believe these differences are due to movement execution related changes in PMUAC, but most likely do not contribute to whether or not an action occurs. When we analyzed the theta PMUAC for the Go, Conflict, and NoGo trials separately, we observed significantly higher PMUAC during the Conflict trials relative to the Go trials, consistent with

prior work (Zavala *et al.*, 2017). We repeated these analyses using the voltage thresholded spiking data instead of the MUA data (Supplementary Fig. S10B). This confirmed the presence of STN beta phase-spike phase locking at baseline as well as its decrease during the task, however it failed to reproduce the trial type dependent differences. Analyzing the relationship between mPFC phase and STN MUA amplitude did not show any significant PMUAC (data not shown).

Error related differences in STN MUA and mPFC gamma power

We analyzed the average error related changes in MUA amplitude during our task (Supplementary Fig. S11). This analysis revealed that on average STN MUA after an error response was significantly higher than during correct Go and correct Conflict trials ($p < 0.05$, permutation test).

Given that the error trials showed a significant increase in mPFC gamma power relative to baseline that was not present when the correct trials were compared to baseline (Fig. 5C vs Fig. 1E), we were interested in testing whether there were any trial type dependent differences in mPFC gamma power. Supplementary Fig. S12 shows significantly higher gamma power during the error trials relative to the Go trials ($p < 0.05$, permutation test). Surprising, the Conflict trials also showed significantly higher gamma power relative to the Go trials, but the Conflict related gamma increase was not as pronounced as it was in the error trials. These differences between Error, Conflict, and Go trials were very similar to the relative differences observed in the beta band, but they occurred in the opposite direction.

Rule change effects on STN beta power

It is possible that the higher levels of beta power observed during Go trials that followed Conflict trials were not due to the level of conflict on the previous trial, but rather due to the change in task rules that occurred from one trial to the next (move in the direction opposite to the arrow vs

move in the direction of the arrow). If the effects were simply due to the rule change, than a rule change in the opposite direction should produce similar electrophysiological results. We therefore compared STN beta power during correct Conflict trials that followed a correct Go trial to correct Conflict trials that followed a previous correct Conflict trial (Supplementary Fig. S13). We included 28 sessions that had at least five or more Conflict trials that followed previous Conflict trials in this analysis. This comparison showed no significant effect of the previous trial on the STN beta power of the current conflict trial. Notably, the response times of the Conflict trials that followed a Conflict trial were 83.7 ± 18.9 ms faster than those that followed a previous Go trial ($t(28) = 4.42$, $p < 0.001$, paired t-test), consistent with prior work (Gratton *et al.*, 1992). Whereas conflict induces slowing of response times when the subsequent trial is a non-conflict trial, it induces speeding of the response times when the second trial is also a conflict trial (the Gratton effect). Our results suggest that STN beta oscillations may be involved in the post conflict slowing of response times on low conflict trials. The mechanisms underlying the speeding of response times that happens on two consecutive conflict trials, however, seem not to involve STN beta oscillations in our task.

Supplemental Materials and Methods

Intraoperative task and recordings during deep brain stimulation surgery

We made intraoperative recordings in 22 participants (19 males; 58.0 ± 1.5 (mean ± SEM) years old) undergoing deep brain stimulation (DBS) surgery of the subthalamic nucleus (STN) for Parkinson's disease. The study was conducted in accordance with an NIH IRB approved protocol, and all participants gave their written informed consent to take part in the study. Participants received no financial compensation for their participation. Parkinson's disease medications were stopped on the night before surgery (12 h preoperatively). We recorded while participants were alert, at rest and supine, and in an OFF state in the operating room. Sample-size estimation computations were not conducted prior to data collection. Instead, during a two year period (April 2014-April 2016) data were collected during all DBS cases in which the patient was willing and able to participate in the task. The goal was to collect between 10-20 recording sessions during this 24 month time period. Approximately one DBS case was done per month resulting in 22 total cases.

As per routine DBS surgery, we used intraoperative microelectrode recordings to identify the STN based on firing rate and pattern (increased spiking activity and background noise relative to the more dorsal zona incerta and thalamus). We simultaneously advanced three targeting electrodes, separately spaced 2 mm apart, during each recording session (placed along a central, 2mm lateral, and 2mm anterior trajectory; Fig. 1C). Each targeting electrode consisted of a microelectrode contact and a macroelectrode contact positioned 3 mm dorsal to the microelectrode tip (Alpha Omega, Alpharetta, GA). Macroelectrode contacts were within the STN if the corresponding microelectrode contact was greater than 3 mm ventral to the dorsal border of the STN (identified by increased spiking activity and background noise relative to the more dorsal zona incerta and thalamus). We restricted all analyses only to signals captured from electrode contacts positioned within the STN. Raw signals were sampled at 1.5024 and 24.0345 kHz from macro and microelectrode contacts, respectively, and stored using a MicroGuide Pro data acquisition system (Alpha Omega Co., Alpharetta, GA). Mean coordinates of the central microelectrode recording sites during the behavioral task, referenced to the mid- commissural point, were x = 12.1 0.2, y = -4.9 0.8, and z = 4.3 0.3 for left electrode recordings, and x = -11.8

0.3, y = 5.8 0.9, and z = 5.0 0.3 for right electrode recordings. These coordinates correspond to left and right STN on the Schaltenbrand-Wahren brain atlas.

During the operative procedure, we acquired simultaneous intracranial EEG (iEEG) recordings from a subdural strip electrode temporarily placed through the DBS burrhole (PMT Corporation, Chanhassen, MN). We placed a six-contact mPFC strip electrode consisting of a single row of six platinum contacts (2.3 mm exposed diameter with 1 cm inter-contact spacing) in a direct anterior direction from the burr hole. The electrodes were placed over the superior frontal gyrus and therefore over the superior portion of the medial prefrontal cortex. Thus, we refer to the area from which we recorded as the medial prefrontal cortex (mPFC). We confirmed contact localization using intraoperative x-ray (Supplementary Fig. S1). In 20 of the 34 sessions, we also placed an eight-contact lateral PFC strip electrode in a direction that was angled approximately 60 degrees lateral to the direction of the mPFC strip electrodes as a part of a separate study. We did not include recordings from these lateral contacts in any analyses presented here. In two of the experimental sessions, we did not implant any iEEG strip electrodes. All subdural strip electrodes were removed after completion of the behavioral task on each side.

In order to estimate locations of the subdural contacts, we co-registered the post-op CT with the pre-op T1 weighted MRI using the publicly available software packages AFNI (http://afni.nimh.nih.gov) (Cox, 1996; Saad and Reynolds, 2012). Specifically, we deobliqued and skull-stripped the pre-operative T1 weighted images, center-aligned the CT to the MRI, and finally, performed an affine transformation using AFNI's 'align-epi-anat.py' command with a local Pearson correlation cost function. We deobliqued and inverted the intensity of the MRI, and center aligned the CT to the MRI. If the registration was unsuccessful, we reprocessed the CT to remove air voxels, and attempted alignment again. On the co-registered image, we identified the location of the burrhole on the pre-operative MRI (Supplementary Fig. S1). We estimated contact locations by measuring every 1 cm directly anterior from the burrhole for the six contact mPFC strips. As this procedure does not provide accurate electrode contact locations, we only used these estimates to confirm that subdural strip electrode locations approximately corresponded to the mPFC. In order to compare electrode locations across subjects, we relied on surface-based registration (Saad and Reynolds, 2012), using FreeSurfer

(http://surfer.nmr.mgh.harvard.edu) (Fischl, 2012) and AFNI to reconstruct and normalize the pial surfaces of each participant. We plotted each electrode at its corresponding mesh location on the standard N27 ”Colin” brain (Holmes *et al.*, 1998), then, for visualization purposes, projected these electrodes to an approximated dural surface and imposed each strip’s original geometric alignment. In this manner, electrodes appeared in analogous anatomical locations while laying visibly above the pial surface and in the correct geometric layout. Electrode projection estimates were unavailable for 6 of the 22 participants because one of the participants was not implanted with iEEG electrodes, and 5 additional participants had CT scans that did not include the top of the skull. Thus, the location of the burrhole could not be determined for these 5 participants.

Behavioral task

Participants made all movements with the hand contralateral to the side of intra-operative recording for that session. All participants who completed the task were sufficiently able to control the joystick. Patients with a tremor that was severe enough to interfere with control of the joystick were unable to perform the task and therefore excluded from the study.

150 ms following each response, subjects were given feedback for that trial. If the participant correctly moved the joystick in the appropriate direction for the Go or Conflict trials or if the subject correctly withheld a response in the NoGo trials, we presented a green smiley face and the word "CORRECT" in the center of the screen. If the subject moved the joystick in the incorrect direction for the Go and Conflict trials or moved the joystick in any direction for the NoGo trials, we presented a red sad face and the word "WRONG." We subsequently refer to these trials with an incorrect response as 'error' trials. The duration of the feedback stimulus was 500 ms. Following the feedback, we displayed a blank screen for a duration randomized between 1000 ± 100 ms, before presenting the warning cue for the subsequent trial.

