Title: Thalamocortical contributions to hierarchical cognitive control

Authors: *Xitong Chen[1,2,3], Stephanie C. Leach[1,2,3], Juniper Hollis[1,2,3], Dillan Cellier[1,2,3], Kai Hwang[1,2,3]

[1]Department of Psychological and Brain Sciences, [2]Cognitive Control Collaborative, [3]Iowa Neuroscience Institute, The University of Iowa

***Corresponding Author:**
Xitong Chen
xitong-chen@uiowa.edu
Department of Psychological and Brain Sciences
The University of Iowa
G60 PBSB
340 Iowa Ave
Iowa City, IA 52242-1407

Number of pages: 36
Number of figures: 8
Number of words: 10471

Keywords: Thalamus, thalamocortical interaction, functional connectivity, control representations, cognitive control

Acknowledgments: Research reported in this publication was supported by the Iowa Neuroscience Institute and the National Institute of Mental Health Award Number R01MH122613. This work was conducted on an MRI instrument funded by 1S10OD025025-01.

Abstract

Cognitive flexibility relies on hierarchically structured task representations that organize task contexts, relevant environmental features, and subordinate decisions. Despite ongoing interest in the human thalamus, its role in cognitive control has been understudied. This study explored thalamic representation and thalamocortical interactions that contribute to hierarchical cognitive control in humans. We found that several thalamic nuclei, including the anterior, mediodorsal, ventrolateral, and pulvinar nuclei, exhibited stronger evoked responses when subjects switch between task contexts. Decoding analysis revealed that thalamic activity preferentially encodes task contexts within the hierarchical task representations. To determine how thalamocortical interactions contribute to task representations, we developed a thalamocortical functional interaction model to predict task-related cortical representation. This data-driven model outperformed comparison models, particularly in predicting activity patterns in cortical regions that encode context representations. Collectively, our findings highlight the significant contribution of thalamic activity and thalamocortical interactions for contextually guided hierarchical cognitive control.

Introduction

Cognitive control enables humans to adjust their behaviors flexibly in response to changing circumstances. For example, the sound of a ringing phone can elicit different reactions depending on the context. At home, one might answer, while driving might cause one to ignore it for safety. Everyday life is filled with scenarios that require us to modify our actions based on different contexts. How does our control system facilitate such contextually guided control?

One influential framework posits that cognitive control is supported by hierarchically organized task representations that are encoded by the functional organization of the prefrontal cortex (Badre, 2008; Badre & D'Esposito, 2007; Badre & Nee, 2018). In this framework, task representations are structured to encode contexts, goals, task-relevant features, and potential actions (Schumacher & Hazeltine, 2016). Furthermore, representations can be organized in a hierarchical manner, where more abstract contextual representations exert stronger influence over action representations to support cognitive flexibility (Cellier et al., 2022; Voytek et al., 2015). Anatomically, it was proposed that this abstraction gradient is supported by a rostro to caudal organization in the lateral prefrontal cortex (Badre & D'Esposito, 2007; Christoff et al., 2009; Koechlin et al., 2003; O'Reilly, 2010; O'Reilly et al., 2002).

To support flexibility, task representations are updated as context changes. One computation model suggests that the cortico-striatal-thalamic circuit is involved in updating cortical task representations (Badre & Nee, 2018; Chatham & Badre, 2015; O'Reilly & Frank, 2006). According to this model, the basal ganglia receives control signals from the prefrontal cortex, and then uses these signals to disinhibit the thalamus to influence cortical representations via thalamocortical inputs (Badre & Nee, 2018; Chatham & Badre, 2015). Inspired by this model, several neuroimaging studies focused on characterizing the interaction between prefrontal cortex and the striatum (Badre & Frank, 2012; Chatham et al., 2014; Nee & Brown, 2013).

Notably, the human thalamus is a critical but understudied component of this circuitry. Structurally, the basal ganglia does not directly project to the cortex, therefore its influence on cortical representations must be mediated by the thalamus (Alexander et al., 1986; Alexander & Crutcher, 1990; Haber & Mcfarland, 2001). Despite this anatomical relationship, many theoretical models omit it from the model. Furthermore, most existing models are primarily built on the known projections from the pallidum to the ventrolateral thalamus for allowing motoric representations to be influenced by higher-order control processes (DeLong, 1990; Mink, 1996; O'Reilly & Frank, 2006; Parent & Hazrati, 1995; Sidibé et al., 1997). It remains unclear whether higher order thalamic nuclei, including the anterior, mediodorsal, and pulvinar nuclei, which have dense connectivity with frontal and parietal regions (Goldman-Rakic & Porrino, 1985; Shibata & Naito, 2005; Yeterian & Pandya, 1985), are involved in hierarchical cognitive control in humans. This is because most past neuroimaging studies did not provide specific details on the thalamic functional anatomy. Furthermore, recent findings from animal models demonstrated that thalamocortical interactions can signal context changes (Hummos et al., 2022; Rikhye et al., 2018; Schmitt et al., 2017). In humans, the thalamus is ideally positioned to influence cortical representations via its extensive one-to-many and many-to-one thalamocortical connectivity motifs (Shine et al., 2023). Specifically, the anterior, medial, and posterior thalamus exhibit

strong converging connectivity with multiple frontal-parietal systems that have been implicated in control-related functions (Hwang et al., 2017). The behavioral significance of this thalamocortical connectivity profile is further affirmed by recent functional neuroimaging studies (Chen et al., 2023; Hwang et al., 2022) and lesion evidence (Hwang et al., 2021, 2022). However, it remains unclear how thalamocortical connectivity supports hierarchical cognitive control.

The current study focused on elucidating the role of the human thalamus in hierarchical cognitive control. We designed a paradigm with different levels of hierarchical task representations. Instead of using different tasks to establish hierarchy, we equated stimuli and rule complexity across different levels of hierarchical representations within one behavioral task design (Cellier et al., 2022). We had three main objectives. First, we aimed to characterize the detailed functional anatomy of thalamic activity that supports hierarchical cognitive control. Second, we aimed to determine the interaction between thalamic activity and cortical control representations. To accomplish this, we utilized a novel, data-driven, thalamocortical activity flow mapping analysis to test if thalamocortical interactions can predict cortical activity patterns related to cognitive control (Chen et al., 2023; Cole et al., 2016, 2021; Hwang et al., 2021). Third, we decoded the cognitive representations encoded by thalamic activity to determine the hierarchical task representation thalamocortical interactions most strongly contribute to.

Results

Hierarchical control task and behavioral results

We designed a behavioral task to investigate hierarchical cognitive control (Figure 1a). In each trial, a task cue was presented for 0.5 second, followed by a picture probe for 2.5 seconds. Subjects had to respond to the probe by applying one of two specific response rules: the "scene rule" (determining whether the presented picture was a scene) or the "face rule" (determining whether the presented picture was a human face). The appropriate rule selection was based on different features of the cue, organized hierarchically. Specifically, the texture of the cue (solid vs. hollow) referred to as the “context”, representing the highest level of attributes. This context information determined which mid-level “feature” subjects should select to determine the appropriate “response rule”. Each feature from the mid-level (shape or color) was then associated with either the face or scene rule. The face and scene rule then determines the visual stimuli category subjects should decide to attend to generate the correct motor response. This organization introduced three hierarchical levels of task switching: extra-dimension switch (EDS), inter-dimension switch (IDS), and Stay (Figure 1b). In EDS trials, subjects updated the highest-level task context. In IDS trials, subjects switched between task-relevant features within the same task context. In Stay trials, subjects performed the same task for two trials in a row. We hypothesized that as the hierarchical task switching demand increases from Stay to IDS to EDS, response time (RT) would also increase.

Results from the one-way (Conditions: EDS, IDS and Stay) repeated measures ANOVA (rmANOVA) showed that the hypothesized hierarchical structure of task representations significantly modulated subject’s RT ($F(2,116)=156.21$, $p<0.0001$, $\eta^2=0.11$, Figure 1c). Subjects responded slowest for the EDS condition (mean ± SD: 1.16s ± 0.22; EDS vs. IDS: $t(58)=8.36$, $p<0.001$, Cohen’s d=0.44; EDS vs. Stay: $t(58)=16.70$, $p<0.001$, Cohen’s d=0.83), followed by the IDS (mean ± SD: 1.07s ± 0.20; IDS vs. Stay: $t(58)=10.60$, $p<0.001$, Cohen’s d=0.41) and Stay (mean ± SD: 0.99s ± 0.19) conditions. Similar effects were observed for accuracy ($F(2,116)=13.70$, $p<0.0001$, $\eta^2=0.05$; Figure 1d). Accuracy for the EDS condition (mean ± SD: 0.94 ± 0.05) was significantly lower than the Stay condition (mean ± SD: 0.97 ± 0.03; $t(58)=-6.10$, $p<0.001$, Cohen’s d=-0.51), but not significantly different to the IDS condition (mean ± SD: 0.95 ± 0.04; $t(58)=-1.88$, $p=0.19$, Cohen’s d=-0.20). The accuracy for the IDS condition was significantly lower than Stay ($t(58)=-3.12$, $p=0.008$; Cohen’s d=-0.36). This observed pattern of reaction time and accuracy results suggest a hierarchical organization of task representations, consistent with findings from our previous published study that employed the same task with EEG (Cellier et al., 2022).

