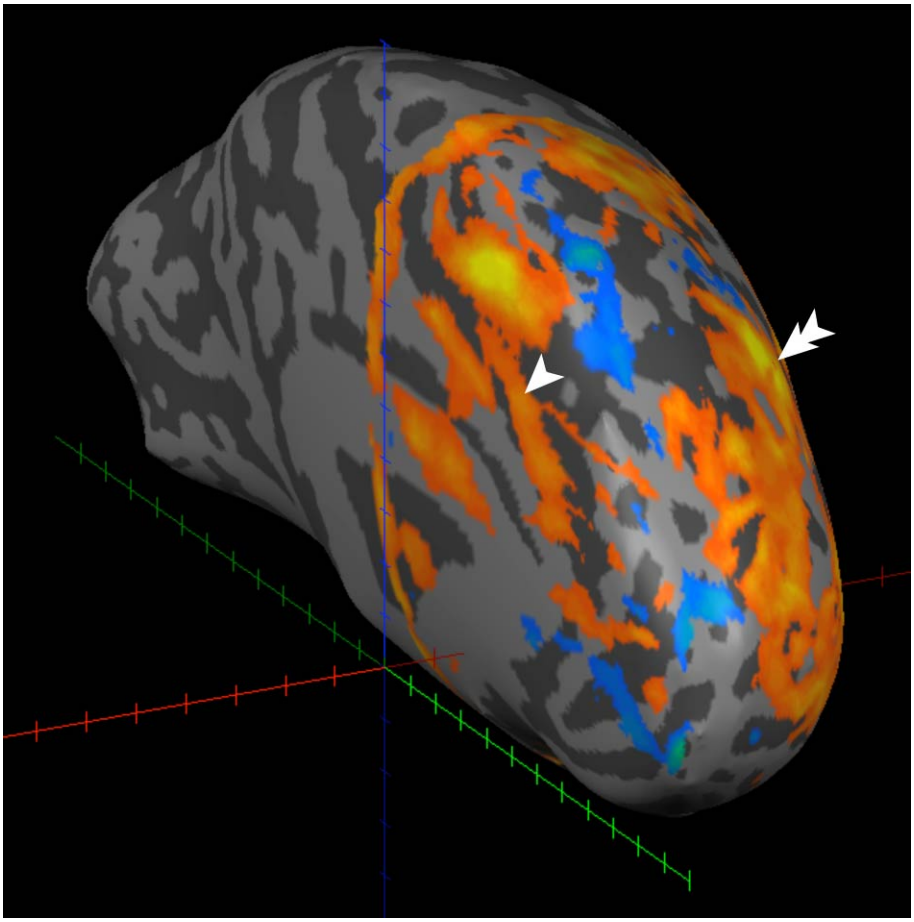
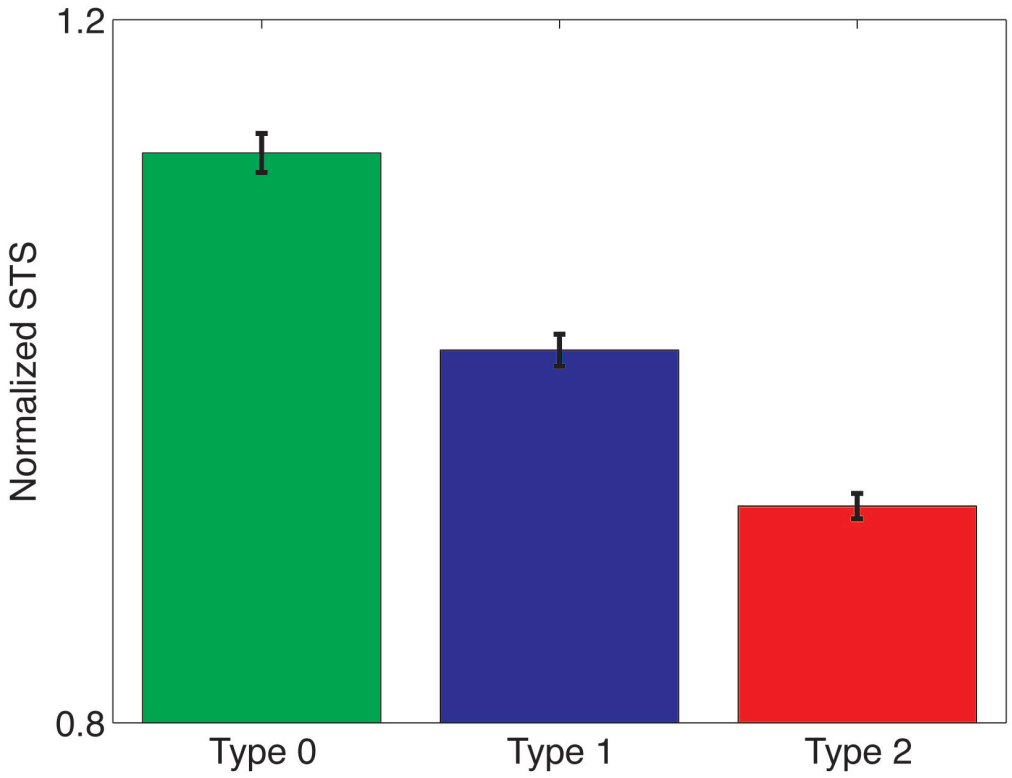
Supplementary Figures