Three strategies were employed to encourage the subjects to respond in a timely manner. First, the subjects were encouraged ahead of time to respond quickly while trying to minimize errors. Second, if the subjects took longer than 1500 ms to respond, their response was not recorded and the feedback they received consisted of a yellow neutral face and the phrase, "Please respond

faster". Third, the arrows were only displayed on the screen for 1000 ms. We discarded all trials with response times less than 300 ms and greater than 1500 ms.

On the day before surgery, participants performed a complete session in order to familiarize themselves with the task. During the operative procedure, most participants performed one session while we captured recordings from the left STN and a second session while we recorded from the right STN. Ten participants did not complete the second session because of fatigue. This resulted in 34 total intra-operative recording sessions included in the analysis. Most of the analyses we conducted included only correct trials. Accordingly, when we analyzed any across-trial changes in behavior and electrophysiology, only trials in which the current trial and the previous trial were both correct were included in the analysis. When we specifically analyzed the error related changes in electrophysiology, we only included the 20 sessions in which participants committed 5 or more errors. Due to the low number of error trials, we included any Go, Conflict, or NoGo trial in which the participant answered incorrectly in our error analysis. The mean number of error trials for these 20 sessions was 12.2 ± 0.4. When we specifically analyzed the post-error related changes that occurred on the correct Go trials that followed an error trial, we only included the 15 sessions in which participants committed 5 or more errors and had 5 or more correct Go trials that followed an error trial. The mean number of errors for these 15 sessions was 7.7 ± 0.3.

LFP and iEEG power

We performed all analyses using MATLAB (Mathworks, Natick, MA). We extracted local field potential (LFP) activity from each macroelectrode and iEEG activity from each subdural contact. We bandpass filtered both signals between 1 and 500 Hz, notch filtering at 60 Hz, and downsampled the data to 1 kHz. We referenced the macroelectrode and iEEG signals by subtracting the signals of adjacent electrodes. For each session, this generated three referenced bipolar LFP channels for the STN and five referenced bipolar signals for the mPFC strip. We henceforth refer to these bipolar channels as electrode contacts. Prior to any subsequent analysis, we discarded all trials exhibiting a clear artifact in the LFP or iEEG trace.

In order to obtain magnitude and instantaneous phase information in the frequency domain, we convolved the LFP signals captured from the STN and iEEG signals captured from the subdural contacts from each trial with complex valued Morlet wavelets (wave number 6). We used 47 logarithmically spaced (8 scales/octave) wavelets between 2 and 107 Hz and convolved each wavelet with 3000 ms of LFP data from each trial. For cue-locked analyses, we analyzed LFP signals from 1000 ms before to 2000 ms following arrow presentation. For response-locked analyses, we analyzed LFP signals from 1500 ms before to 1500 ms after the response. We used a 1000 ms buffer on both sides of the clipped data to eliminate edge effects. We squared the magnitude of the continuous-time wavelet transform to generate a continuous measure of instantaneous power for each frequency. We determined the z-scored power from each channel and frequency using the mean and standard deviation of the power recorded from that channel during a baseline period. We defined the baseline period as the 500 ms preceding the presentation of the fixation cue in each trial, and used the mean power during this baseline period to normalize all cue and response locked analyses. During this time, the participants were staring at a blank screen between two adjacent trials. Though several of our findings show significant differences during the time period in between adjacent trials, none of these differences extended into this baseline period. Overall, very similar difference between trial types were obtained when we reanalyzed the data using the 500 ms period of the actual warning cue as our baseline (data not shown).

For each STN recording, we averaged the normalized power from macroelectrodes that were within the STN as identified during the operative procedure. Because we did not have access to intraoperative computed tomography (CT) imaging, we were unable to use patient specific landmarks to accurately localize individual iEEG contacts. Thus, for each iEEG strip, we averaged the normalized power from all five bipolar channels recorded from that strip. This procedure resulted in one spectrogram for the STN and one for the mPFC during each trial.

Statistical analysis

To assess differences in spectral power between conditions across recording sessions, we first calculated the trial-averaged normalized power for all three conditions in each region in each

experimental session. In order to prevent the 3:1:1 ratio of Go:Conflict:NoGo trials from affecting our results, we subsampled the correct Go trials prior to taking the average. Subsampling was achieved by randomly selecting a subset of the correct Go trials to match the number of correct Conflict trials and then calculating the average across only that subset of correct Go trials. We repeated this process 1000 times, and calculated the mean subsampled value across those 1000 iterations to compare the mean response during the Go trials to the mean response during the Conflict trials. Due to the difference in error rates between the Conflict and the NoGo trials, we also subsampled the correct NoGo trials to match the number of correct Conflict trials. We were therefore left with an average normalized power for Go, Conflict, and NoGo trials of similar trial number at each time point and frequency for each session. We also used subsampling when comparing the error trials to correct trials or when comparing the Go trials that followed a previous Go trial to the Go trials that followed a Conflict trial. To test for any trial type related differences in power, we performed random-effects statistical analyses in each region across sessions. For each comparison, our null hypothesis was that across sessions, there was no difference in normalized power between trial types (Go versus Conflict, NoGo versus Go, NoGo versus Conflict, etc). In order to assess overall changes in power during the task regardless of condition, we used a similar statistical analysis. In this case, our null hypothesis was that the average power across all trials was not different from zero. We tested these hypotheses using a non-parametric permutation procedure in which the session is the unit of observation (Maris and Oostenveld, 2007).

In each region, we computed the true mean difference across sessions between the two conditions being compared (Go versus Conflict, NoGo versus Go, NoGo versus Conflict, all trials versus baseline, etc) for every time point and frequency. We then randomly permuted the condition specific averages for each session and recomputed the mean difference across sessions. We repeated the permutation 1000 times to generate an empirical distribution of possible mean differences that were all equally probable under the null hypothesis. For every time-frequency point, we compared the true mean difference to the mean and standard deviation of the corresponding point in the empirical distribution to generate a p-value. This p-value represents the likelihood that the true mean difference for each time-frequency point represents a departure

from the null hypothesis. However, this p-value for each time-frequency point does not take into account the multiple comparisons that are made across all time points and frequencies.

To correct for multiple comparisons across all time points and frequencies, we used a cluster correction method based on exceedance mass testing (Maris and Oostenveld, 2007). This method assumes that a true effect at any time-frequency point is likely to be observed across multiple time points and frequencies. We defined time-frequency clusters by thresholding the across-session p-values derived from the statistical analysis described above. Any contiguous time-frequency points with a p-value less than 0.05 were included in each cluster. For each identified cluster, we defined a cluster statistic to be the sum of the z-scores, derived from the p-value using a normal cumulative distribution function, for all time-frequency points within that cluster.

We calculated clusters using the true data, and for each of the 1000 permutations of the session-specific trial averages. We used the maximum cluster statistic of each permutation to create an empirical distribution for significance testing. We determined whether a true cluster test statistic was significant by comparing it to the empirical distribution of maximum cluster test statistics. In this manner, significant clusters can arise from large differences between trial types that extend over a small number of frequencies or over a small time period, or from smaller differences that involve a larger number of time-frequency points. We considered cluster test statistics with $p < 0.05$ to be significant and corrected for multiple comparisons.

We also tested for significant differences in power within two specific frequency bands of interest (2 - 5 Hz and 8-30 Hz) based on the overall changes in power observed during the task relative to the baseline period. We used the same permutation procedure described above, but first averaged spectral power across the frequency band of interest prior to calculating any true or surrogate differences between conditions. In this case, clusters were based on contiguous time points exhibiting significant differences across sessions.

Finally, we tested whether there was a relationship between single-trial changes in normalized power and response time. We first averaged, for each trial, the normalized power during a selected time period of interest. For the theta band, this time window was the beginning of each

trial (arrow onset) to the response. For the beta band, the time window of interest was the first 500 ms after the arrow onset. These time windows were chosen as they corresponded to the points in the task that showed the most robust differences between trial types. We also repeated the analysis using the baseline period time window as a control. At each of the time windows of interest, we correlated the average value for each trial's power with that trial's response times using Spearman's correlation. The resulting correlation coefficients were then averaged across macroelectrode contracts for the STN and all bipolar contacts for the iEEG recordings prior to being averaged across recording sessions. We used a two-tailed, one-sample t-test to determine if the mean correlation was significantly different from zero across sessionss and a two-sample t-test to determine if there was a significant difference in the correlation coefficients between Go and Conflict trials.

STN-cortical phase coherence

To estimate the time-varying inter-site phase coherence, between the STN and the mPFC, we used the continuous time wavelet transform data extracted above for each trial and each frequency (see 'LFP and iEEG Power'). We first calculated the difference at each time-frequency point between the instantaneous phase (projected on the complex plane) in the LFP signal and the instantaneous phase in the iEEG signal. To generate a continuous-time estimate of phase coherence, we calculated the magnitude of the average difference over 250 ms sliding windows (step size 1 ms) (Lachaux *et al.*, 2002). This results in a time-varying estimate of phase coherence for every frequency, for every trial, and for every pair of electrodes.

We determined the z-scored phase coherence for each pair of electrodes at each time-frequency point by comparing the continuous measure of phase coherence to the mean and standard deviation of the phase coherence recorded from that pair of electrodes during the baseline period. As with the analysis of power, we subsampled the Go and NoGo trials to match the number of correct Conflict trials and then averaged the z-scored phase coherence across trials. We averaged the z-scored phase coherence across all contacts within the iEEG strip and then averaged across all macroelectrodes that were within the STN as identified during the operative procedure. We therefore generated a single average phase coherence spectrogram between the mPFC and the

STN for each condition for every session. In order to compare differences in phase coherence between conditions across all sessions, we used the same across-session permutation procedure described above.