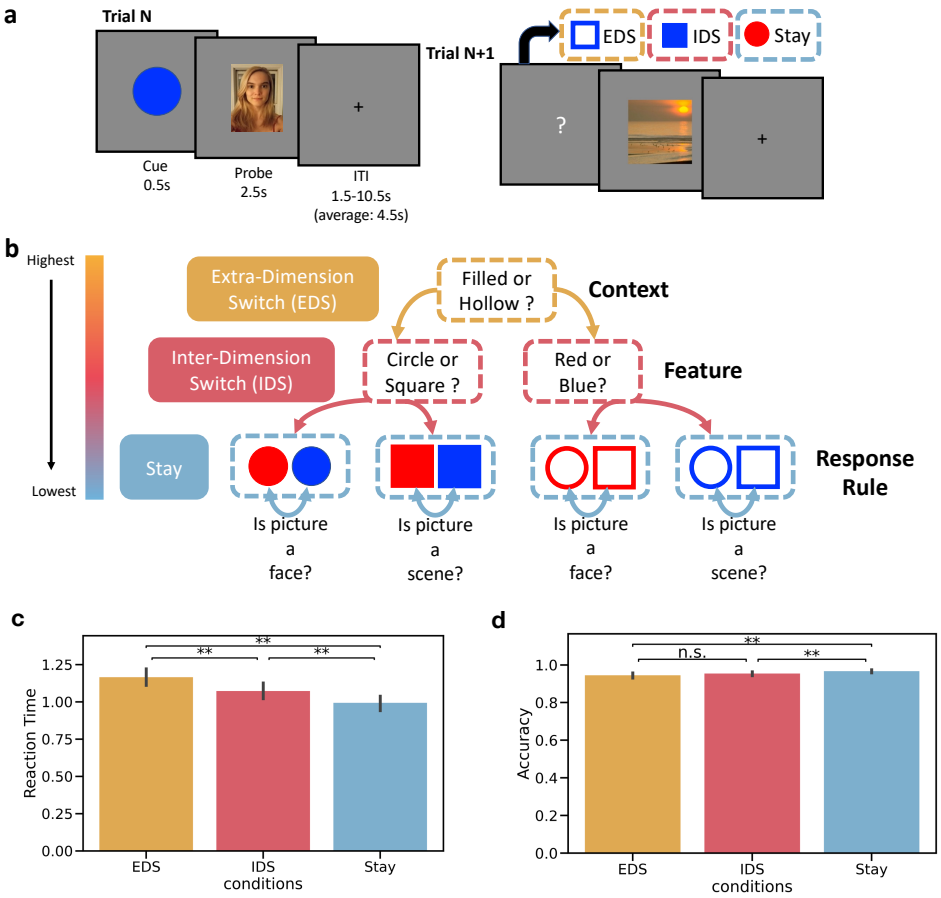

Figure 1. Hierarchical Control Task. *(a) The hierarchical control task, in which we introduced different levels of hierarchical task switching between trials. (b) The hierarchy of task representations. (c) Reaction time observed across the three hierarchical levels of task-switching. (d) Accuracy across the three hierarchical levels of task-switching. ** p<0.01. n.s. non-significant.*

Cortical evoked responses

We found that the behavioral performance was influenced by the demand of hierarchical task switching. Subsequently, we explored how hierarchical task switching modulated cortical and subcortical activity. We first investigated cortical evoked responses elicited by different hierarchical task switching conditions. This analysis identified a set of cortical regions including the middle frontal gyrus, the inferior frontal sulcus, the precentral sulcus, the intraparietal sulcus, the postcentral sulcus, the insula, the medial frontal gyrus, the posterior cingulate gyrus, the lateral occipital gyrus, the fusiform gyrus, the posterior cingulate cortex, and the inferior temporal gyrus (EDS, IDS and Stay conditions; Figure 2a). We then contrasted the magnitudes of evoked responses between different task switching conditions (e.g., EDS-Stay, EDS-IDS and IDS-Stay; Figure 2b). Specifically, we examined EDS versus Stay and EDS versus IDS to determine cortical regions involved in switching of task context, and IDS versus Stay for switching of task feature. We found stronger evoked responses in the frontoparietal and the temporal regions in response to changing task contexts (EDS-Stay and EDS-IDS, Figure 2b), included areas such as the rostral middle frontal gyrus, the inferior frontal sulcus, the insula, the precentral sulcus, the postcentral sulcus, the intraparietal sulcus, the medial frontal gyrus, the posterior cingulate gyrus, the precuneus, the cuneus, and the middle temporal lobe. We did not

observe significant effects in these regions during switches in feature level task information (IDS-Stay). Instead, when contrasting task switching that involved feature level task information (IDS-Stay), we found significant clusters primarily distributed in the premotor cortex and the intraparietal sulcus. These cortical regions also showed significant modulations in response to updating context information (EDS-IDS). This pattern revealed an asymmetrical hierarchy consistent with previous studies (Badre & D'Esposito, 2007), in which several regions in the caudal frontal cortex are more involved in contextual update, whereas premotor regions are responsive to both contextual and feature level updates of task representations.

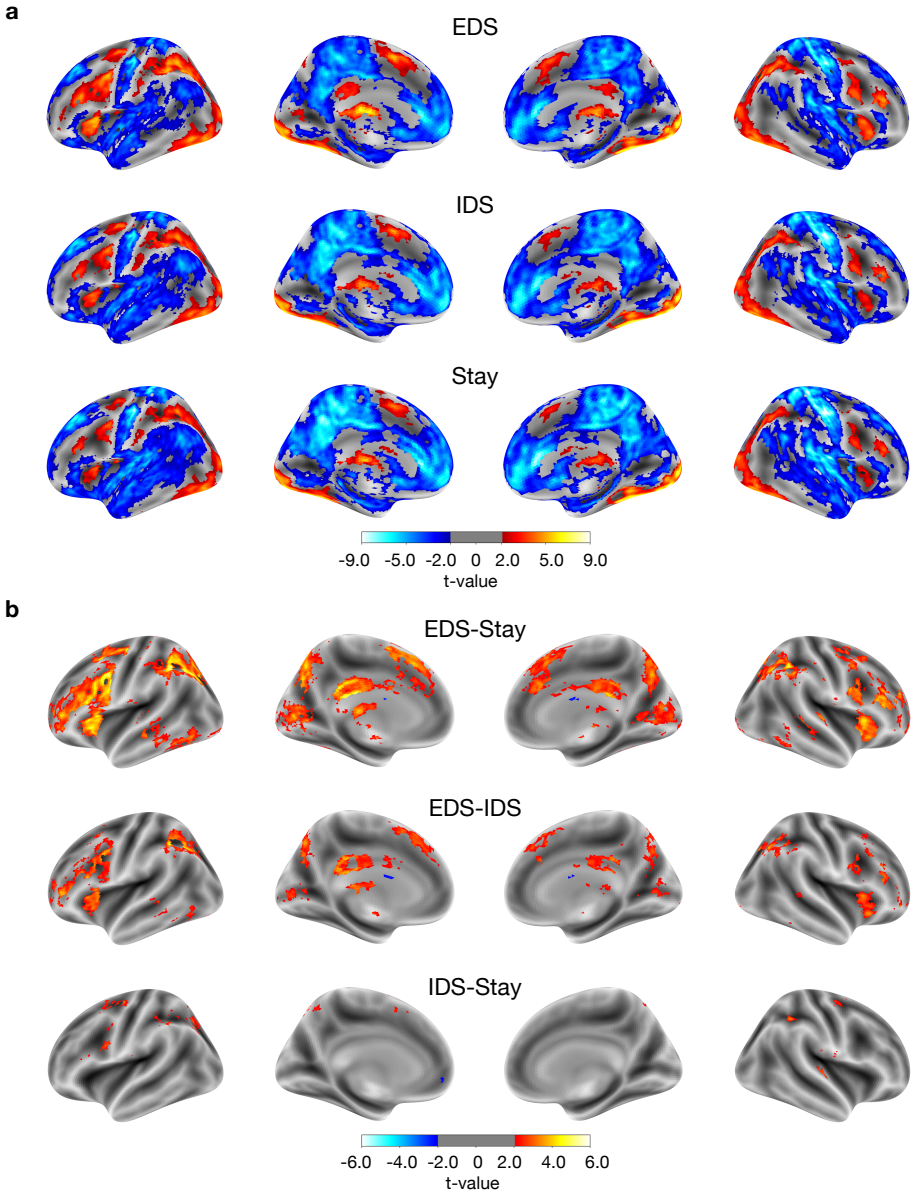

Figure 2. Cortical evoked response to hierarchical task-switching *(a) Cortical evoked responses for the three task-switching conditions. (b) Contrast between conditions. Clustered corrected at p<0.05.*

Thalamic activity profile responding to hierarchical task-switching

We then investigated the detailed functional anatomy of thalamic activity in response to different levels of hierarchical task switching (EDS, IDS and Stay conditions; unthresholded activity maps: Figure 3a, thresholded activity maps at cluster corrected $p < .05$: Figure 3b). Significant evoked responses were observed in the anterior, the ventral, the medial and the posterior regions of the thalamus for all task switching conditions (Figure 3a&b). The negative response patterns observed in the medial and the posterior region of the thalamus may be attributed to a significant undershoot in the hemodynamic response functions (HRFs) of these nuclei (see Supplementary Materials S1).

Our results revealed activity patterns within the thalamus for updating different hierarchical task representations. Thalamic voxels in the anterior, the ventroanterior, the ventromedial, the ventrolateral, the mediodorsal and the intralaminar nuclei showed stronger evoked responses more selective for updating context representations (EDS-IDS; Figure 3c). In addition, we observed that voxels in the anterior, the ventroanterior, the ventrolateral, the mediodorsal, the intralaminar, the medial pulvinar, the lateral pulvinar and the ventral posterior nuclei of the thalamus showed stronger evoked responses for the EDS-Stay contrast (Figure 3c). We did not observe any significant clusters for the IDS-Stay contrast, but only for the EDS-IDS and EDS-Stay contrasts. This suggests that the thalamus may be more selectively involved in hierarchical cognitive control that involves updating context representations.

In summary, the hierarchical cognitive control modulates the evoked response of thalamus. Additionally, thalamic nuclei in anterior, ventral and medial regions displayed a preference of updating context-level task representation in hierarchical cognitive control task.