Supplementary Figure S1



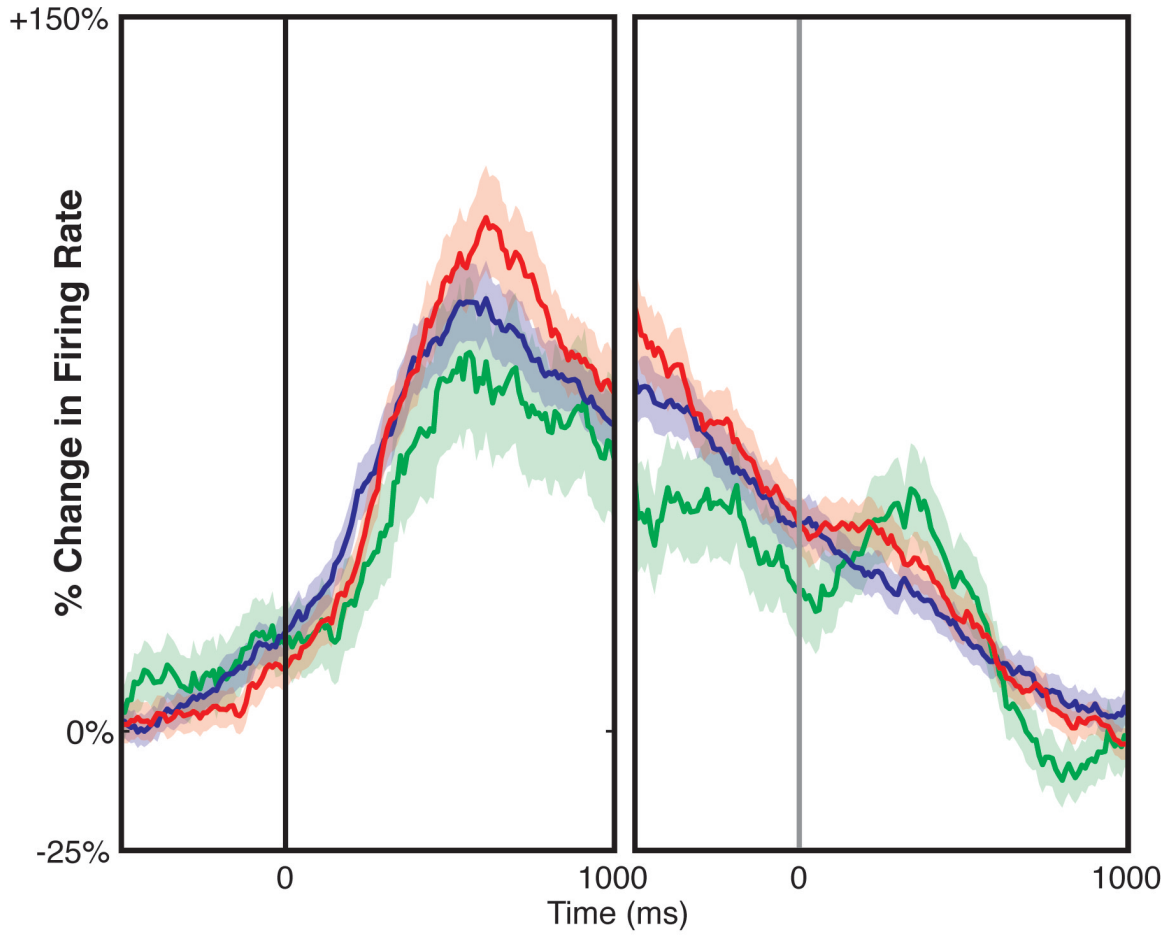
Supplementary Fig. 1. Functional MRI during the MSIT. Expanded view of the left hemisphere (gyri in light gray, sulci in dark gray) showing BOLD fMRI activation during the MSIT. A significant ($p < 0.05$ corrected for multiple comparisons) increase in signal is evident over the dACC (single arrowhead), as well as over the DLPFC (double arrowhead).