Spiking activity

We extracted spiking activity by bandpass filtering microelectrode recordings between 0.3 and 3 kHz and resampling the filtered signals at 24 kHz. For the purposes of plotting spiking activity raster plots and generating a time series of spiking activity (Fig. 2), we identified spike events by manually setting a negative or positive voltage threshold depending on the direction of the voltage deflection (Plexon Offline Sorter, Inc., Dallas, TX). Given the difficulty of isolating single unit activity in the STN (Weinberger *et al.*, 2006; Sharott *et al.*, 2014), we used a previously outlined technique (Stark and Abeles, 2007) to identify STN multiunit activity (MUA). We measured the MUA by clipping extreme values in the band passed filtered data (larger or smaller than the mean 2 STDs) and computing the root mean squared (RMS). We computed the RMS by squaring the data, low pass filtering at 100 Hz, downsampling to 1000 Hz, and taking the square root. Finally, we smoothed the resulting MUA data using a Gaussian kernel (standard deviation 50 ms). Though most of our analysis of STN spiking involved the MUA data, repeating the analysis using the more traditional voltage threshold method produced very similar patterns of activity for task responsive recordings (see Supplementary Fig. S4 and Supplementary Fig.S5). For this confirmatory analysis, we calculated the continuous-time firing rates for each recording by smoothing the spike train (generated by manual thresholding, see above) from each trial (1 ms bins) with a Gaussian kernel (standard deviation 50 ms). The MUA analysis was chosen over this method as it seemed less prone to noise in the recording and did not depend on a manually chosen arbitrary threshold being applied to all time periods of the recording. It is important to note, however, that MUA activity is thought to reflect a composite of the highly focal firing rate of neurons, bursting, recruitment, and synchronization effects (Moran *et al.*, 2008).

We extracted 3000 ms of MUA data from each trial for each microelectrode. We excluded all trials with an MUA greater or less than 15 standard deviations from the average MUA across all

trials. To generate a normalized average MUA, we compared MUA for each trial to the mean and standard deviation of the MUA firing during the baseline period and then averaged across trials. To determine whether an individual microelectrode recording exhibited a significant change in MUA during the task, our null hypothesis was that the average MUA across all correct trials (Go, Conflict, NoGo) was not different than zero. For each recording, we used the same permutation procedure described above to compare the true mean continuous MUA across all trials to an equally sized trial- by-time matrix of zeros. For these within-subject comparisons, however, the unit of observation being permuted was the individual trial. Any MUA recording with a contiguous cluster of p-values < 0.05 was considered responsive.

From the 34 STNs we included in our analyses, we identified 49 MUA recordings (from 22 different STNs) that were within the borders of the STN. 39 (79.6%) of the recordings (from 20 different STNs) showed an increase in MUA firing during the task. Out of the 10 recordings that did not show a significant increase in MUA during the task, 5 showed a significant decrease, and 5 showed no changes. Nevertheless, for the across-session analysis of trial-type related MUA changes all 49 recordings were included. Rather than treat each of these 49 recordings as independent, however, we chose to combine recordings that were acquired from the same STN. Had we not done so, we would run the risk of artificially increasing the number of MUA recordings included in our statistical analysis as MUAs recorded from the same STN were more than likely not independent. Indeed, we found that there was often a similar pattern of task related MUA changes in different electrodes recorded from the same STN. To address this issue, we average the normalized MUA of responsive electrodes recorded from the same STN. Thus, each of the 22 STN recording sessions with spiking activity resulted in just one measurement of MUA activity that we subsequently used in our across-session analyses when comparing trial types.

To determine if the STN MUAs exhibited a significant difference between trial types (Go versus Conflict, NoGo versus Go, NoGo versus Conflict, etc) across sessions, for every STN we calculated the mean difference in firing at every time point between trial types. We then calculated the average difference across all 22 STNs and compared the difference to an empirical distribution generated by permuting the condition labels of each session's average data 1000

times. We corrected for multiple comparisons across time using the same nonparametric clustering-based procedure described above.

Phase-MUA Coupling

To quantify the degree of coupling between macro electrode LFP phase and micro electrode MUA amplitude, we used a modified version of the techniques developed by Canolty and colleagues (2006) to quantify phase-amplitude coupling. The traditional method of calculating phase amplitude coupling involves calculating the instantaneous phase values for one frequency during a specified time window and then calculating the instantaneous amplitude at a higher frequency for the same time period. The phase information is subsequently converted into the complex space using the Euler transform, and each complex value phase vector is then multiplied by the amplitude of the high frequency oscillation for the corresponding time point. The magnitude of the average vector is then calculated across the entire time period to get a measurement of the degree of coupling between the low frequency phase and the high frequency amplitude. Finally, the resulting phase amplitude coupling value is normalized by calculating the z-score relative to the mean and standard distribution of surrogate phase amplitude coupling values generated by shuffling the amplitude time series prior to multiplying it by the phase time series. To quantify the degree of phase-MUA coupling (PMUAC) in our data, we modified this procedure by replacing the instantaneous power of a high frequency oscillation with the magnitude of the MUA microelectrode recording and analyzed its coupling to the phase of the corresponding LFP macroelectrode recording for each session.

$$PMUAC = \left| \frac{1}{n} \sum_{t=1}^{n} a_t e^{i\varphi_t} \right|$$

Where n signifies the total number of time points, a_t the modulated amplitude and φ_t the phase of the modulating frequency at time point t; i is the imaginary operator. A shortcoming of this method is that non-uniform distributions of phase values can bias the results if the mean phase vector is of non-zero amplitude. This can happen, for example, when a task related stimulus or a motor response resets the phase values to a similar phase across all trials (sometimes called ’inter trial phase clustering’ or ’inter trial phase coupling’, (Cohen, 2014)). Whether the inter trial phase clustering bias increases or decreases the estimated phase amplitude coupling value

depends on whether the mean phase vector points in a similar or opposite direction as the mean phase amplitude coupling vector. The method developed by Van Driel and colleagues (2015) corrects for this bias by subtracting the mean phase vector prior to averaging across all phase-times-amplitude vectors.

$$dPMUAC = \left| \frac{1}{n} \sum_{t=1}^{n} a_t (e^{i\varphi_t} - \bar{\Phi}) \right|$$

where $\bar{\Phi}$ is defined by

$$\bar{\Phi} = \frac{1}{n} \sum_{t=1}^{n} e^{i\varphi_t}$$

The remaining steps of this method (including the shuffling of power values relative to phase values to generate a surrogate distribution) are the same as those used in the original method developed by Canolty and colleagues.

To generate a continuous metric of dPMUAC, we first down-sampled all STN LFP and MUA data to 100 Hz to decrease computation time. We obtained the oscillatory phase of the macro electrode by convolving the raw LFP signals with either the complex valued Morlet wavelets when looking at all frequencies (wave number 6, 47 logarithmically spaced wavelets between 2 and 107 Hz) or with the Hilbert transform when analyzing the changes in the theta and beta bands. We then windowed each trial into 250 ms long time periods (step size 10 ms). For each 250 ms window, we concatenated the phase and amplitude values for all trials generating a phase time series and an amplitude time series for that time window that was n*250 ms long, where n is the number of trials in that session. We then used the dPMUAC calculation described above to generate a phase-MUA coupling value for that time window. We normalized the resulting value by permuting the amplitude information across trials. Whereas in the true case, phase values from a given trial were multiplied by an instantaneous MUA amplitude value from that same time point in that same trial, in each permuted case we assigned to each phase value an instantaneous MUA amplitude value from the same time point in a different trial drawn at random from the remaining trials. We permuted amplitude information 1000 times resulting in a distribution of 1000 surrogate values for each time window. Finally, we compared the true dPMUAC values with the mean and standard deviation of the distribution of permuted values to generate normalized dPMUAC values for each time window.

We calculated the task related changes in normalized dPMUAC by subtracting for each time point the normalized dPMUAC value observed during the baseline period. We determined if individual time points exhibited significant changes from baseline by using the same across-session permutation procedure described above (Maris and Oostenveld, 2007). As with our analysis of trial type related differences in MUA, all MUA recordings from the same STN recording session were averaged together prior to any across session comparisons.

We also executed this process separately for the Go, Conflict, and NoGo trials. As with the analysis of power, we subsampled the Go and NoGo trials such that n matched the number of correct Conflict trials. This yielded for each time window a debiased phase-MUA coupling value for each trial type ($dPMUAC_{Go}$, $dPMUAC_{Conflict}$, $dPMUAC_{NoGo}$). When normalizing these values to their respective surrogate distributions, we permuted MUA amplitude information separately for Go, Conflict, and NoGo trials to prevent any biases that may emerge from non-uniform phase or power distributions caused by the trial type (i.e., $dPMUAC_{Go}$ values were normalized using only Go trials, $dPMUAC_{Conflict}$ values were normalized using only Conflict trials, etc). In this manner, we normalized dPMUAC values observed for a given trial type by the probability of observing dPMUAC by chance given the distribution of phases and amplitudes observed during that trial type. This procedure resulted in a continuous spectrogram of normalize dPMUAC measurements for each of the three conditions for each recording session. We determined if individual time points exhibited significant trial-type related differences across all sessions by using the same across-session permutation procedure described above (Maris and Oostenveld, 2007).