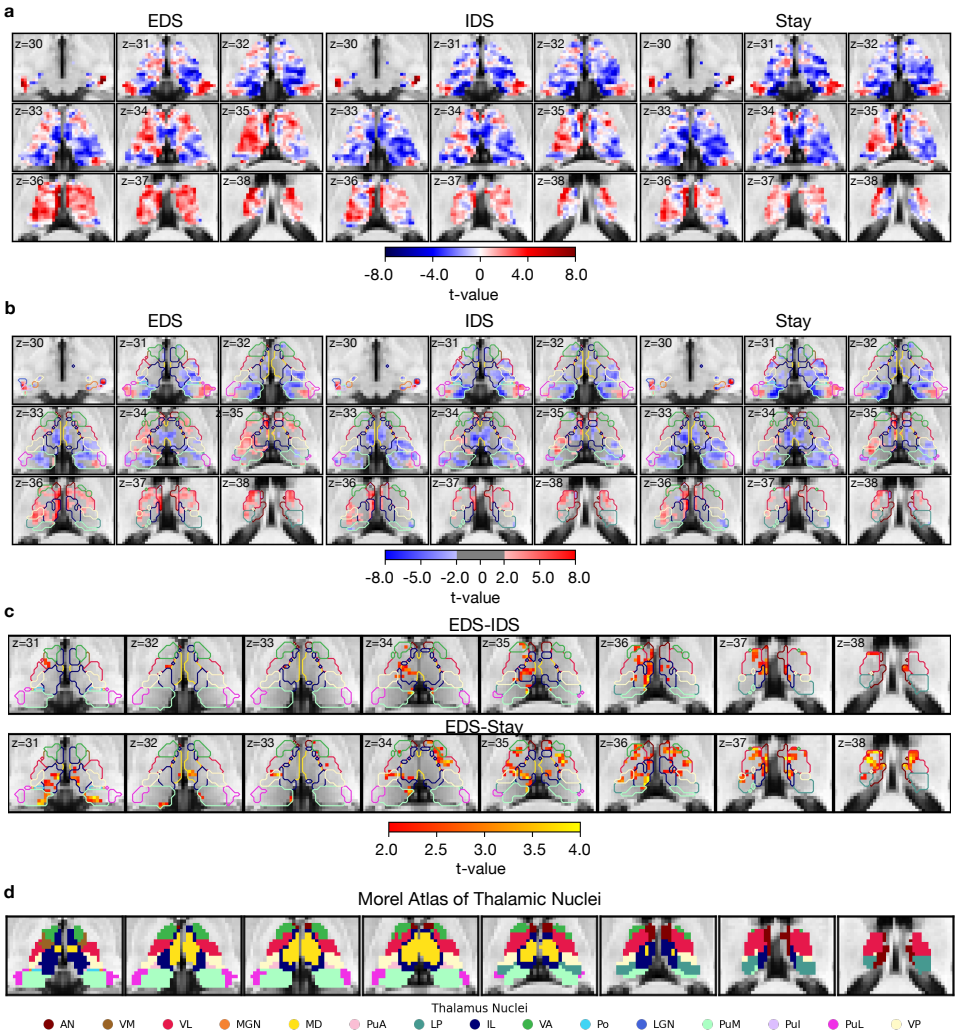

Figure 3. Thalamic voxel-wise evoked responses in response to hierarchical task-switching *(a)Unthresholded thalamic evoked response patterns (t-value) for the three hierarchical task-switching conditions. (b)Thresholded thalamic evoked response patterns. (c) Contrast between conditions, cluster corrected at p<0.05. The color of contour in (b) and (c) refers to the nuclei described in (d). (d) Morel atlas of thalamic nuclei. Abbreviations for thalamic nuclei are as follows: AN, anterior nucleus; VM, ventral medial nucleus; VL, ventral lateral nucleus; MGN, medial geniculate nucleus; MD, mediodorsal nucleus; PuA, anterior pulvinar nucleus; LP, lateral posterior nucleus; IL, intralaminar nucleus; VA, ventral anterior nucleus; Po, posterior nucleus; LGN, lateral geniculate nucleus; PuM, medial pulvinar nucleus; PuL, lateral pulvinar nucleus; VP, ventral posterior nucleus.*

Thalamic voxels selectively encode contextual information

Our results showed that thalamic activity is most strongly associated with updating contextual information during hierarchical cognitive control. We next seek to determine which level of hierarchical task representation is encoded in the thalamus. Utilizing multi-voxel pattern analysis, we investigated whether we could decode context, color, shape, or response rule from multi-voxel thalamic activity patterns. We observed above chance decoding of contextual information throughout the thalamus (solid vs. hollow; Figure 4a). Highest decoding accuracy was found in the anterior (0.52±0.004), the ventroanterior (0.52±0.004), the ventrolateral (0.52±0.004), the mediodorsal (0.52±0.004), the intralaminar (0.52±0.004), the medial pulvinar (0.52±0.004), and the lateral pulvinar nuclei (0.52±0.004; Figure 4b). These thalamic nuclei also

found to show strongest evoked response during the contextual update condition (EDS). Decoding performance for other information, color (blue vs. red), shape (square vs. circle) and task (face vs. place tasks), was significantly weaker, with few significant voxels in anterior and medial regions (Figure 4b). But these voxels did not survive after cluster-correction. These results suggest that thalamic activity is involved in encoding contextual information during hierarchical cognitive control.

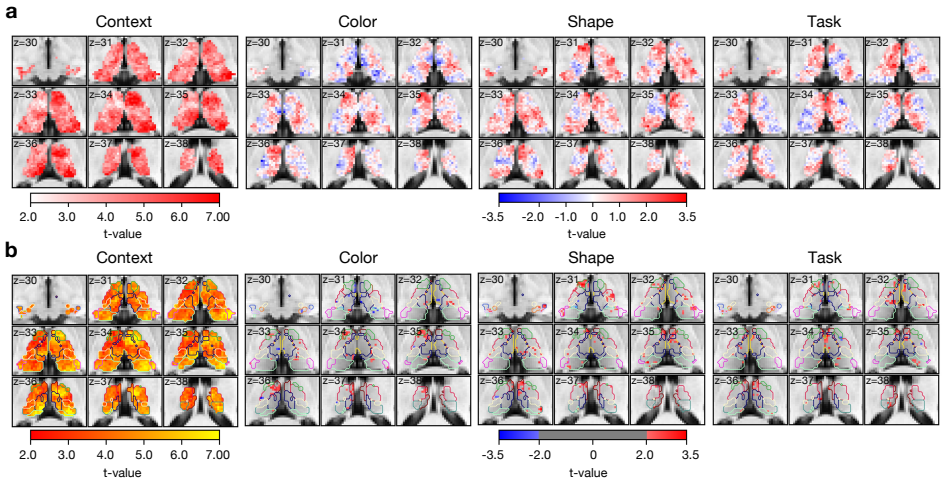

Figure 4. Thalamic activity encodes task contexts *(a) Unthresholded decoding performance (t-value), compared against chance (50%) (b) Thresholded decoding performance (t-value) at p<0.05. Context was cluster-corrected at p<0.05. Color, Shape and Task were uncorrected.*

Thalamocortical interaction model predicts cortical activity patterns

Given the strong reciprocal and non-reciprocal connections between the thalamus and the cortex, significant thalamocortical interactions may take place during hierarchical cognitive control. Prior studies using computational modeling (Hummos et al., 2022) and animal models (Rikhye et al., 2018; Schmitt et al., 2017) have suggested that thalamocortical interactions instantiate cortical task representations. Therefore, we further examined how thalamocortical interactions in humans contribute hierarchical task representations to support cognitive flexibility. To this end, we modified a previously developed activity flow mapping procedure to determine the relationship between thalamic evoked responses, thalamocortical functional connectivity, and cortical representations (Cole et al., 2016, 2021; Ito et al., 2017). In brief, this method tests whether a data-driven thalamocortical model can predict cortical activity patterns by weighting evoked response patterns of thalamic voxels based on their functional connectivity (FC) with cortical ROIs (see Methods for details). This approach allows us to assess the extent to which thalamocortical interactions influence cortical task activity patterns associated with hierarchical task representations.

We first tested whether this thalamocortical model can predict whole-brain activity patterns associated with different hierarchical task-switching conditions (Figure 5a). Briefly, we calculated the dot product of voxel-wise evoked response patterns within the Morel's thalamus mask (Krauth et al., 2010) and their FC matrix between thalamic voxels and 400 ROIs defined

by the Schaefer's cortical parcellation (Schaefer et al., 2018). We employed a split-half cross-validation procedure to generate predicted activity patterns for the 400 ROIs (see Methods for details). To evaluate prediction accuracy, we compared these predicted patterns to observed cortical activity patterns, using Pearson correlation normalized by noise ceiling (see Methods for details). We then compared the performance of our model against two null models that assumed neither the spatial pattern of thalamic evoked response nor thalamocortical functional connectivity pattern carried task-related information. The results demonstrated that our model significantly outperformed the two null models across all conditions (Figure 5b&c). However, there was no statistically significant difference between the three conditions ($F(2,116)=1.43$, $p=0.24$; EDS (mean ± SD): 0.50±0.25, IDS (mean ± SD): 0.46±0.20, Stay (mean ± SD): 0.49±0.18).

The above analysis included all thalamic voxels. To investigate the specific contributions of different thalamic subregions to model prediction, we simulated thalamic lesions by systematically removing 20% of thalamic voxels, which we referred to as "virtual lesioning voxels," from the activity flow mapping analysis. These voxels were selected based on their ranking by the amplitudes of evoked responses, ranging from the 1st to the 80th percentile. Our findings consistently showed that the removal of voxels with the strongest evoked responses resulted in a notable decrease in prediction accuracy across all three hierarchical task-switching conditions (Figure 5d). After averaging the reduction magnitude across thalamic nuclei, we found that thalamic voxels contributing most significantly to prediction accuracy were primarily in the anterior, the ventrolateral, mediodorsal, the intralaminar, the ventral posterior, and the medial pulvinar nuclei (Figure 5e&f). Notably, these thalamic voxels overlapped with those showing strong evoked response modulations (EDS-Stay and EDS-IDS; Figure 3c). In summary, the above results demonstrated the utility of thalamocortical activity flow prediction. Having demonstrated the utility of this approach, next we built thalamocortical models to predict voxel-wise activity patterns within specific cortical ROIs that encode different task representations.