Supplementary Figure S2



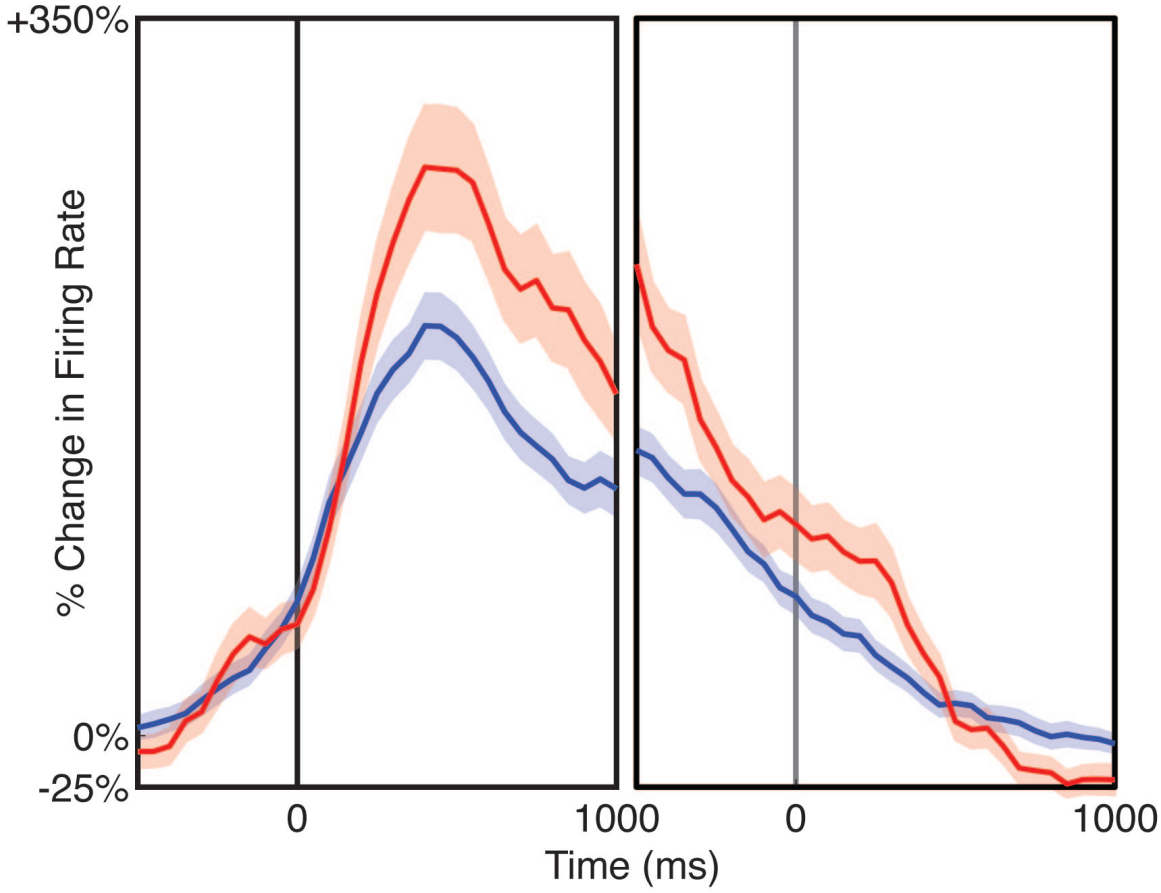
Supplementary Fig. 2. Reaction time data analyzed using the "speed of target selection" (STS) method of Mansouri et al. (2007)[24]. Selection speed decreased with increasing cue interference ($F(2,1546) = 107.7$, $p < 1x10^{-20}$, ANOVA).

Supplementary Figure S3



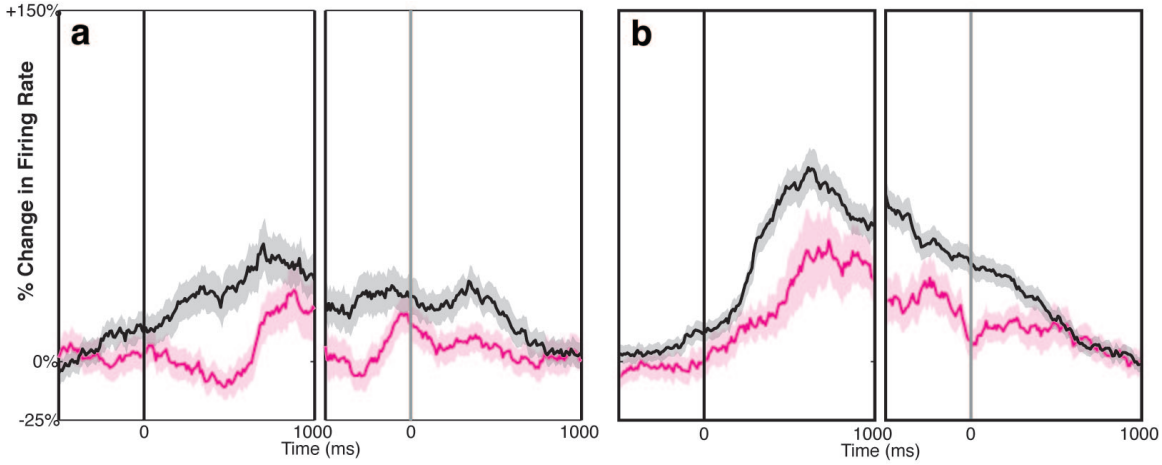
Supplementary Fig. 3. Average firing of all recorded neurons during Type 0 (green), Type 1 (blue), and Type 2 (red) trials, aligned to the cue (black vertical line) and to the choice (gray line). Dorsal ACC neurons increased firing with increasing cue interference. Shading indicates s.e.m. (n=3561).