Finally, we repeated the phase-spike coupling analysis using the thresholded spiking data instead of the MUA. We used the same methods we have previously described in detail to quantify the interactions between individual spike events and oscillatory phase (Zavala *et al.*, 2017). Briefly, for every detected spike event, we retained the instantaneous beta phase in the corresponding LFP recording. We then calculated the amplitude of the mean phase vector for all spike events in all trials during a 250 ms time window (10 ms step). Finally, we normalized this value to a permuted dataset generated by scrambling across trials in the same way that we did for the

phase-MUA coupling analysis. We used the same across-subject permutation procedure described above (Maris and Oostenveld, 2007) to test for any trial type related differences across sessions.

Supplementary References

Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, et al. High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. Science 2006; 313: 1626–1628.

Cohen MX. Analyzing Neural Time Series Data: Theory and Practice. Cambridge, MA: MIT Press; 2014.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. Int. J. 1996; 29: 162–173.

van Driel J, Cox R, Cohen MX. Phase-clustering bias in phase–amplitude cross-frequency coupling and its removal. J. Neurosci. Methods 2015; 254: 60–72.

Fischl B. FreeSurfer. NeuroImage 2012; 62: 774–781.

Gratton G, Coles MGH, Donchin E. Optimizing the use of information: Strategic control of activation of responses. J. Exp. Psychol. Gen. 1992; 121: 480–506.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 1998; 22: 324–333.

Lachaux J-P, Lutz A, Rudrauf D, Cosmelli D, Le Van Quyen M, Martinerie J, et al. Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. Neurophysiol. Clin. Neurophysiol. 2002; 32: 157–174.

Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 2007; 164: 177–190.

Moran A, Bergman H, Israel Z, Bar-Gad I. Subthalamic nucleus functional organization revealed by parkinsonian neuronal oscillations and synchrony. Brain 2008; 131: 3395–3409.

Saad ZS, Reynolds RC. SUMA. NeuroImage 2012; 62: 768–773.

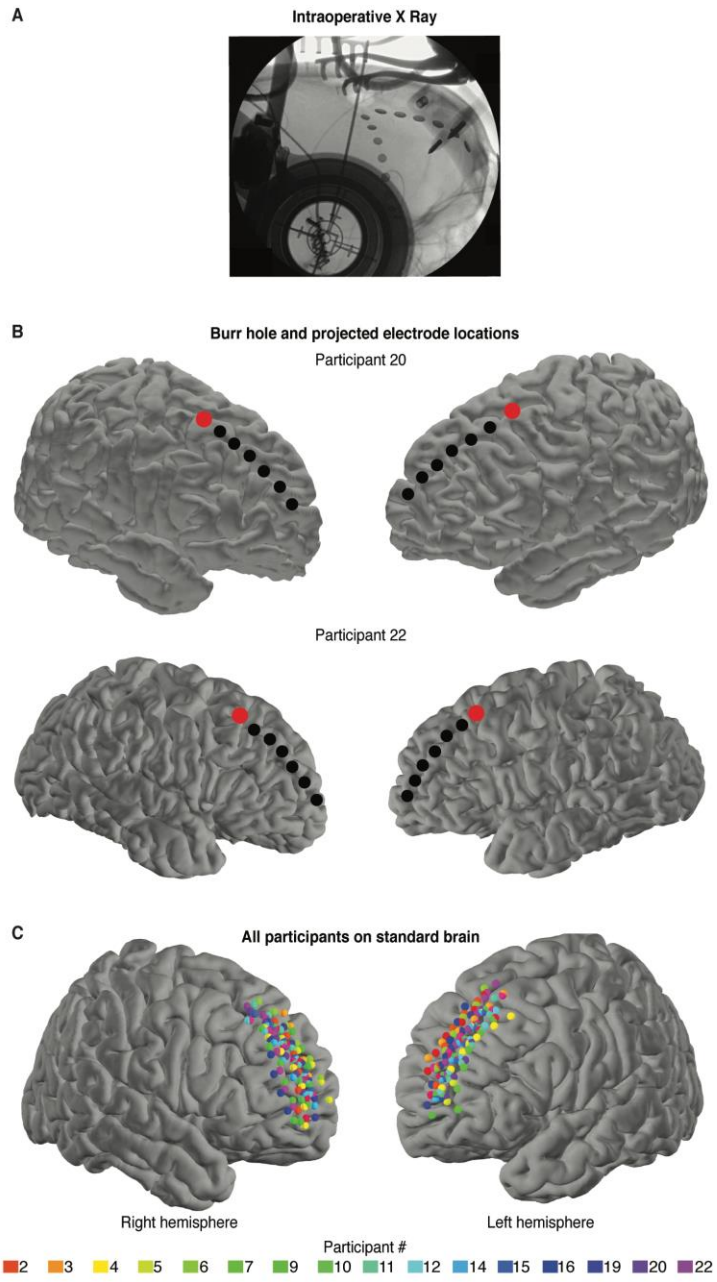
Sharott A, Gulberti A, Zittel S, Jones AAT, Fickel U, Münchau A, et al. Activity Parameters of Subthalamic Nucleus Neurons Selectively Predict Motor Symptom Severity in Parkinson's Disease. J. Neurosci. 2014; 34: 6273–6285.

Stark E, Abeles M. Predicting Movement from Multiunit Activity. J. Neurosci. 2007; 27: 8387–8394.

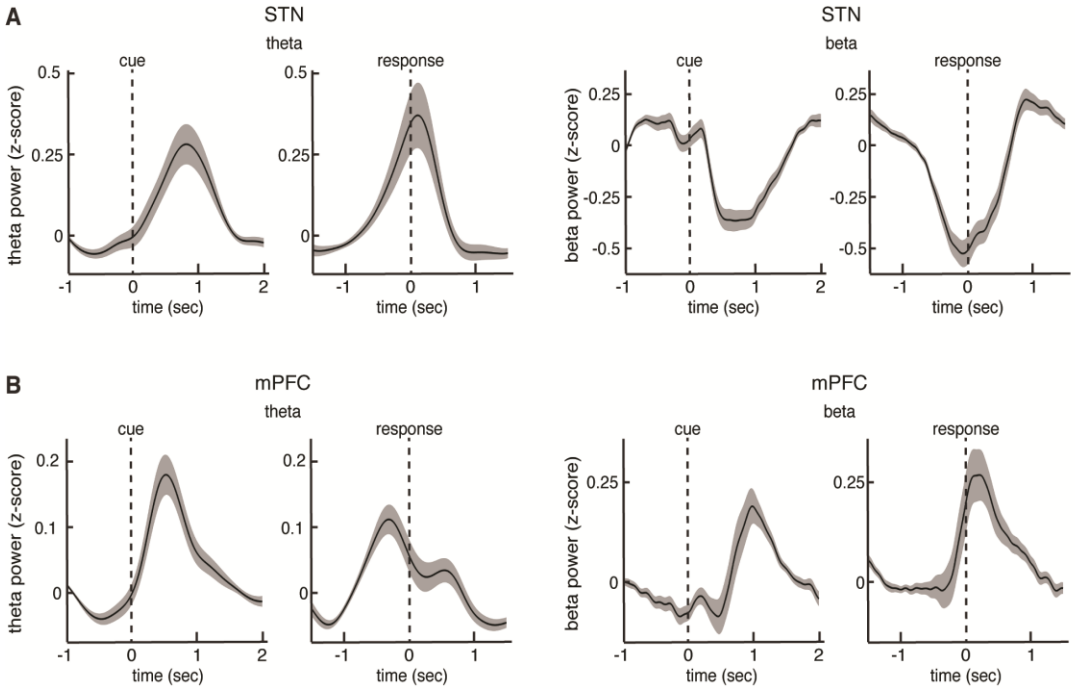
Weinberger M, Mahant N, Hutchison WD, Lozano AM, Moro E, Hodaie M, et al. Beta Oscillatory Activity in the Subthalamic Nucleus and Its Relation to Dopaminergic Response in Parkinson's Disease. J. Neurophysiol. 2006; 96: 3248–3256.

Zavala B, Damera S, Dong JW, Lungu C, Brown P, Zaghloul KA. Human Subthalamic Nucleus Theta and Beta Oscillations Entrain Neuronal Firing During Sensorimotor Conflict. Cereb. Cortex 2017; 27: 496–508.