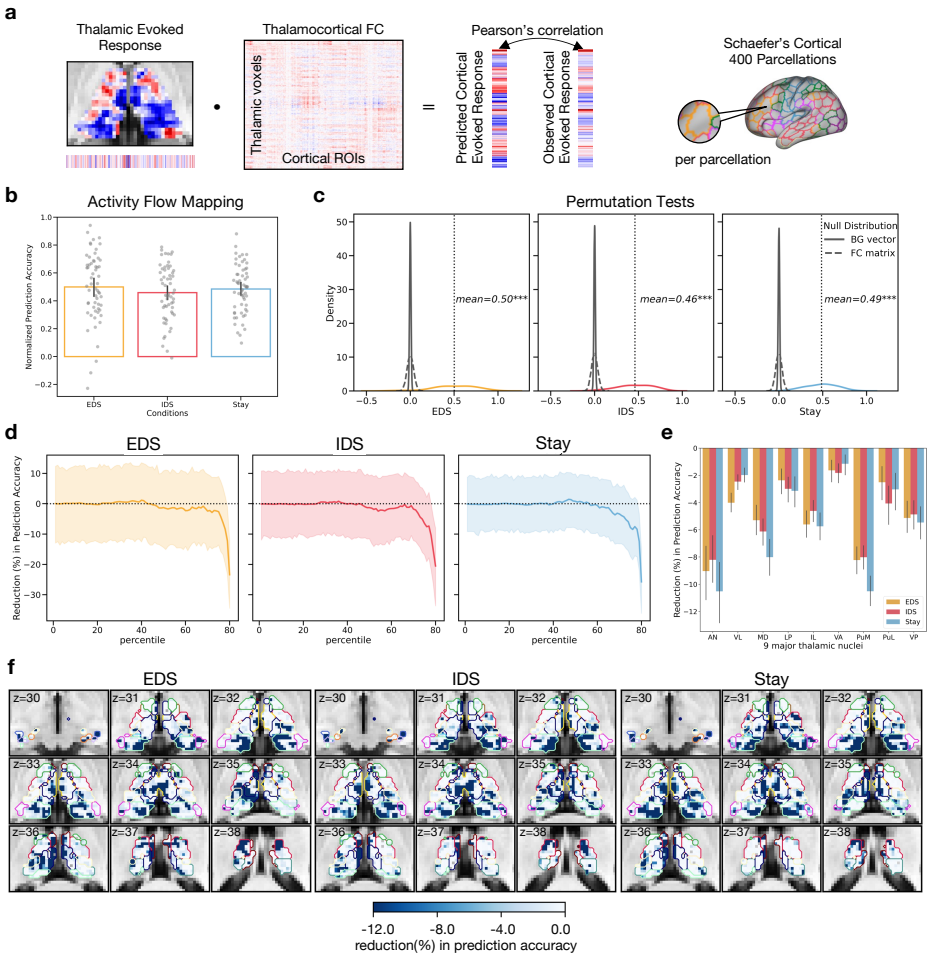

Figure 5. Thalamocortical activity flow mapping *(a) Diagram of activity flow mapping analysis, adapted from (Cole et al., 2016, 2021). The thalamus and brain parcellation masks that were used to extract the evoked response and time series data for all activity flow mapping analysis in current study. Schaefer's 400 cortical parcellations (Schaefer et al., 2018) were used to calculate voxel-to-parcellation thalamocortical activity flow. (b) The prediction accuracy of voxel-to-ROI thalamocortical activity flow for the hierarchical task-switching conditions. The error bar represents the 95% confidential interval. (c) Null models. (d) Reduction in prediction accuracy for the three hierarchical task-switching conditions. (e) Averaged reduction magnitude in prediction accuracy in 9 major thalamic nuclei. (f) Thalamic topography of reduction in prediction accuracy.*

Thalamocortical interactions update cortical control representations

Our results from the thalamocortical activity flow analysis so far have demonstrated that thalamocortical interactions can predict cortical activity patterns. However, considering the hierarchical organization of task representations, it remains unclear which specific level of representations is influenced by thalamocortical interactions. To address this question, we first conducted a whole-brain representational similarity analysis (RSA) to identify the specific representation encoded by each cortical region. Subsequently, we examined the pattern of thalamocortical interactions with each cortical region using the previously mentioned thalamocortical activity flow model.

We first performed a ROI RSA (Figure 6a&b) to identify cortical ROIs that encode different constituent representations within the hierarchical task (Figure 6a). In each cortical ROI, we fitted five representation similarity models (Figure 6a) and other control regressors (see Methods

for details) to assess different representations. The representation models encompass the following: Context (indicating whether the cue was classified as filled or hollowed), Context × Shape, Context × Color (reflecting whether the task-relevant feature of the cue is shape-based or color-based depending on the context), Task (indicating whether the cue is associated with a face or scene response rule), and Decision (specifying the integrated task representation guiding subject responded with their left or right finger).

We found that different regions encode different constituent representations of the task. Specifically, context representation was encoded in the left rostral prefrontal cortex (PFC), the left inferior frontal gyrus, the left dorsal medial PFC, the left temporoparietal junction, the left middle and the inferior temporal gyrus, the left precuneus, the right dorsal medial PFC and the right middle temporal gyrus. Task-relevant feature representations given the task context (Context × Color and Context × Shape) were encoded in the left rostral PFC, the medial dorsal PFC, the sensorimotor cortex, the precuneus, the temporal cortex and the occipital cortex. We did not observe positively significant cortical ROIs related to task performed representation. Decision representation was encoded in the dorsal lateral PFC (DLPFC), the insular cortex, the sensorimotor cortex, the intra parietal sulcus, and the anterior temporal cortex (Figure 6c).

Having determined the cortical regions that preferentially encode different representations, we then asked if thalamocortical interactions can predict voxel-wise activity patterns within these regions (Figure 6b), and whether there is a difference in prediction between different task switching conditions. To address this question, we performed voxel-to-voxel thalamocortical activity flow mapping analysis for cortical ROIs that encode different levels of task representations. We first compared our voxel-to-voxel thalamocortical interaction model to two null models. These comparison models assumed neither the spatial pattern of thalamic evoked response nor the thalamocortical functional connectivity pattern carried task-related information. We found that our thalamocortical interaction model for all three task-switching conditions outperformed the two null models across cortical regions representing different levels of task representations (see Supplementary Materials S2). The results suggest that the thalamocortical interaction model has better than chance performance on predicting different levels of cortical task representations during the hierarchical task-switching. However, we found no significant difference in prediction accuracy between EDS, IDS, and Stay conditions, therefore for the next prediction analysis we used the evoked responses from all conditions combined, which models across these three conditions.

We then compared the prediction accuracy between ROIs that encode different levels of task representations. The goal is to determine whether the thalamocortical model can better predict cortical regions that encode a specific constituent representation. One-way rmANOVA showed that there is a significant difference between different cortical representations ($F(3,174)=31.62$, $p<0.0001$, $\eta^2=0.16$; Context (mean ± SD): 0.36±0.33, Context × Color (mean ± SD): 0.20±0.34, Context × Shape (mean ± SD): 0.19±0.35, Decision (mean ± SD): 0.26±0.35; Figure 6d). Specifically, the thalamocortical model displayed highest prediction accuracy in predicting context representation (Context vs. Context × Color: $t(58)=6.04$, $p<0.0001$, Cohen's d=0.87; Context vs. Context × Shape: $t(58)=7.24$, $p<0.0001$, Cohen's d=1.04; Context vs. Decision : $t(58)=5.80$, $p<0.0001$, Cohen's d=0.51), then followed by predicting decision cortical representation (Decision vs. Context × Color: $t(58)=2.94$, $p=0.03$, Cohen's d=0.41; Decision vs.

Context × Shape: $t(58)=4.51$, $p=0.0002$, Cohen's d=0.60). We did not find significant difference of prediction accuracy between the two feature cortical representations (Context × Color vs. Context × Shape: $t(58)=2.49$, $p=0.09$, Cohen's d=0.20).

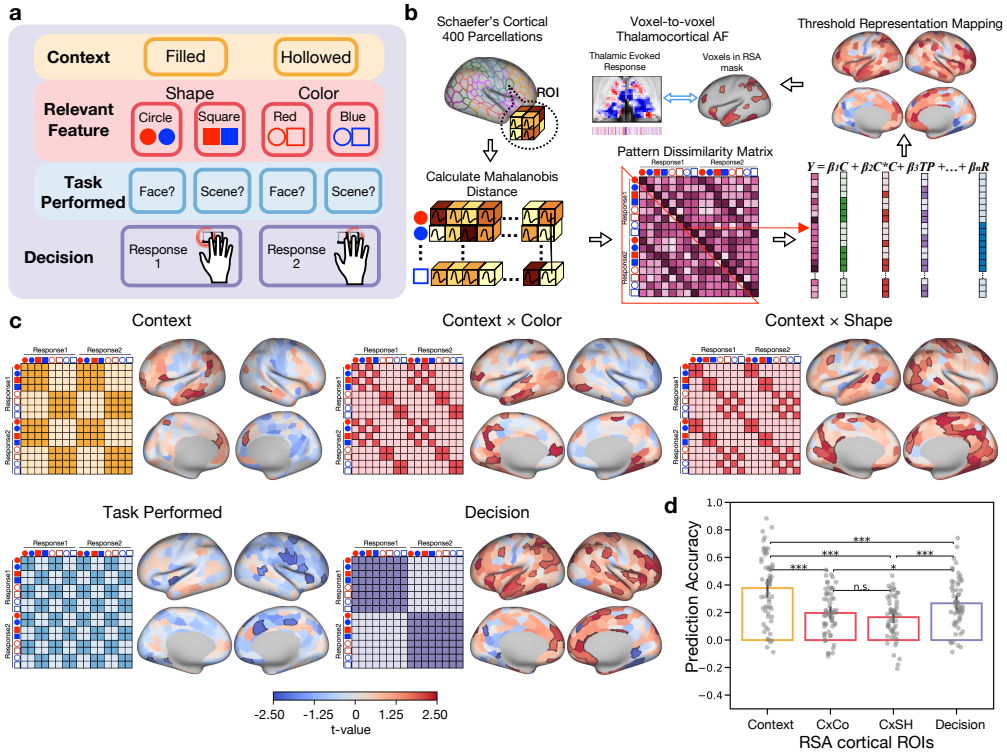

Figure 6. Thalamocortical updates the cortical control representation. *(a) The hierarchical structure of task representations. (b) Activity flow mapping and ROI-RSA procedures. Within each ROI, we calculated the Mahalanobis distance of cortical activity pattern between different types of cues to evaluate the neural dissimilarity matrix. The lower triangle of the neural dissimilarity matrix was extracted and regressed onto an intercept and different representational models. The resulted beta coefficients were subjected to group-level analysis by computing voxel-wise t-test across subjects (3dttest++ in AFNI). The thresholded results were used as mask for the further voxel-to-voxel thalamocortical interaction model. (c) Five pattern dissimilarity matrix that modeled different representations. Cortical regions that encode different representations within the hierarchical task representation. The white outline in each cortical map depicts statistically significant clusters that used in the voxel-to-voxel thalamocortical activity flow mapping. (d) The prediction accuracy of voxel-to-voxel thalamocortical activity flow model for different hierarchical task representations. Multiple comparisons were corrected by the Bonferroni procedure. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.*