Supplementary Figure S4



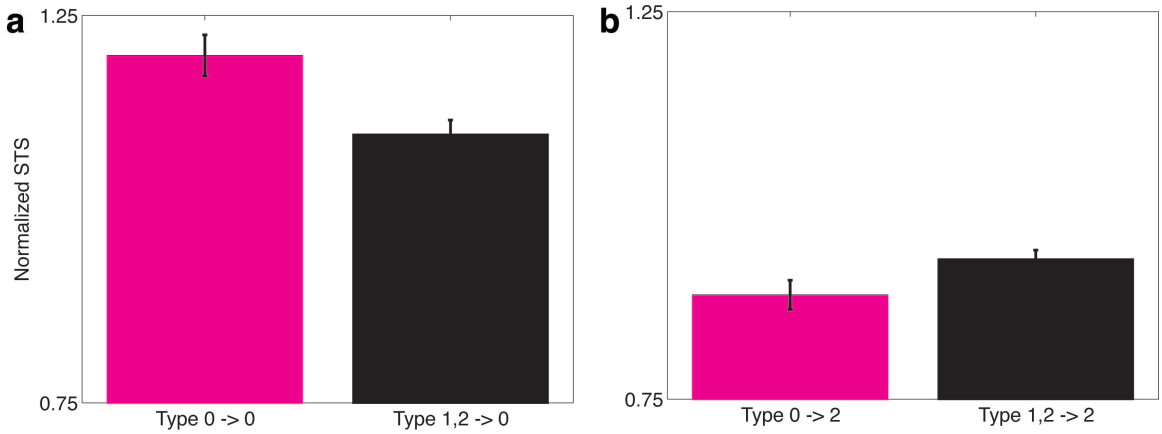
Supplementary Fig. 4. Dorsal ACC neurons encode conflict rather than number of potential responses. To determine whether dACC neuron firing encodes the amount of conflict within the cue or particular responses to the cue, we compared trials with the same number of possible responses, but differing amounts of conflict. Between trials with two possible responses, the firing rate was higher in the presence of two types of conflict (red) rather than one (blue). Average firing rates within a 500 – 1000 ms window were greater in the higher conflict trials (p = 6.4 x 10^{-3}, Mann-Whitney test). Shading indicates s.e.m. (n=1121).

Supplementary Figure S5



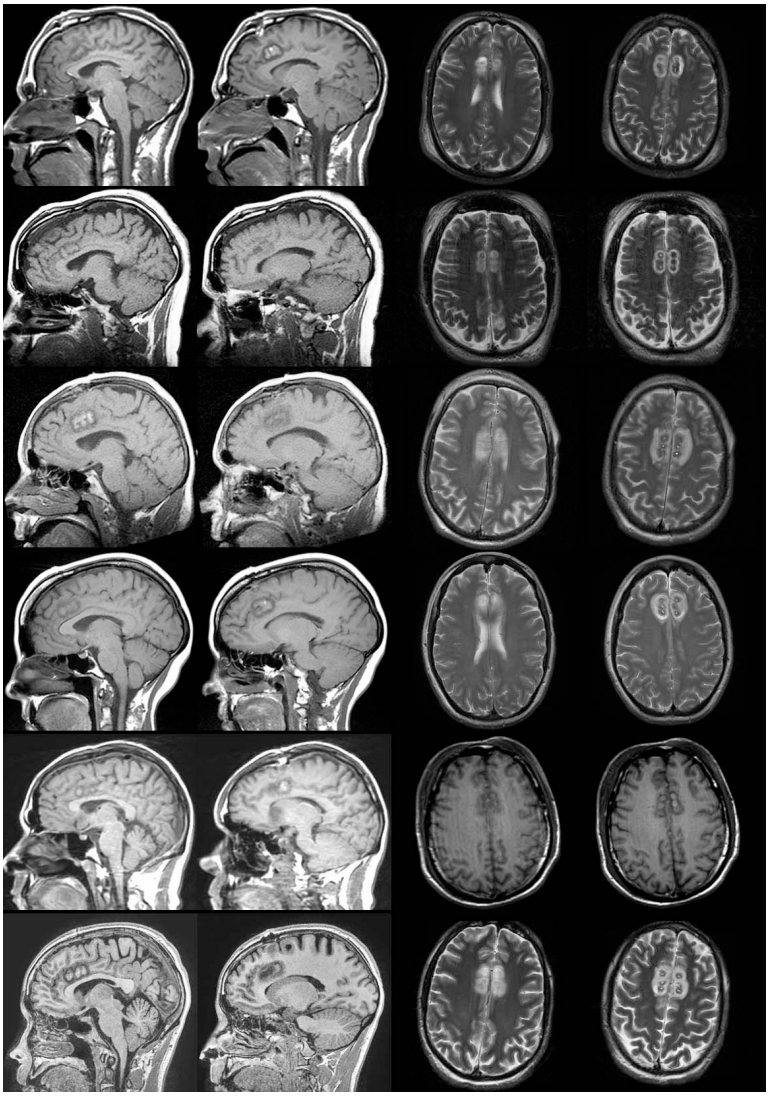
Supplementary Fig. 5. Effect of the previous trial on total dACC population firing. (**A**) Effect of preceding trial on non-interference (Type 0) trials. Current non-interference (Type 0) trials were segregated based on whether they were immediately preceded by an interference (Type 1 or 2) trial or another non-interference trial. When preceded by an interference trial (1,2→0; black line), cue-responsive activity increased more rapidly than when preceded by a non-interference trial (0→0; purple line). (**B**) Similarly for high interference (Type 2) trials, activity was higher when preceded by interference trials (1,2→2; black) than by non-interference trials (0→2; purple). Thus in both circumstances firing rates were higher when the previous trial contained interference. Shading indicates s.e.m. (n=2050).