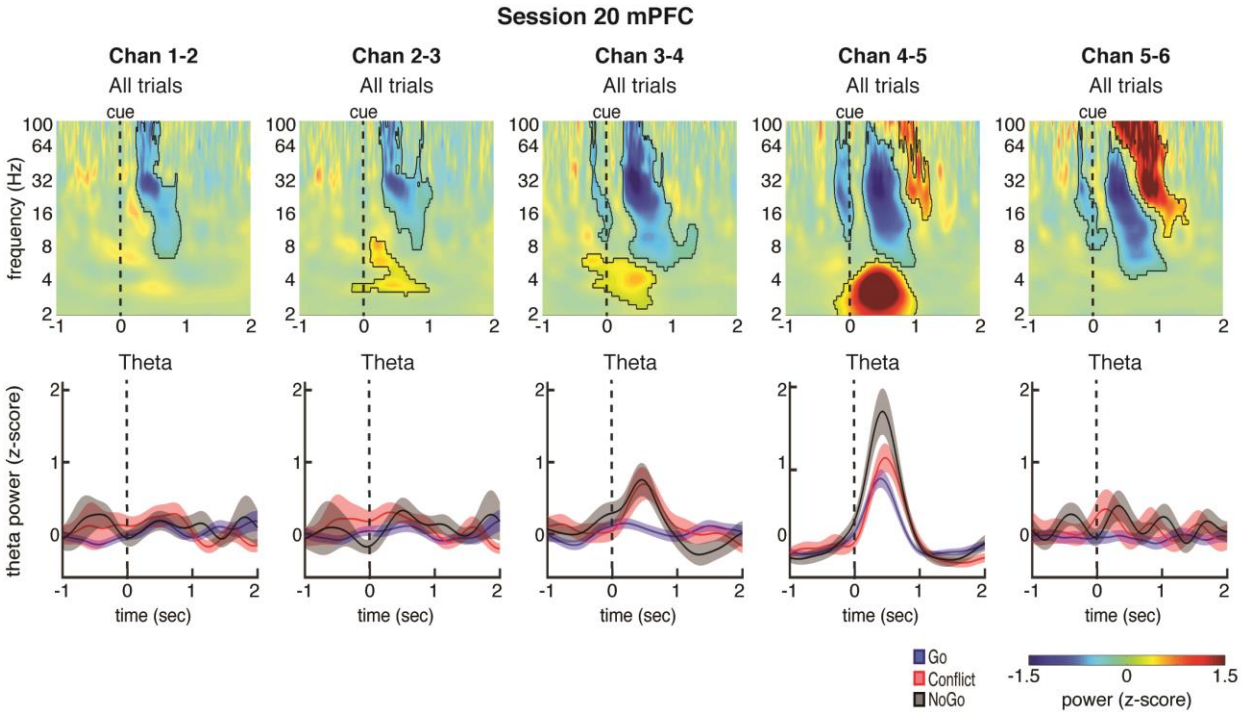
Supplementary Figure Legends



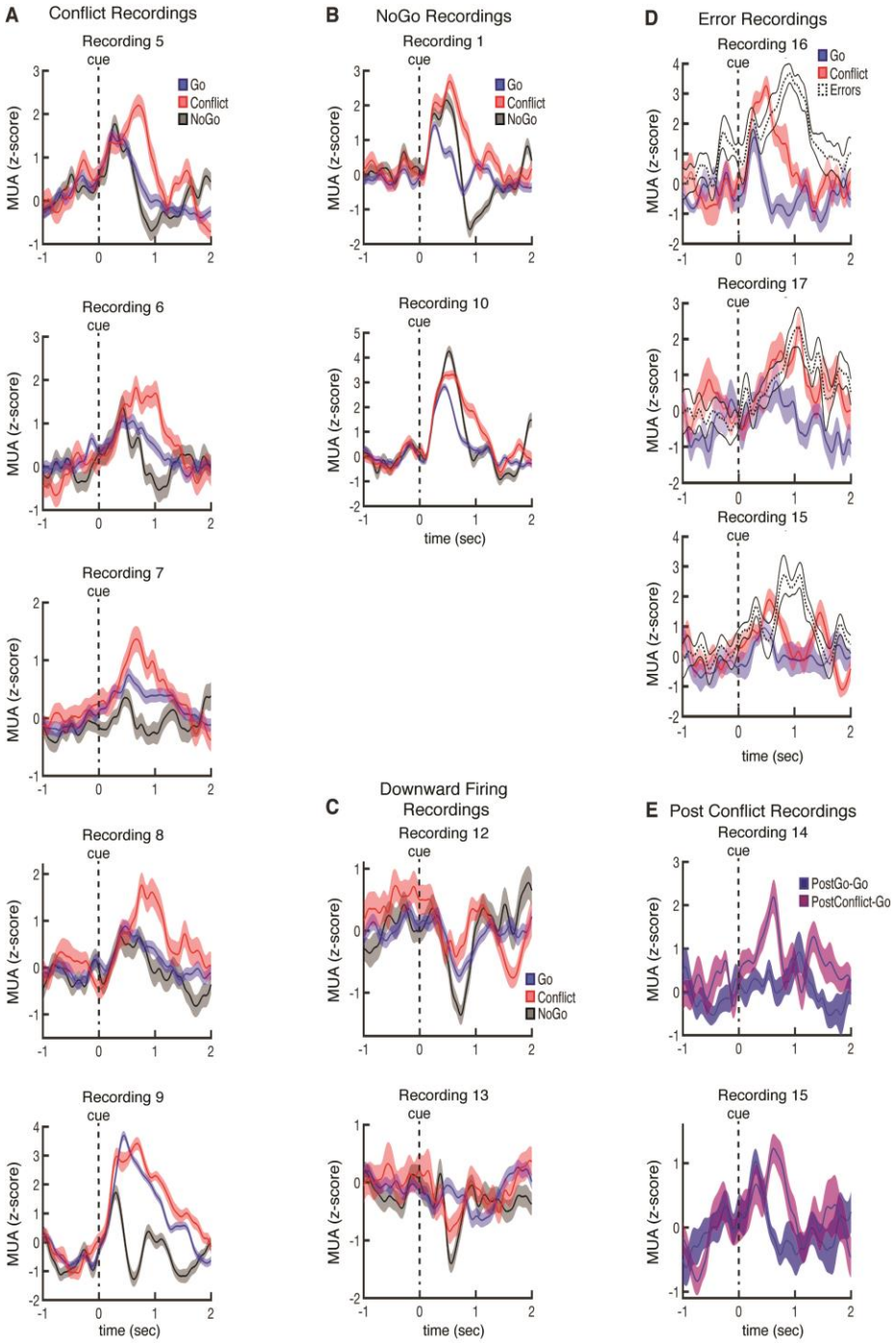
Supplementary Figure S1. iEEG electrode locations. (A) Intraoperative X-ray for one participant shows the location of anterior and lateral PFC electrodes relative to the burr hole. (B) Reconstructed brain surface for two participants. The burr hole location, identified by coregistering the pre-operative MRI with the postoperative CT, is indicated by the red dot. Estimated iEEG electrode locations for the mPFC are indicated by black dots. (C) Projected location of the iEEG electrodes for 16 of the 22 of the participants plotted on a standard brain. Electrode projection estimates were unavailable for 6 of the participants.



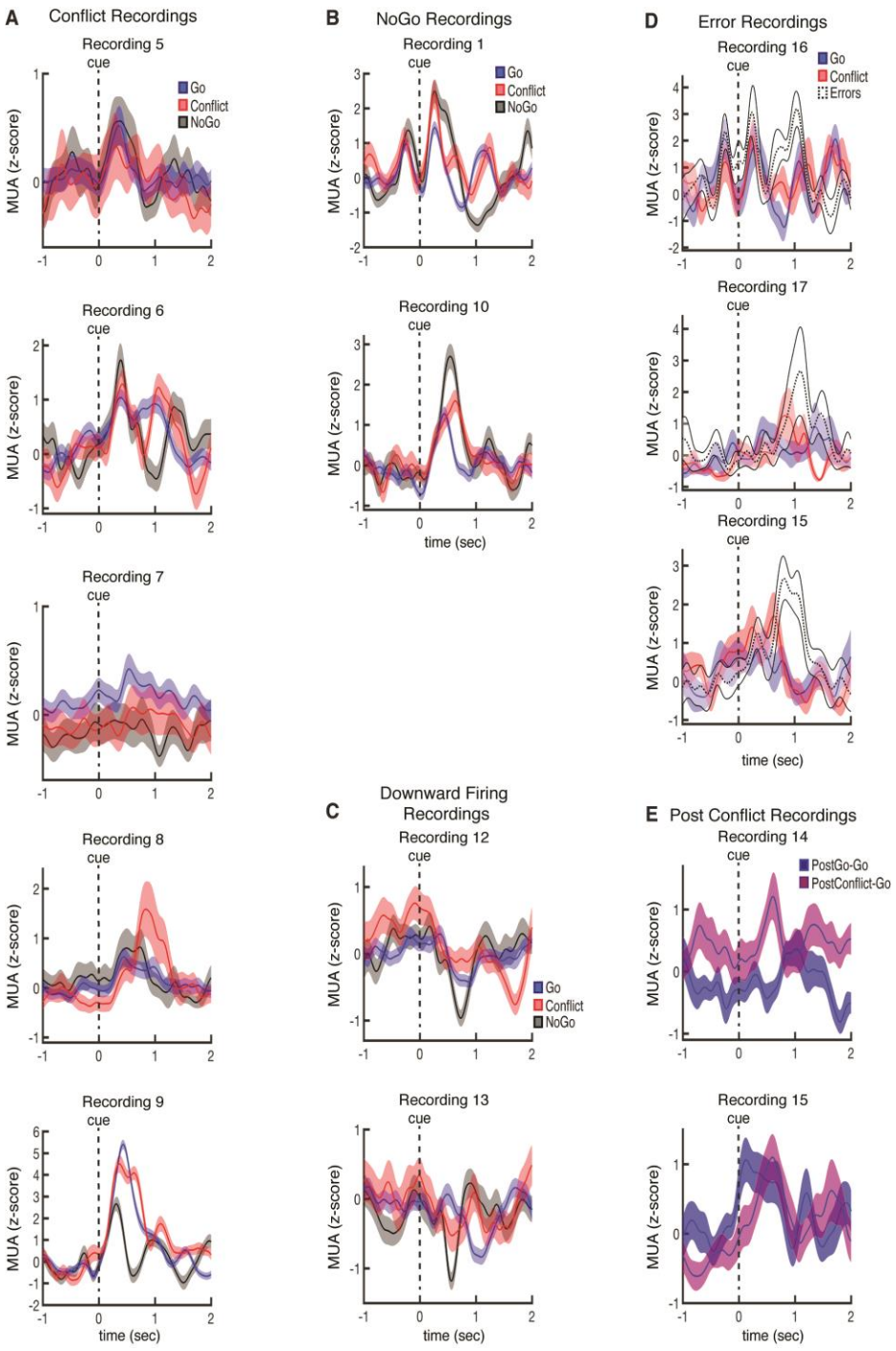
Supplementary Figure S2. Task related changes in STN and mPFC theta and beta power relative to baseline. (A) Same data as Fig. 1D, but plotted separately for the STN theta (top) and STN beta bands (bottom). (B) Same data as Fig. 1E, but plotted separately for the mPFC theta (top) and mPFC beta bands (bottom).