To further understand how different thalamic subregions influence cortical representations, we performed a virtual lesion analysis. We found that removing thalamic voxels from the thalamocortical activity flow model resulted in a decrease in the model's ability to predict voxel-wise activity patterns in cortical ROIs that encode all levels of hierarchical task representations (Figure 7). We found that removing voxels from the anterior, the ventrolateral, the mediodorsal,

the intralaminar, and the medial pulvinar, led to a consistent reduction in model's performance across different levels of cortical representations.

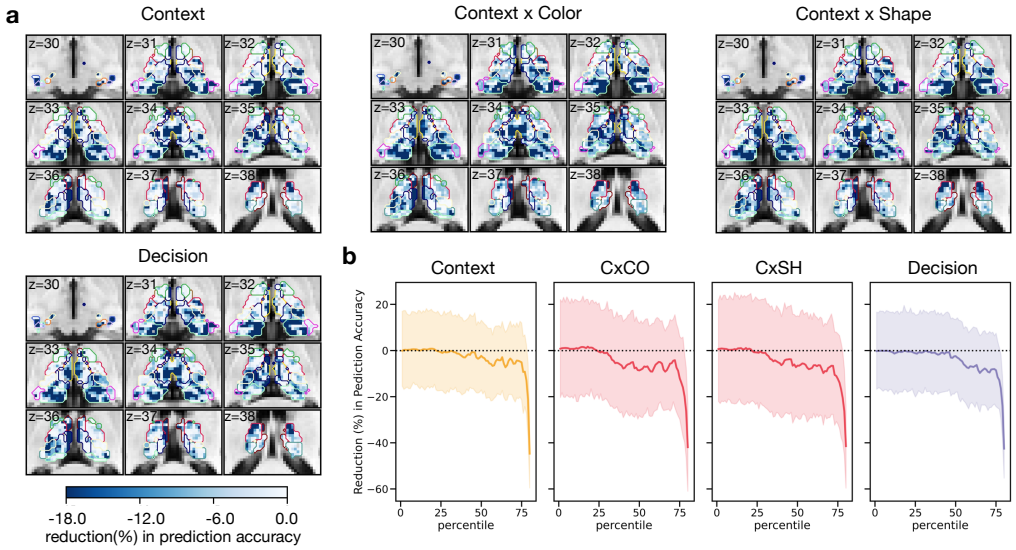

Figure 7. Thalamic topography of reduction in model prediction. *(a) The group averaged results of thalamic topography of reduction in model performance when predicting various cortical representations. (b) Reduction in prediction accuracy for the different cortical representations after lesioning thalamic voxels ranking by the amplitudes of evoked responses, ranging from the 1st to the 80th percentile.*

Evoked activity in the basal ganglia (BG) and BG-cortical activity flow analysis

Our findings have demonstrated that thalamocortical interactions can predict activity patterns associated with hierarchical cognitive control. Another subcortical region likely critical for hierarchical control is the BG (Badre & Nee, 2018; Nee & Brown, 2013). The BG has been hypothesized to inhibit or disinhibit thalamocortical circuits when cortical representations need to be updated; however, it does not have direct projections to the cortex, therefore the cortical representations of task information is likely to involve thalamocortical interactions that are potentially under the influence of the BG inhibitory output. It is unknown whether the thalamus and the BG will show similar or distinct response and activity flow prediction patterns, therefore we investigated the role of the BG in predicting cortical activity patterns. Our prediction was that the BG will exhibit a similar response profile as the thalamus given the known frontal-BG involvement in hierarchical cognitive control (Badre & Nee, 2018; Nee & Brown, 2013). However, considering that the output of BG, especially the globus pallidus nucleus, is mediated by the thalamus to influence cortical functions, we predicted that it would be less effective at predicting cortical activity patterns compared to the thalamus.

We first investigated how hierarchical task-switching modulates evoked responses in the BG. We found that the caudate nucleus and the putamen showed significant evoked responses to different hierarchical task-switching conditions (EDS, IDS and Stay conditions; unthresholded: Figure 8a; thresholded activity maps at cluster corrected $p < .05$: Figure 8b) but not the globus pallidus. When comparing these conditions, the caudate nucleus displayed a stronger evoked

response for EDS-IDS and EDS-Stay contrasts (Figure 8c). No significant clusters were observed for the IDS-Stay contrast. Similar to the patterns observed in the thalamus and the cortex, as the demand for hierarchical cognitive control increased, the caudate nucleus exhibited much stronger activity during EDS compared to IDS and Stay conditions.

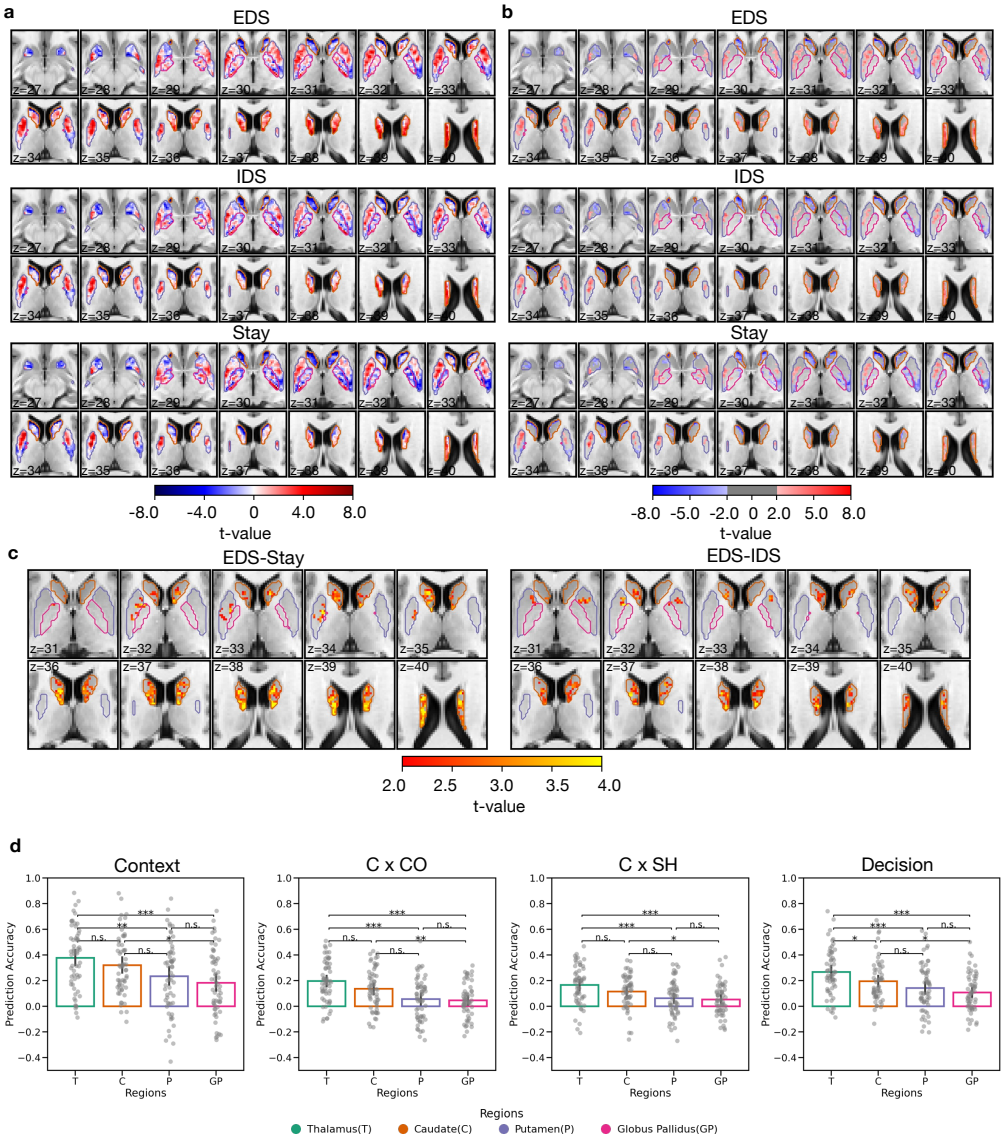

Figure 8. BG involvement into hierarchical task switching. *(a) Unthresholded BG evoked responses (t-value) in three hierarchical task-switching conditions. (b) Thresholded BG evoked response in the three hierarchical task-switching conditions. (c) Statistically significant clusters in BG evoked responses were observed in EDS-Stay and EDS-IDS contrasts. There were no statistically significant results of IDS-Stay. (b)(c) The colored voxels depict the cluster-based corrected significant results at p<0.05. (d) Comparing the prediction accuracy of the three conditions between thalamocortical and BG-cortical models for parcellations identified from Figure 6(c). The error bar represents the 95% confidential interval.*

While the BG also exhibited evoked response modulations by different hierarchical task-switching conditions, we expect the BG-cortex interaction model to have weaker performance in predicting task-related cortical activity when compared to the thalamocortical model. This is because the BG modulates thalamocortical interactions without directly projecting to the cortex.