Supplementary Figure S6



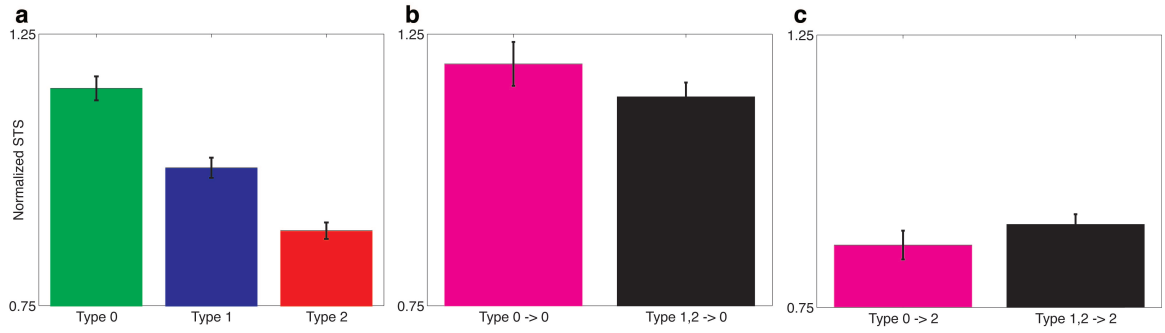
Supplementary Fig. 6. Trial-by-trial modulation of speed of target selection (STS). We analyzed trial-to-trial behavioral variations as in Figure 3, but using STS instead of RT. (**A**) Again, current non-interference (Type 0) trials were segregated based on whether they were immediately preceded by interference (1,2→0; black) or non-interference (0→0; purple) trials. STS was more rapid for current non-interference trials when preceded by non-interference trials (p = 1.7 x 10^{-3}, t-test). (**B**) STS was more rapid for current high interference (Type 2) trials when preceded by interference (1,2→2; black) than by non-interference (0→2; purple) trials (p = 0.04, t-test). Error bars represent s.e.m. (n=816).

Supplementary Figure S7



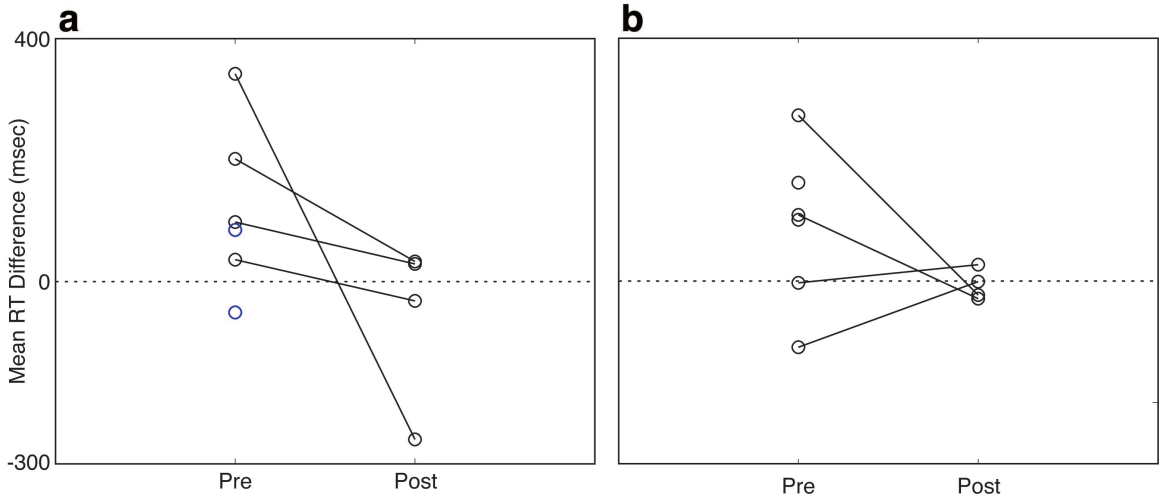
Supplementary Fig. 7. Lesion location across all six subjects. Immediate (within 24 hours) postoperative MRIs are shown for each subject, one row per subject. The first two panels in each row are sagittal T1-weighted images of the left hemisphere, from medial to lateral. The third and fourth panels in each row are T2-weighted axial images (except the fifth subject, in whom only T1-weighted axial images were available), from inferior to superior. The location and size of the bilateral triple lesions are consistent across the subjects.