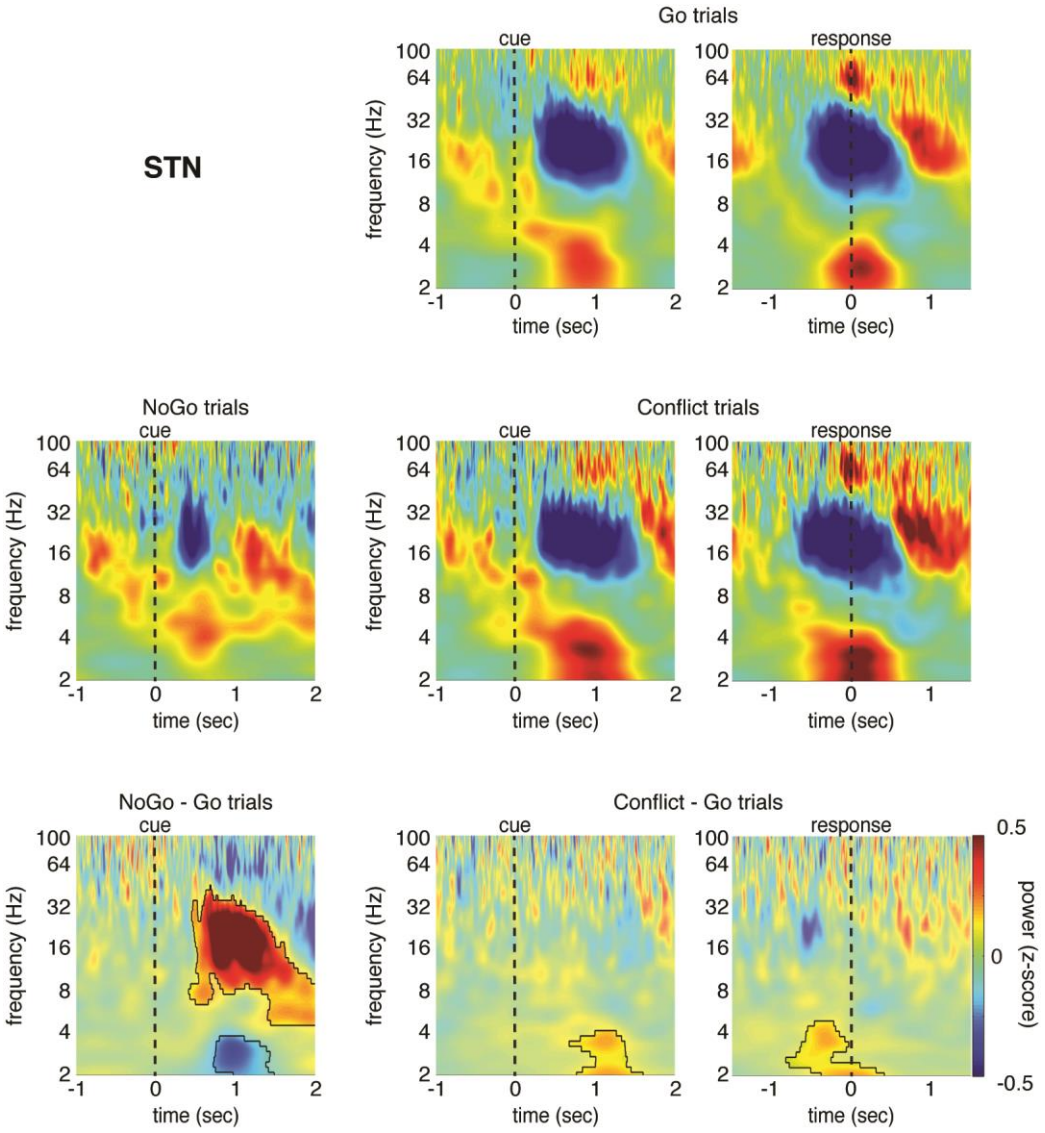
Supplementary Figure S3. Within-contact analysis of task related changes in mPFC power for one session. (Top) Normalized oscillatory power averaged across all correct trials. Each column corresponds to one bipolar contact recorded from the mPFC iEEG electrode strip during recording session 20. Data are aligned to arrow onset (t=0); mask indicates time-frequency regions exhibiting significant differences from baseline ($p < 0.05$, corrected for multiple comparisons, permutation test). (Bottom) Same as (Top) but restricted to the theta band and analyzed separately for the Go, Conflict, and NoGo trials. The theta power changes are most pronounced in bipolar contacts 3-4 and 4-5, which also show the greatest difference between trial types.



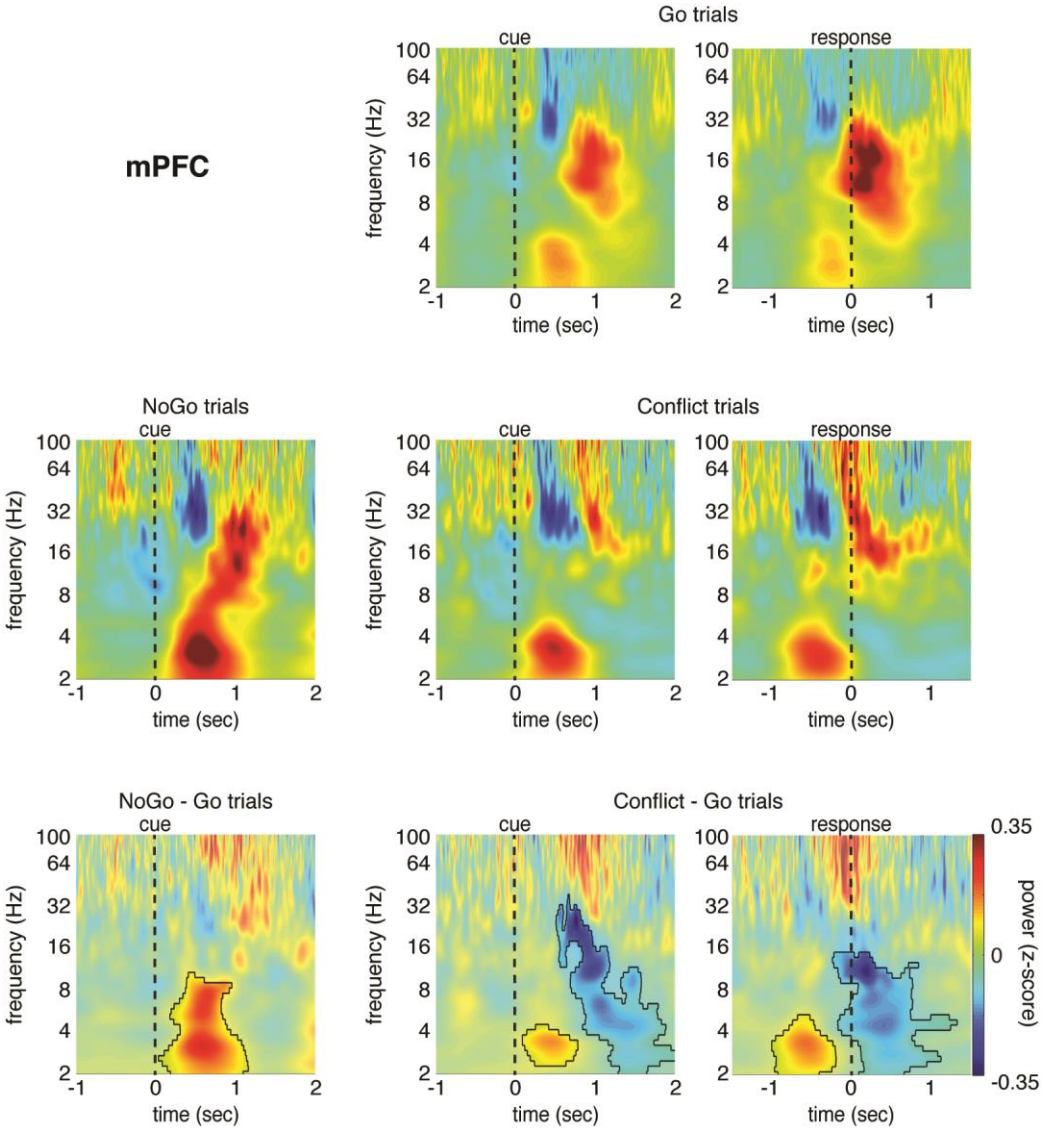
Supplementary Figure S4. Individual examples of heterogeneous STN MUA patterns during the task. (A). Five MUA recordings showing the same pattern as that seen when all MUA recordings were averaged (Conflict MUA > Go MUA > NoGo MUA). (B) Two MUA recordings showing higher MUA for the NoGo condition. (C) Two MUA recordings showing a decrease in activity during the task. (D) Three MUA recordings showing higher activity following errors. (E) Two MUA recordings showing a higher activity during the Go trials that followed conflict (postConflict-Go trials).