A BG-cortical activity flow model can therefore serve as a comparison model for our thalamocortical model. Instead of using the thalamic evoked response and thalamocortical FC matrix, we predicted the voxel-wise cortical activity pattern for each of representation cortical areas by calculating the dot product of voxel-wise BG evoked response and voxel-wise FC between BG voxel and every voxel within any given cortical region (BG model performance see Supplementary Materials S3&S4). The resulting predicted cortical activity pattern was compared with the observed cortical activity pattern using Pearson correlation, normalized by the noise ceiling (see Methods for details). Consistent with our prediction, our findings demonstrated that thalamus model exhibited superior performance to all BG-cortical models when predicting cortical representations that encodes integrated task representation (Decision; Figure 8d). When predicting cortical representations encoding contextual (Context) and relevant feature (Context × Color, Context × Shape) representations (Figure 8d), thalamus outperformed putamen and globus pallidus model (Figure 8d).

Discussion

The goal of the current study is to elucidate the role of the human thalamus in hierarchical cognitive control. Specifically, our study aimed to investigate the role of the thalamus in encoding and updating task representations within a multi-level hierarchical task structure, encompassing context, task-relevant features, and decision rules. We made three key findings. First, we observed increased evoked activity in the anterior, the medial, the mediodorsal, and the posterior thalamus during hierarchical cognitive control, particularly in response to updating contextual representations. Second, these thalamic subregions predominantly encoded context representations within the hierarchical task organization. Finally, we found that these thalamic regions interacted primarily with cortical areas that encode context representation, including the rostral PFC, the medial PFC, the temporoparietal junction and the temporal cortex. Specifically, thalamocortical interactions were able to predict voxel-wise activity patterns in these cortical regions. Our results suggest that the thalamus, along with its associated thalamocortical interactions, encode and update context representations during hierarchical cognitive control.

The representational structure of hierarchical cognitive control enables adaptive and flexible cognitive control. Task representations consists of numerous constituent representations, that are at least partially encoded by different cortical regions (Badre et al., 2009; Badre & D'Esposito, 2007). In this study, we designed a task where subjects performed with a hierarchically organized task representation. Critically, our design allowed us to dissociate different representations without comparing conditions with varying stimuli complexity, difficulty, or response requirements (Cellier et al., 2022). With this approach, we mapped cortical areas that encode different levels of task representations. We identified that the context-level representations were predominately encoded in the rostral PFC, the medial PFC, the temporoparietal junction and the temporal cortex. Task-relevant representations were found to be encoded in the DLPFC, the medial dorsal PFC, the sensorimotor cortex, the precuneus, the temporal cortex and the occipital cortex. Furthermore, representations pertaining to the decision of the task were localized in the DLPFC, the insular cortex, the sensorimotor cortex, the intra parietal sulcus, and the anterior temporal cortex. Given that cortical functions are intimately linked to the thalamus (Shine et al., 2023), our study investigates the function and representation encoded by the thalamus within this characterized hierarchical task structure.

Our results from MVPA analysis revealed that the thalamus preferentially encode contextual information, but not other task-relevant information within the hierarchical task structure. Notably, several thalamic nuclei, including the anterior, the ventroanterior, the ventrolateral, the mediodorsal, the intralaminar, the medial pulvinar and the lateral pulvinar nuclei, showed strong context coding. This findings is consistent with prior studies of human imaging and non-human animal models, which found that in behavioral tasks where subjects have to apply distinct response strategies under different contexts, the mediodorsal thalamus primarily encode context representations (Hummos et al., 2022; Rikhye et al., 2018).

The selective encoding of context representation by the thalamus likely constrains how thalamocortical processes support hierarchical control. We found that the several thalamic nuclei also exhibited stronger evoked responses during task-switching conditions that involved updating the context representation (EDS). Significant task-evoked activity in multiple thalamic nuclei,

including the anterior, the ventrolateral, the mediodorsal, the lateral posterior, the intralaminar, the ventroanterior, the medial pulvinar, the lateral pulvinar, and the ventral posterior nuclei. These thalamic nuclei showed most pronounced increase in evoked activity when subjects updated the highest-order task context. Furthermore, task-switching that updated other constituent representations within the hierarchical task structure did not elicit stronger evoked responses in the thalamus, further demonstrating the specificity of thalamic contextual coding.

A recently developed computational model has provided a hypothetical mechanism for how the thalamus can encode contextual representations (Hummos et al., 2022). In this model, neurons in the mediodorsal thalamus receive converging inputs from prefrontal neural populations that encode different high-dimensional representations of task rules. Through this converging input, thalamic neurons extract a summary signal from multiple converging cortical inputs, encoding a low-dimensional, compressed representation abstracted from the high-dimensional inputs. This observation is further consistent with our prior fMRI studies, where we found that relative to the cortex, the thalamus encodes low-dimensional structures of cognitive activity (Hwang et al., 2022), and that thalamic subregions showing the strongest context coding also exhibit converging connectivity (hub-like connectivity) with multiple frontoparietal structures (Hwang et al., 2017, 2021).

In the cortex, we observed stronger evoked responses in several frontoparietal regions during hierarchical task switching. Our findings are consistent with prior research (Badre et al., 2009; Badre & D'Esposito, 2007; Bahlmann et al., 2015; Nee & Brown, 2013), suggesting that updating higher-level contextual information engages regions such as the rostral PFC, the posterior parietal, the posterior cingulate, and the insular areas, while updating lower-level task-relevant information predominantly involves the premotor and the parietal regions. To further investigate how thalamic functions may interact with these cortical regions, we developed a novel activity flow mapping analysis (Cole et al., 2016, 2021; Ito et al., 2017). This method models thalamic evoked responses and thalamocortical connectivity patterns as input features to predict cortical activity patterns. We then compared our model with two null models, which posited that neither the thalamus nor thalamocortical connectivity carries task-specific information. We found that our thalamocortical interaction model significantly outperformed the two null models, successful in predicting cortex-wide activity patterns associated with our task.

To further determine the specific cortical task representation influenced by thalamocortical interactions, we adapted the model to predict voxel-wise activity patterns within cortical ROIs encoding different levels of task representation. Our results showed that the thalamocortical model most accurately predict voxel-wise activity patterns in cortical ROIs that encode the context representation, outperforming ROIs encoding other representations within the hierarchical task structure. This finding is further consistent with our prior findings demonstrating the thalamic preference for encoding the context of hierarchical task representations, and showed increased activity during task-switching conditions that involved context update.

A logical question concerns which thalamic subregions contributed to predicting context related cortical activity patterns? Subsequent simulated lesion analysis revealed that certain thalamic subregions had stronger contribution, particularly the anterior, the ventrolateral, the mediodorsal,

the intralaminar, and the medial pulvinar nuclei. The removal of these thalamic nuclei from the thalamocortical model led to a decrease in the performance of model performance, and the decrement was more pronounced than other thalamic subregions. These thalamic subregions contributing most to activity flow model performance spatially align with those whose evoked responses were mostly strongly modulated by contextual task switching. We speculate that these identified thalamic subregions are particularly well-suited for contextually guided cognitive control due to their hub-like connectivity property, capable of extracting context representations via its converging inputs, and ideally positioned to influence cortical representations via its distributed connectivity (Hwang et al., 2017). Indeed, we found that removing these thalamic subregions from the activity flow prediction model also negatively affected the prediction of activity patterns in ROIs encoding other task-related features. This finding suggests that in cortical regions not strongly encoding context representations, receiving contextual influence from the thalamus may also be critical for the selective processing of task-relevant features. Anatomically, it is long known that the anterior, the mediodorsal, and the pulvinar nuclei have a strong anatomical relationship with frontoparietal regions that we found to encode context representations (Goldman-Rakic & Porrino, 1985; Shibata & Naito, 2005; Yeterian & Pandya, 1985). For example, the anterior and mediodorsal thalamus have projections to both the lateral and medial prefrontal cortex, including the medial superior frontal gyrus, the superior frontal gyrus and the middle frontal gyrus (Ketz et al., 2015; Sun et al., 2018; Zikopoulos & Barbas, 2006). The pulvinar nucleus mainly projects to the prefrontal, the posterior parietal, the temporal, and the extrastriate areas (Grieve et al., 2000; Jones et al., 1979; Robinson & Petersen, 1992; Yeterian & Pandya, 1985).

The BG, also a crucial part of the cortico-striatal-thalamic circuit, is proposed to act as a gate for updating task representation in the cortex (Chatham et al., 2014; Nee & Brown, 2013; O'Reilly & Frank, 2006; Rac-Lubashevsky & Frank, 2021). While the basal ganglia's response to hierarchical task-switching was found to be similar to that of the thalamus, its ability to predict cortical activity patterns was significantly weaker than the thalamocortical model. Our previous work also observed a better performance for the thalamocortical model when compared to the BG-cortical model in predicting the cortical activity related to a working memory task (Chen et al., 2023) and cognitive tasks across multiple behavioral domains (Hwang et al., 2022). These results suggested that the thalamus has a more direct impact on cortical activity pattern than BG in predicting the hierarchical task representations. In the traditional BG-thalamic gating model, the ventrolateral and ventroanterior nucleus is the major target of BG inhibitory output (DeVito & Anderson, 1982; Sidibé et al., 1997). Other thalamic nuclei, including several that we identified, such as the mediodorsal, the anterior, and the pulvinar nucleus, may also play an essential role in hierarchical task switching by receiving inhibitory outputs from other basal ganglia structures such as the substantia nigra (Ilinsky et al., 1985) and the ventral pallidum (Root et al., 2015; Young et al., 1984). More recently, in rodent models, it has also been demonstrated that the mediodorsal thalamus is capable of directly inhibiting and exciting cortical context representations (Rikhye et al., 2018). Incorporating these findings with our own, the thalamocortical interaction likely plays an active role beyond merely receiving inhibitory output from BG. While we found that thalamocortical interactions involving multiple thalamic subregions can predict cortical activity patterns associated with context representations, due to the limitations of fMRI, we were not able to discern whether it is exciting or inhibiting cortical representations under different task conditions.