Supplementary Figure S8



Supplementary Fig. 8. Abolition of behavioral adaptation following a targeted dACC lesion. As in Figure 4, we analyzed behavioral responses following cingulotomy. Rather than using RT, we used STS, as in Supplementary Figure 6. (**A**) Modulation of STS by trial type following cingulotomy. STS followed a dose-response pattern ($F(2,569) = 46.9$, $p < 1x10^{-20}$, ANOVA) similar to that before the lesion (Figure S2). Behavioral adaptations (the influences of previous trial identity on current trial reaction times, Figure 3, S6), however, were abolished for both **(B)** non-interference ($p = 0.21$, t-test) and **(C)** high-interference ($p = 0.26$, t-test) trials. Error bars represent s.e.m. (n=572).

Supplementary Figure S9



Supplementary Fig. 9. Individual RT results for the six subjects before and after cingulotomy. Each point represents the difference in raw RT (ΔRT) between different pairs of trial transitions. (**A**) Difference in raw RT between current Type 0 trials preceded by Type 1,2 trials and Type 0 trials (ΔRT = 2→0 minus 0→0). Five of the six subjects showed a conflict adaptation effect in the expected direction before cingulotomy (ΔRT > 0, implying that RT was longer in 2→0 trials than 0→0 trials). All four of the subjects with post-lesion data demonstrated reduction in the adaptation effect (decrease in ΔRT). (**B**) Difference in raw RT between current Type 2 trials preceded by Type2 trials and Type 0 trials (ΔRT = 0→2 minus 2→2). Four of the six subjects showed the expected conflict adaptation effect before cingulotomy (ΔRT > 0, implying RT was longer in 0→2 trials than 0→0 trials). The two of these who had post-lesion recordings showed abolishment of the adaptation effect. Of the two remaining subjects, one demonstrated no appreciable effect, and one had an opposite effect that was abolished after the lesion. Abbreviations: Pre, pre-lesion; Post, post-lesion.

Supplementary Methods

Surgical procedure

The surgical procedure was performed using standard stereotactic techniques. A Cosman-Roberts-Wells (Integra, Plainsboro, NJ) stereotactic frame was affixed to the patient under local anesthesia, and a high-resolution MRI obtained. The target for the left posterior lesion (2 cm posterior to the most anterior point of the frontal horn of the lateral ventricle, 0.7 cm lateral to midline, and 0.5 cm superior to the corpus callosum) was programmed into a neuro-navigation computer (StealthStation, Medtronic, Minneapolis, MN) and the stereotactic frame was then appropriately set. The patient was positioned semi-recumbent, the surgical area prepared, and sterile drapes applied. Local anesthetic was infiltrated, a coronal skin incision was performed, and bilateral burr holes drilled 1.5 cm lateral to midline, and 10.0 cm posterior to the nasion. A computerized microelectrode drive controlled by a neurophysiology system (Alpha Omega, Alpharetta, GA) was affixed to the frame. Following dural opening, microelectrodes were lowered using the computerized drive in increments of 0.01 mm. The position of the tip of the electrodes was also monitored in real time using the stereotactic neuronavigation system. Following microelectrode recordings, a thermocoagulation electrode with a 10 mm exposed tip (Cosman Medical, Burlington, MA) was lowered to the target. Lesions were performed by heating the electrode to 85 degrees Celsius for 60 seconds. Two more pairs of lesions were then created, each 7 mm anterior and 2 mm inferior to the previous lesion.

Functional magnetic resonance imaging (fMRI)

Functional MRI was performed using a 3.0 T scanner (Allegra; Siemens AG, Munich, Germany) and head coil. The task was presented on a screen visible via a tilted mirror, and controlled using MacStim 2.6 software (WhiteAnt Occasional Publishing, West Melbourne, Australia). Scans were acquired with the following specifications: 15 coronal sections; 64 x 64 matrix; 3.125 mm^2 in-plane resolution; 5 mm thickness with 0 mm skip; 30 ms echo time; 1500 ms repetition time; 90 degrees flip angle; 20 cm^2 field of view.

During fMRI, only Type 0 and 2 trials (no interference and both types of interference; see Figure 1B) were employed. The task was run in a block design. Each block consisted of 24 trials of the same type. One run consisted of 8 alternating blocks, with an additional five visual fixation blocks interspersed.

Post-lesion behavior

Immediately after creation of the cingulate lesions, subjects again performed the MSIT task. These task sessions were identical in all respects to the pre-lesion sessions except that we collected only behavioral data. Four of the six patients participated in these post-lesion sessions.

Reaction times

Reaction time (RT) was defined as the interval between the onset of the cue and the subject's button-press. To allow for inter-subject comparisons, we normalized subject RTs relative to their individual distributions. Normalized RTs were defined as the number of standard deviations (z-scores) above or below the bottom 10^{th} percentile of a subject's RT distribution. The choice of each subject's reference point for the normalization (bottom 10^{th} percentile vs. median or mean) is arbitrary and does not affect the result, since it simply represents a rigid translation of the normalized values. We chose the bottom 10^{th} percentile so that values were positive, to facilitate visual comparisons.