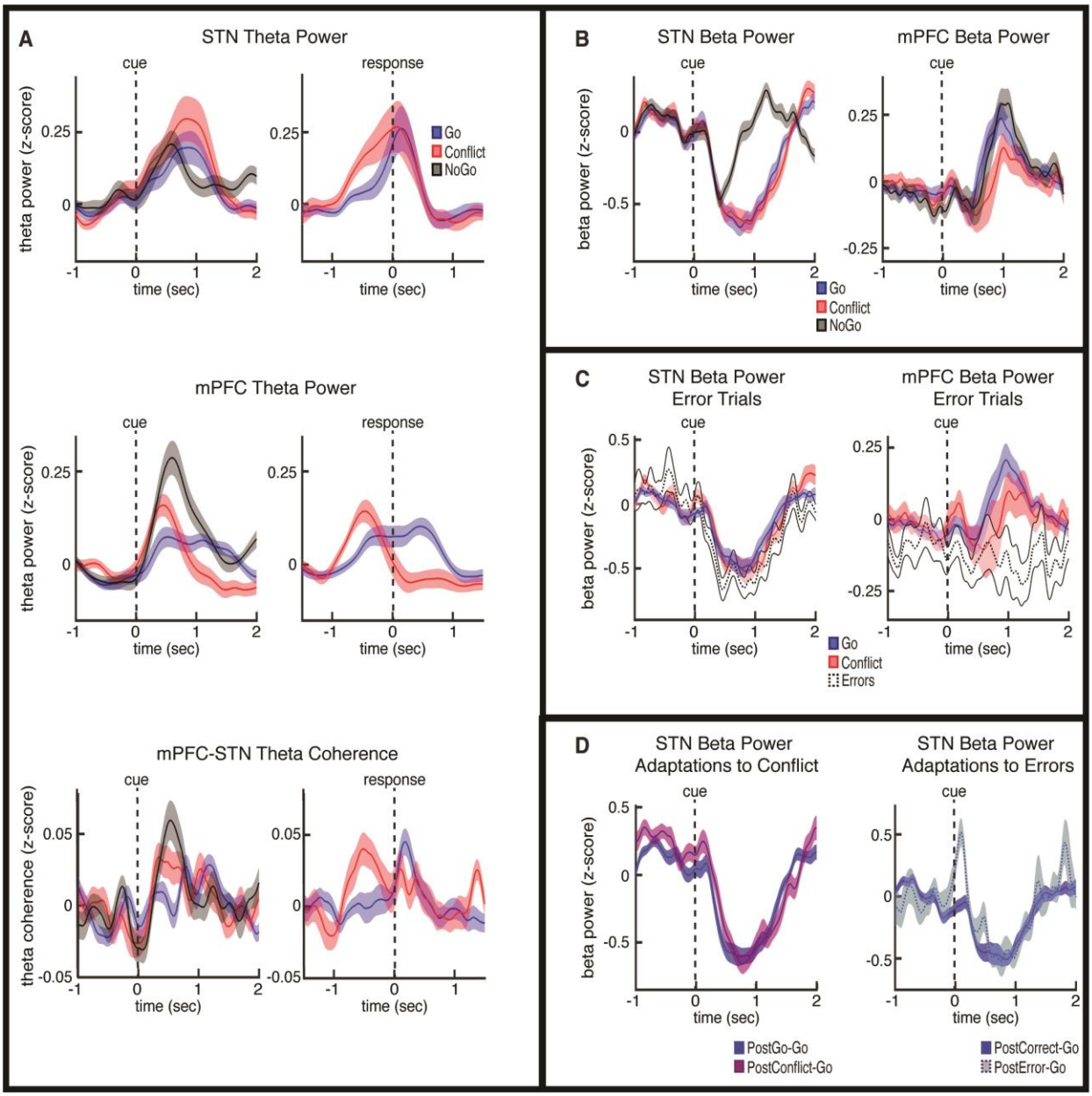
Supplementary Figure S5. Individual examples of heterogeneous STN firing patterns during the task: thresholded spiking analysis. Same data as Supplementary Fig. S4, but rather than use MUA to quantify spiking we used traditional voltage thresholding methods to detect individual spike events.



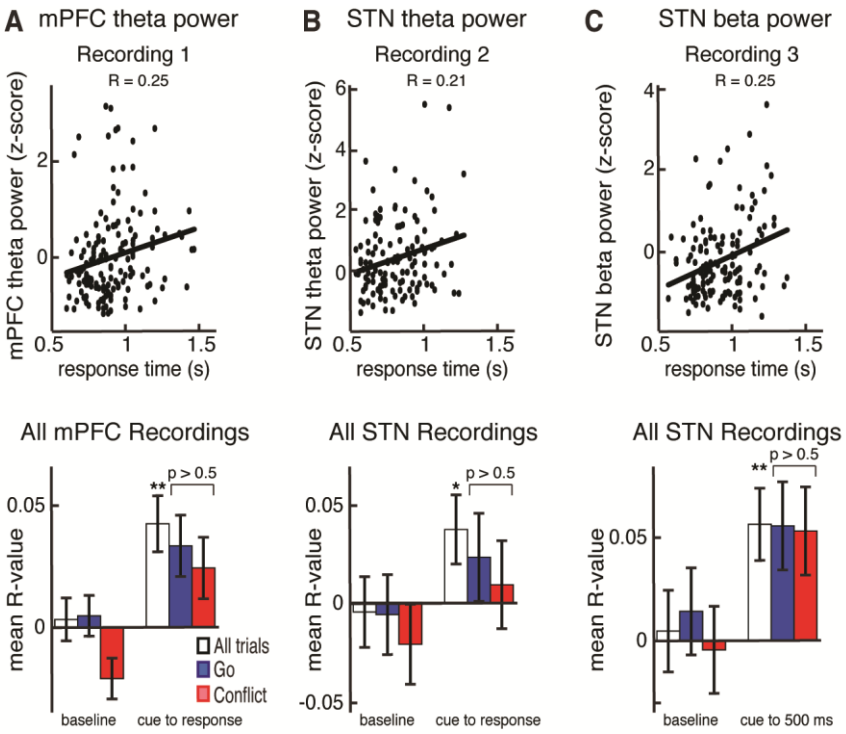
Supplementary Figure S6. Summary of trial type related differences in STN power. Same data from Fig. 1D but plotted separately for the Go, Conflict, and NoGo trials. All three trials types are plotted aligned to arrow onset (left, t=0). The Go and Conflict trials are also plotted aligned to the motor response (right, t=0). Mask indicates time-frequency regions exhibiting significant differences between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test). NoGo trials were associated with increased beta power relative to Go trials, and Conflict trials were associated with increased theta power.