To summarize, we provided a detailed characterization of how the human thalamocortical system contributes to hierarchical cognitive control. Our results are consistent with the interpretation of the thalamus as a central and active node contributing to brain-wide information processes. The functional hubs within the thalamus, which specialize in processing contextual information, play a key role in encoding and updating context representations to facilitate flexible, contextually guided cognitive control. This function is vital for the brain's capacity to adapt and respond to evolving task requirements.

Methods

Subjects

Seventy-four healthy adult subjects participated in our functional magnetic resonance imaging (fMRI) study. Fifteen subjects were excluded from further data analysis based on the following criteria: five subjects were excluded because of poor behavioral performance (accuracy < 0.7); one subject was excluded because of failure in the response equipment; one subject voluntarily dropped out midway through the experiment; eight subjects were rejected for excessive head motion (mean framewise displacement (FD) mean > 0.4 and more than 33.3% of data removed because of excessive head motion). Therefore, a total of 59 subjects (44 females and 15 males, age: 22.32±4.71) were included for data analyses. All subjects were right-handed, had normal or corrected-to-normal vision, and reported no history of a neurological conditions. All subjects provided written informed consent in accordance with procedures approved by the Institutional Review Boards at the University of Iowa (IRB-01-201808855).

Hierarchical cognitive control task

In our behavioral task, subject utilized a hierarchically structured task representation to determine which response rule to adopt to respond to the probe stimuli (Cellier et al., 2022). In each trial, one of eight different colored geometric cues was presented for 0.5s, and then was immediately followed by a probe image of either a face or a scene. The probe image was presented for 2.5s. Subjects were asked to respond during this time window. Trials were separated by a jittered inter-trial-interval that varied between 1.5s to 10.5s (average: 4.5s) with a "+" at the center of monitor (Figure 1a). Subjects were required to respond to the probe picture based on one of two response rules: judging whether the picture depicted a face (face rule) or a scene (scene rule). The correct response rule depended on three attributes of the cue (fill texture: solid or hollow, shape: square or circle, color: blue or red), and these attributes were organized in a hierarchical fashion (Figure 1b). The highest level of attributes, referred to as the "context," evaluated the fill texture of the cue (solid vs. hollow). This context information determined which mid-level "feature" subjects should utilize to determine the response rule. Specifically, a solid cue prompted the participant to judge whether the shape of the cue was a circle or a square, while a hollow cue instructed them to judge whether the color of the cue was blue or red. This mapping was counterbalanced across subjects. Each feature from the mid-level was then associated with either the face or scene rule. For instance, a circle corresponded to the face rule, while a square corresponded to the scene rule. Once the correct response rule was determined, participants pressed a button to indicate "yes" or "no" based the probe picture. The responding buttons were counterbalanced across subjects. Each subject completed 8 runs (about 9min/run) in total, and each run contained 51 trials.

The hierarchical structure of our behavioral task allowed us to introduce three levels of task-switching operations that update task representations: extra-dimension switch (EDS), inter-dimension switch (IDS), and Stay (Figure 1a&b). For EDS subjects were required to update the highest-level of task context. IDS, on the other hand, switched between task-relevant features under the same task context. Lastly, during Stay trials subjects maintained the same task representation for two consecutive trials.

Magnetic resonance imaging

All participants were scanned at the Magnetic Resonance Research Facility at the University of Iowa, utilizing a GE SIGNA Premier 3T MRI Scanner with a 48-channel head coil. Structural images were acquired using a multi-echo MPRAGE sequence (TR = 2348.52ms; TE = 2.968ms; flip angle = 8°; field of view = 256 * 256; 200 sagittal slices; voxel size = 1mm^3). Functional images were acquired using an echo-planar sequence sensitive to blood oxygenated level-dependent (BOLD) contrast (TR = 1800ms; TE = 30ms; flip angle = 75°; voxel size = 2.5mm^3).

fMRI preprocessing

All fMRI data were preprocessed using fMRIPrep version 20.1.1 (Esteban et al., 2019) to reduce noise and transform the data from the subject's native space to the ICBM 152 Nonlinear Asymmetrical template version 2009c for group analysis (Fonov et al., 2011). The preprocessing steps included bias field correction, skull-stripping, co-registration between functional and structural images, tissue segmentation, motion correction, and spatial normalization to the standard space. As part of the preprocessing, we included rigid-body motion estimates, cerebral spinal fluid (CSF), and white matter (WM) noise components obtained through the component-based noise correction procedure (Behzadi et al., 2007) in the regression model to reduce the influences from noise and artifacts. We did not perform any spatial smoothing.

Behavioral data analysis

We performed one-way repeated measure ANOVA with three levels (EDS, IDS and Stay) to examine the effects of hierarchical task-switching on response time and accuracy. Incorrect trials were not included in for the reaction time analysis. We applied Bonferroni's correction method with a family-wise error rate (FWER) of 0.05. We reported Cohen's d as the metric of effect size.

fMRI data analysis

Evoked responses

The analysis of evoked responses was conducted using AFNI (Cox, 1996). To estimate the evoked response for individual subject, we performed voxel-wise general linear model (GLM) analysis, which combined a generalized least squares regression with voxel-wise restricted maximum likelihood (REML) estimate of an ARMA (1,1) temporal correlation model (3dDeconvolve and 3dREMLfit). For each voxel, the GLM was constructed and fitted to the preprocessed timeseries with the task regressors and nuisance regressors (see fMRI Preprocessing section). All task regressors were modeled with 9 basic TENT functions that lasted from time 0 to 14.4 seconds post cue, separately for each condition. High motion volumes (FD > 0.4mm) were removed via the censoring option in 3dDeconvolve, and subjects with more than 33.3% of the data censored were dropped from further analysis. The group analysis involved examining the three task-switching conditions (EDS, IDS, and Stay), as well as contrasting pairs of task-switching conditions (EDS-IDS, IDS-Stay, and EDS-Stay). This was

performed using a linear mixed model with a local estimate of random effects variance (3dMEMA). To correct for multiple comparisons, we used a whole-brain, family-wise error cluster correction procedure (3dClustSim), where the smoothness was estimated using the autocorrelation function option.

Functional connectivity (FC)

To estimate task-state thalamocortical FC, we calculated the background connectivity (Cole et al., 2019) between thalamus and cortices by using the residual from the GLM procedure (see Evoked Responses section). For each participant, we extracted voxel-wise signals (using residuals from GLM) from the thalamus, based on the Morel Atlas (Krauth et al., 2010), and from 400 cortical regions of interest (Schaefer et al., 2018) and then performed principal component regression analysis. Subsequently, a principal component analysis (PCA) was applied to the fMRI time series across all thalamic voxels, using the maximum available number of components (for instance, 199 components when 200 time points were present). These components were entered as predictors and cortical ROI data as the dependent variable in a linear regression model. The resulting beta coefficients were then transformed back to original thalamocortical space, which described the linear relationship between thalamus time series and cortical time series.

Multivoxel pattern analysis (MVPA)

We aimed to understand the level at which the thalamus preferentially classifies task-relevant information in a hierarchical structure. We performed a MVPA approach, using a spherical searchlight procedure with a radius of $6mm^3$. We first removed the time series of noise by applying nuisance regressors (see Preprocessing section) for data modeling (3dTproject in AFNI). The Least Squares Sum method (3dLSS in AFNI) was then used to model the trial-by-trial evoked responses for each of the 8 cues. For each searchlight region, we performed a logistic regression classifier to decode the multivoxel activity pattern associated with the context (solid vs. hollow), color (blue vs. red), shape (square vs. circle) and task (face vs. scene) respectively. Our cross-validation strategy involved a leave-one-group-out approach, such that one functional run of data was reserved as test data while the classifiers were trained on other runs. To assess our findings at the group level, we conducted voxel-wise t-test across subjects (3dttest++ in AFNI), comparing the decoded results against to a chance level of 0.5.

Activity flow mapping analysis

Activity flow mapping analysis framework

We modified the activity flow mapping procedure (Cole et al., 2016, 2021; Ito et al., 2017) to determine whether thalamocortical (or BG-cortical) interactions can predict cortical task activity patterns during hierarchical cognitive control. This activity flow model was calculated by computing the dot product between the seed region (e.g., thalamus or BG) evoked response vector and the functional connectivity (FC) matrix between the seed region and cortices (e.g., thalamocortical FC or BG-cortical FC):

$$P_{ctx} = O_{seed} \cdot WFC_{SeedCtx}$$

Here, P_{ctx} refers to the vector of predicted parcellation or voxel-wise cortical evoked response for the voxel-to-parcellation and voxel-to-voxel activity flow procedure respectively. O_{seed} corresponds to the vector of the voxel-wise seed region (e.g., thalamus or BG nuclei) evoked responses, which is the task-specific beta value estimated using subject-level voxel-wise GLM estimates. $WFC_{SeedCtx}$ refers to FC matrix between the seed region and the cortices (e.g., thalamocortical FC or BG nuclei-cortical FC), which is the beta value estimated using subject-level principal component regression analysis.

The model performance was evaluated by the prediction accuracy from each of activity flow mapping procedures. Prediction accuracy for the three hierarchical task-switching conditions (EDS, IDS and Stay) was evaluated using split-half cross validation. We calculated Pearson's correlation between z-transformed observed and predicted cortical evoked responses. Data were split into two parts and correlations were computed between observed evoked responses from the first half and predicted evoked responses from the second half, then vice versa. The average of these two correlations gave us the final prediction accuracy.