We also computed a speed of target selection (STS) as per Mansouri et al. (2007).[25] We inverted response time (defined as the interval between appearance of the cue on the screen and subject button-press) to obtain STS = 1/RT. We then divided by the mean STS across all trials in the behavioral session. In this way, STS was normalized to 1, which represented an intermediate or "average" response speed. Faster responses (*e.g.* on low-conflict trials) are reflected as relative increases in STS (*i.e.* STS > 1), whereas slower responses (*e.g.* on high-conflict trials) are attended by decrements in the STS (*i.e.* STS < 1). Use of the normalized STS, rather than raw RTs, allowed us to view data from different subjects on a common scale and to minimize between-subject differences when pooling data to calculate composite statistics.

Single-unit isolation

Single units were isolated by first building a histogram of peak heights from the raw voltage tracings on each channel. We applied a minimum threshold of three standard deviations to exclude background noise. Action potentials were sorted using waveform principal component analysis. Spike clusters of putative neurons were required to separate clearly from any channel noise, to demonstrate a voltage waveform consistent with that of a cortical neuron, and to have at least 99% of action potentials separated by an inter-spike interval of 1 ms or more.

We recorded an average of 1.1 neurons per microelectrode. When a single channel captured more than one neuron cluster, we required a clear distinction between the two clusters to include either one as a single unit ($p < 0.01$, multivariate ANOVA across the first two principal components). Additionally, we required putative dACC neurons to fire at an average rate of at least 1.0 spikes/sec, to be stably active for at least 25 task trials, and to not demonstrate significant drift over the duration of the recording. We excluded single units not meeting these criteria. We did not use any multi-unit activity.

Single-unit and population responses

We classified neurons into three mutually exclusive categories based on the timing of their peak firing rates with respect to task events. Neurons peaking in activity prior to stimulus presentation, during the stimulus period, and after the button-push were classified as “pre-cue”, “cue”, and “feedback” neurons, respectively.

To facilitate comparisons between neurons with different firing rates, we normalized (divided) neuronal activity by the average neuronal firing rate during a 500 ms window preceding appearance of the fixation point. Population firing rates were computed by averaging these normalized neuronal responses with a 250 ms moving boxcar window.

Supplementary Tables

Supplementary Table 1

	Cue Period Conflict Sensitivity[†]		
	n	**Type 2 > 0**	**Type 2 < 0**
Pre-Cue	12	4 (33%)	2 (17%)
Cue	24	9 (38%)	0
Feedback	23	4 (17%)	4 (17%)
Total	59	17 (29%)	6 (10%)

[†] An individual neuron was deemed "sensitive" to conflict if, in the cue period, in three or more consecutive 250 ms bins advanced in 50 ms increments, the firing rate for Type 2 vs. 0 trials differed at a significance level $p < 0.1$ in each bin.

Individual neurons in all three categories were sensitive to conflict; however, the population of cue neurons was enriched in conflict-sensitive neurons. All of these conflict-sensitive cue neurons fired faster for Type 2 than for Type 0 trials, consistent with the population response. The population response was therefore not an artifact of a few outliers. In the other two neuron populations, on the other hand, several neurons exhibited the opposite directional sensitivity. The cue neuron population was therefore unique in its consistency of conflict-sensitive neurons preferring high- over low-conflict trials.

Supplementary Discussion

Functional MRI results

Preoperative fMRI comparing high (Type 2) to no (Type 0) interference trials showed robust activations over the dACC (Figure 1D). In addition to dACC activation, fMRI also showed an increase in BOLD signal over the dorsolateral prefrontal cortex (DLPFC) (Figure S1), a region known to be involved during tasks engaging working memory and goal-oriented decision-making[30]. In particular, the DLPFC is prominently activated by the MSIT task in healthy volunteers[13-15]. This similarity in fMRI signal distribution suggests that our results can be generalized beyond the confines of our study population.

Population encoding of cognitive interference and trial history

Population-averaged activity of cue neurons correlated with the level of cognitive interference (Figure 2C, D). This dose-dependent encoding of cognitive load was not specific to cue neurons, however. In fact, the population-averaged activity of the entire pool of 59 neurons similarly demonstrated a trend of increased firing rates for high-interference trials during the cue period (Figure S3).

Similarly, the history-dependence of neural activity was not specific to cue neurons. Cue neurons demonstrated increased activity when the current trial had been preceded by an interference-laden trial (vs. a non-interference trial; Figure 3A, B). When responses of all

neurons—not only cue neurons—were pooled, a similar phenomenon was observed (Figure S5A, B). As with the cue neurons, this early activation was present for current trials either with (Figure S5B) or without (Figure S5A) interference.

Thus, the observed population encoding of cognitive interference and the history-dependence of neuronal activity were present both in the subset of cue neurons as well as across the entire pool of recorded dACC neurons.

Supplementary Note 1

Figure 2 shows that dACC neurons increased their firing rate with increasing cognitive interference. This pattern was true within individual neurons and the population average. One possible explanation for this observation is that individual dACC neurons encode the amount of *conflict* between simultaneously present but mutually incompatible response channels evoked by the cue. In this situation, Type 0 trials create the least conflict because of the lack of interference, Type 1 trials have either spatial or distracter interference and therefore create more conflict, and Type 2 trials have both types of interference and therefore create the most conflict.