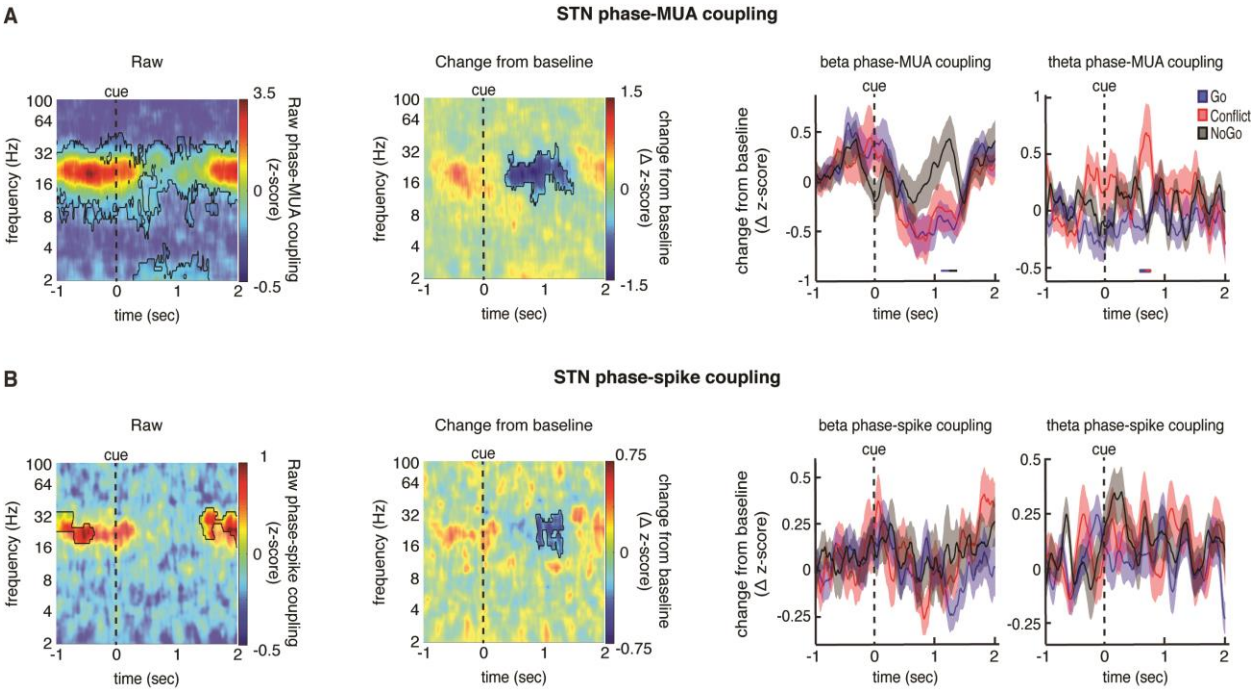
Supplementary Figure S7. Summary of trial type related differences in mPFC power. Same data from Fig. 1E but plotted separately for the Go, Conflict, and NoGo trials. All three trials types are plotted aligned to arrow onset (left, t=0). The Go and Conflict are also plotted aligned to the motor response (right, t=0). Mask indicates time-frequency regions exhibiting significant differences between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test). NoGo and Conflict trials were associated with increased theta power relative to Go trials. After the response was made, Go and and NoGo trials showed beta power increases that were attenuated in the Conflict trials.



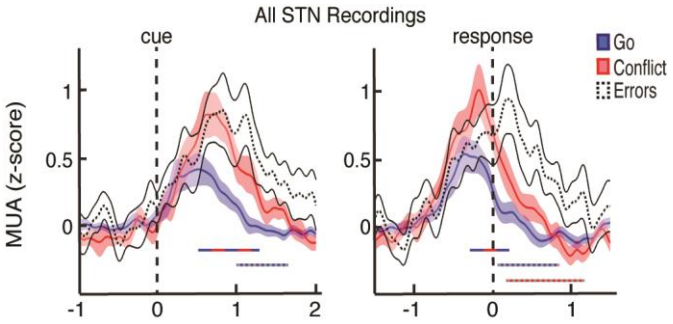
Supplementary Figure S8. Principal findings of the manuscript reproduced using 2-8 Hz and 12-30 Hz frequency ranges. (A) Same as Figure 3, but reanalyzed using 2-8 Hz for the theta band. (B) Same as Figure 4, but reanalyzed using 12-30 Hz for the beta band. (C) Same as Figure 5, but reanalyzed using 12-30 Hz for the beta band. (D) Same as Figure 6, but reanalyzed using 12-30 Hz for the beta band.



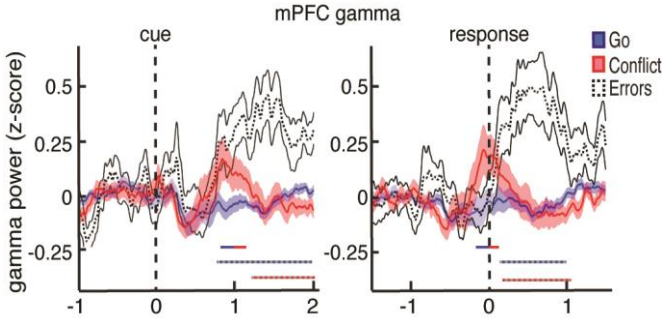
Supplementary Figure S9. Correlations between response time and either theta or beta power. (A) Top. Scatter plot of response time (x axis) and normalized mPFC theta power (y axis) for one exemplar subject. Black line denotes best fit line. Bottom. Mean correlation across all sessions of response time with mPFC theta power during the baseline period (left) and during the task from stimulus onset to response (right). Across-session averages of within-session Spearman correlation coefficient are shown for all correct trials that involved a response (i.e. correct Go and Conflict trials combined) and for the Go and Conflict trials separately. * denotes $p < 0.05$. ** denotes $p < 0.01$. (B) same as (A) but for the STN theta power. (C) same as (A) but for the STN beta power during the first 500 ms of each trial. mPFC theta power, STN theta power, and STN beta power significantly correlated with response time across all correct trials. There were no significant difference between Go and Conflict trials.



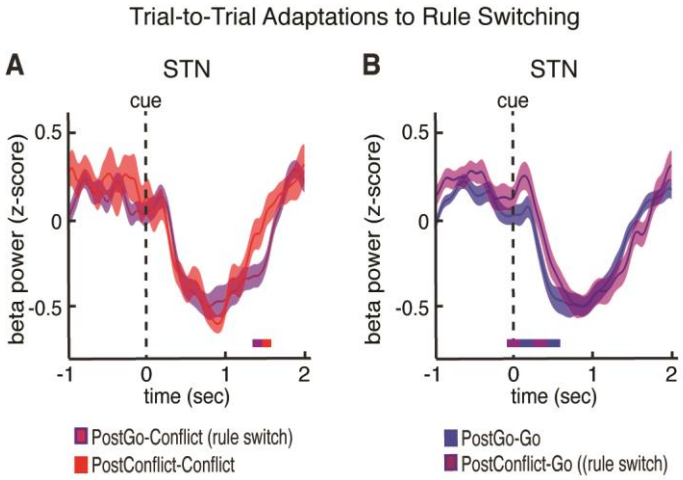
Supplementary Figure S10. Task related changes in STN phase-MUA coupling. (A) Left: MUA amplitude interactions with the phase of all analyzed frequencies. The average across all STN electrodes and all correct trials is shown. Data are aligned to arrow onset (t=0). Phase MUA coupling values for each recording were normalized to a surrogate distribution. Mask indicates time- frequency regions exhibiting significantly positive phase MUA coupling across all sessions when compared to zero ($p < 0.05$, corrected for multiple comparisons, permutation test). Significant phase-MUA coupling was seen in the beta and theta frequency bands. Middle: same data as (A, Left) but the phase-MUA coupling value recorded during a baseline period was subtracted from each time-frequency point. Mask indicates time-frequency regions exhibiting significant differences from baseline ($p < 0.05$, corrected for multiple comparisons, permutation test). Right: Same as (A, Middle), but averaged across the beta band (left) or theta band (right) and plotted separately for Go, Conflict, and NoGo trials. Time points exhibiting a significant difference between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test) are denoted by colored horizontal bars denoting which comparison was made. (B) Same data as (A), but quantified using the threshold detected spiking events rather than the MUA data. Overall, STN beta and theta oscillations significantly entrained STN spiking. Beta band entrainment decreased during the task, and theta band entrainment increased during the Conflict trials.



Supplementary Figure S11. STN multiunit activity during errors. Average continuous-time MUA firing rate for all STN microelectrode recordings that showed spiking activity plotted separately for the Go, Conflict, and error trials. Data are aligned to arrow onset (left) and to the motor response (right, t=0). Time points exhibiting a significant difference between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test) are denoted by colored horizontal bars denoting which comparison was made. The error trials showed an increase in STN MUA that was significantly higher and more prolonged than the correct Go and Conflict trials.



Supplementary Figure S12. Error related differences in mPFC gamma power. (A) Cue and response aligned time evolving gamma power changes averaged over all mPFC electrodes during the Go, Conflict, and error trials. Time points exhibiting a significant difference between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test) are denoted by colored horizontal bars denoting which comparison was made. Following an error, the mPFC showed a significant increase in gamma band power. During Conflict trials, an increase was also observed, and this was significantly different from the Go trials.



Supplementary Figure S13. Changes in oscillatory STN beta power related to across-trial rule switching. (A) Cue aligned time evolving beta power changes averaged over all STN electrodes during the correct Conflict trials that followed a correct Go trial (postGo-Conflict) and during the correct Conflict trials that followed a correct Conflict trial (postConflict-Conflict). Time points exhibiting a significant difference between trial types ($p < 0.05$, corrected for multiple comparisons, permutation test) are denoted by colored horizontal bars. The late differences are due to the differences in response time, resulting a difference in the relative timing of the response when aligned to the cue. (B) The analysis from Figure 6A is reproduced here for ease of comparison, however only sessions that were included in the PostConflict-Conflict analysis are included in this plot. Mean STN beta power is shown for the correct Go trials that followed a correct Go trial (postGo-Go) and during the correct Go trials that followed a correct Conflict trial (postConflict-Go). Only the sessions that had 5 or more postConflict-Conflict trials are plotted in panels (A) and (B). Overall, conflict trials that followed a Go trial did not show any changes in STN beta power early in this task, suggesting that the differences seen during the Go trials in Figure 6A were due to adaptations to Conflict rather than adaptations to a rule switch.