An alternative explanation is that dACC neurons encode particular potential *responses* that are variably activated by the cue. In this scenario, a neuron may fire maximally when its response is cued in the absence of conflict, less when its response is cued in the presence of conflict or when another response is cued in the presence of conflict from its preferred response, and least when another response is cued in the absence of conflict. Trials with more conflict therefore would therefore generate increased firing because of the greater number of overall neurons activated.

To distinguish between these possibilities, we identified trials with the same number of potential responses but differing amounts of conflict. Type 0 trials all had 0 levels of conflict, and 1 possible response. Type 1 trials all had 1 level of conflict (either spatial or distracter), and 2 possible responses (the one correct response, and one competing

response resulting from the conflicting stimulus dimension). Type 2 trials all had 2 levels of conflict, but some had 2 possible responses (*e.g.*, 212 – button 1 as the correct response, and button 2 due to the position of the target number 1), and others had 3 possible responses (*e.g.*, 313 – the same two as in 212, with the addition of button 3 due to the identity of the distracters). Of the 12 Type 2 trials, half had 2 possible responses, and half had 3. We therefore compared Type 1 trials (1 level of conflict, 2 possible responses) to the half of Type 2 trials with 2 levels of conflict and 2 possible responses.

Supplementary Note 2

In addition to the neural-neural correlation described in the main text including all trial type transitions, we also calculated neural-neural correlations for individual trial type transitions. Below is a 3x3 table showing all trial transition types and the neural-neural correlation values for each combination of trials. Each table cell shows the correlation coefficient r (and p-value), with significant correlations denoted with an asterisk.

		Current trial		
		Type 0	**Type 1**	**Type 2**
Previous trial	**Type 0**	0.11 (0.265)	0.16 (0.128)	0.23 (0.002)*
	Type 1	0.15 (0.139)	0.15 (0.006)*	0.15 (0.008)*
	Type 2	0.21 (0.006)*	0.14 (0.019)*	0.14 (0.015)*

Six of the nine correlations are significant. The only non-significant correlations are those between the easiest trial types (0→0, 0→1, 1→0). These trials are the ones in which

firing rates are lowest, as they include the least conflict. The lack of significance in the neural-neural correlation for this transition type may be a consequence of the limited dynamic range of the neuronal activities in these trials, or may simply reflect the fact that there are fewer trials for analysis in these individual comparisons compared to when the data is pooled. The remaining comparisons all show a significant correlation between previous trial and current trial firing rate.

Supplementary Note 3

We inspected the raw analog waveform of each unit recording for evidence of drift and eliminated any neurons that demonstrated drift. To demonstrate that drift was not a significant contributor to our results, we analyzed data during the fixation period of each trial. If the neural-neural correlations that we observed during the cue period were simply a result of drift, then the drift would be evident throughout the trial. We therefore calculated the correlation between the firing rate of the previous and current trial for all trials during a 500 ms window in the fixation period (as we did in the main text in the cue period). There was no correlation during the fixation period ($r = 0.02$, $p = 0.18$). In contrast, the correlation between previous and current trial firing during the cue period was highly significant ($r = 0.12$, $p = 3.3 \times 10^{-16}$), as reported in the main text. Thus drift was not a significant contributor to the neural-neural correlations we observed during the cue period.

Supplementary Note 4

The correlation for Type 1 trials (0,1,2→1) was also significant (r = -0.05, p = 3.6e-2). Type 1 trials have one degree of conflict, whereas Type 2 trials have two. Thus the direction of the correlation (negative) of Type 1 trials is the same as that of Type 2 trials, but the magnitude is smaller.

We can also calculate these neural-behavioral correlations for all individual trial type transitions. Below is a 3x3 table showing all trial transition types and the neural-behavioral correlation values for each combination of trials. Each table cell shows the correlation coefficient r (and p-value), with significant correlations denoted with an asterisk.

		Current trial		
		Type 0	**Type 1**	**Type 2**
Previous trial	**Type 0**	0.13 (0.018)*	-0.07 (0.32)	-0.12 (0.008)*
	Type 1	0.21 (0.002)*	-0.04 (0.29)	-0.03 (0.44)
	Type 2	0.06 (0.17)	-0.06 (0.11)	-0.18 (0.009)*

Again, the correlation coefficients follow the expected pattern in terms of direction of correlation. For current Type 0 (non-interference; 1st column of table) trials, the correlation coefficient was positive, meaning that increased neuronal activity (implying increased conflict) on the previous trial correlated with longer reaction times on the current non-interference trial. Conversely, for current Type 1or 2 (interference trials; 2nd and 3rd columns of table) trials, the correlation coefficient was negative, meaning that

increased neuronal activity correlated with shorter reaction times on the current interference trial. Thus although not all correlations met statistical significance (likely for the same reasons as described in Supplementary Note 2), the signs of all 9 correlations followed the expected pattern. It would be highly unlikely for the correlations to assume the expected direction by chance alone.