Noise ceiling

Note that since the upper bound for the prediction accuracy is limited by the measurement noise of cortical and thalamic evoked responses, prediction accuracy was further normalized by the split-half noise ceiling (King et al., 2023), taking into account the split-half model reliability:

$$r_{nc} = \frac{\sqrt{vU}}{\sqrt{vU + vE}}$$

Here, r_{nc} refers to the split-half noise ceiling metric, and vU is the variance of true signal, the predicted cortical patterns in first four runs, whereas vE is the variance of noise, the difference of predicted cortical patterns between first and last four runs. The reported split-half prediction accuracy was calculated as the split-half cross-validated prediction accuracy normalized by split half noise ceiling. We did not observe any significant difference of noise ceiling metric across three hierarchical task-switching conditions for any type of activity flow procedure (see Supplementary Materials S5).

Null models

To evaluate the performance of the activity flow mapping analysis, we compared the prediction accuracy of the model to two null models. The first null model was constructed by randomly shuffling each subject's seed region evoked responses (O_{seed} in the equation) with 5,000 random permutations. The second null model was constructed by randomly shuffling elements within the FC matrices between the seed region and cortices ($WFC_{SeedCtx}$ in the equation) with 5,000 random permutations. These two null distributions satisfied the null hypothesis that neither the spatial pattern of seed region evoked responses nor the patterns of functional connectivity between seed region and cortices carry task-specific information.

Effect of simulated thalamic lesions on activity flow mapping

To determine which thalamic subregions contribute to predicting cortical activity patterns, we systematically removed 20% of thalamic voxels from the activity flow mapping analysis based on the strength of evoked response amplitudes (the absolute value ranked from weakest to strongest). We stepped the removal from the 1st to 80th percentile in 1 percentage increments. Activity for removed voxels was set to zero before calculating the predicted patterns. The reduction percentage in prediction accuracy was calculated as the difference of prediction accuracy before and after the removal of thalamic voxels.

Thalamocortical interactions and cortical control representations

Representational Similarity Analysis (RSA)

To investigate brain regions involved in encoding different constituent representations within the hierarchically structured task representations, we performed ROI-based RSA looping through 400 cortical ROIs (Schaefer et al., 2018). We initially conducted voxel-wise GLM analysis, using 9 basic TENT functions to model the impulse response of 8 different cues from time 0 to 14.4 seconds following each cue, as implemented using the 3dDeconvolve. Within the ROI, we averaged the signals over 9 TRs for each voxel and then calculated the neural dissimilarity between voxel-wise patterns for each pair of cues (8 cues in total). The neural dissimilarity was computed as one minus the spatial Mahalanobis distance. We then fitted the following representational models onto the observed dissimilarity matrix: (1) Context: whether the cue is classified as filled or hollowed. (2) Task-relevant feature: the interaction between context and color; the interaction between context and shape. (3) Task performed: whether the cue is associated with the face or scene response rule. (4) Decision: specifying the integrated task representation guiding subject responded with their left or right finger. (5) Other control regressors include the main effect of shape and color; the two-way interactions: shape × task performed; color × task performed; color × shape; context × task performed; the three-way interactions: context × shape × task performed; context × color × task performed.

The lower triangle of the neural dissimilarity matrix was extracted and regressed onto an intercept and the lower triangle of the representational models, resulting in one beta coefficient per variable, per subject, per ROI. Group-level analyses were conducted by computing voxel-wise t-test across subjects (3dttest++ in AFNI). To control the false positive rate (FPR) globally, the option "-Clustsim" was used to calculate the cluster-based thresholding determined by 10,000 3d t-statistic maps from randomization (Cox et al., 2017).

Voxel-to-voxel thalamocortical activity flow on cortical regions representing hierarchical task representation

To test whether the thalamocortical activity flow model can predict cortical regions that found to encode specific representations, we conducted RSA to identify cortical regions responsible for encoding context, task-relevant features, task performed, and decision. Subsequently, we tested

whether the thalamocortical activity flow model could successfully predict the voxel-wise activity patterns in the identified ROI for the EDS, IDS, Stay and All (which models evoked response of all task-switching conditions) conditions. We first compared the model performance on these four conditions relating to hierarchical switching (EDS, IDS, Stay and All conditions) to two null models respectively (see Null models). We then performed a one-way repeated measure ANOVA to examine the effects of hierarchical task switching (EDS, IDS and Stay conditions) on model performance. To examine differences in model performance across different cortical representations, we performed a one-way repeated measure ANOVA specifically on the All condition. We applied Bonferroni's correction method with a family-wise error rate (FWER) of 0.05. We reported Cohen's d as the metric of effect size. To determine which thalamic subregions contribute to predicting cortical activity patterns of different levels of task representation, we performed simulated thalamic lesions on activity flow mapping (see Effect of simulated thalamic lesions on activity flow mapping section).

Reference

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266–271. https://doi.org/10.1016/0166-2236(90)90107-L

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, *9*(1), 357–381. https://doi.org/10.1146/annurev.ne.09.030186.002041

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2007). Functional Magnetic Resonance Imaging Evidence for a Hierarchical Organization of the Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *19*(12), 2082–2099. https://doi.org/10.1162/jocn.2007.19.12.2082

Badre, D., & Frank, M. J. (2012). Mechanisms of Hierarchical Reinforcement Learning in Cortico–Striatal Circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. https://doi.org/10.1093/cercor/bhr117

Badre, D., Hoffman, J., Cooney, J. W., & D'Esposito, M. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience*, *12*(4), 515–522. https://doi.org/10.1038/nn.2277

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Bahlmann, J., Blumenfeld, R. S., & D'Esposito, M. (2015). The Rostro-Caudal Axis of Frontal Cortex Is Sensitive to the Domain of Stimulus Information. *Cerebral Cortex*, *25*(7), 1815–1826. https://doi.org/10.1093/cercor/bht419

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Cellier, D., Petersen, I. T., & Hwang, K. (2022). Dynamics of Hierarchical Task Representations. *Journal of Neuroscience*, *42*(38), 7276–7284. https://doi.org/10.1523/JNEUROSCI.0233-22.2022

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal Output Gating during Selection from Working Memory. *Neuron*, *81*(4), 930–942. https://doi.org/10.1016/j.neuron.2014.01.002

Chen, X., Sorenson, E., & Hwang, K. (2023). Thalamocortical contributions to working memory processes during the n-back task. *Neurobiology of Learning and Memory*, *197*, 107701. https://doi.org/10.1016/j.nlm.2022.107701

Christoff, K., Keramatian, K., Gordon, A. M., Smith, R., & Mädler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain Research*, *1286*, 94–105. https://doi.org/10.1016/j.brainres.2009.05.096

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726. https://doi.org/10.1038/nn.4406

Cole, M. W., Ito, T., Cocuzza, C., & Sanchez-Romero, R. (2021). The Functional Relevance of Task-State Functional Connectivity. *Journal of Neuroscience*, *41*(12), 2684–2702. https://doi.org/10.1523/JNEUROSCI.1713-20.2021

Cole, M. W., Ito, T., Schultz, D., Mill, R., Chen, R., & Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. *NeuroImage*, *189*, 1–18. https://doi.org/10.1016/j.neuroimage.2018.12.054

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI Clustering in AFNI: False-Positive Rates Redux. *Brain Connectivity*, *7*(3), 152–171. https://doi.org/10.1089/brain.2016.0475

DeLong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*, *13*(7), 281–285. https://doi.org/10.1016/0166-2236(90)90110-V

DeVito, J. L., & Anderson, M. E. (1982). An autoradiographic study of efferent connections of the globus pallidus in Macaca mulatta. *Experimental Brain Research*, *46*(1), 107–117. https://doi.org/10.1007/BF00238104

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), Article 1. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. https://doi.org/10.1016/j.neuroimage.2010.07.033

Goldman-Rakic, P. S., & Porrino, L. J. (1985). The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *Journal of Comparative Neurology*, *242*(4), 535–560. https://doi.org/10.1002/cne.902420406

Grieve, K. L., Acuña, C., & Cudeiro, J. (2000). The primate pulvinar nuclei: Vision and action. *Trends in Neurosciences*, *23*(1), 35–39. https://doi.org/10.1016/S0166-2236(99)01482-4

Haber, S., & Mcfarland, N. R. (2001). The Place of the Thalamus in Frontal Cortical-Basal Ganglia Circuits. *The Neuroscientist*, *7*(4), 315–324. https://doi.org/10.1177/107385840100700408

Hummos, A., Wang, B. A., Drammis, S., Halassa, M. M., & Pleger, B. (2022). Thalamic regulation of frontal interactions in human cognitive flexibility. *PLOS Computational Biology*, *18*(9), e1010500. https://doi.org/10.1371/journal.pcbi.1010500

Hwang, K., Bertolero, M. A., Liu, W. B., & D'Esposito, M. (2017). The Human Thalamus Is an Integrative Hub for Functional Brain Networks. *Journal of Neuroscience*, *37*(23), 5594–5607. https://doi.org/10.1523/JNEUROSCI.0067-17.2017

Hwang, K., Shine, J. M., Bruss, J., Tranel, D., & Boes, A. (2021). Neuropsychological evidence of multi-domain network hubs in the human thalamus. *eLife*, *10*, e69480. https://doi.org/10.7554/eLife.69480

Hwang, K., Shine, J. M., Cole, M. W., & Sorenson, E. (2022). Thalamocortical contributions to cognitive task activity. *eLife*, *11*, e81282. https://doi.org/10.7554/eLife.81282

Ilinsky, I. A., Jouandet, M. L., & Goldman-Rakic, P. S. (1985). Organization of the nigrothalamocortical system in the rhesus monkey. *Journal of Comparative Neurology*, *236*(3), 315–330. https://doi.org/10.1002/cne.902360304

Ito, T., Kulkarni, K. R., Schultz, D. H., Mill, R. D., Chen, R. H., Solomyak, L. I., & Cole, M. W. (2017). Cognitive task information is transferred between brain regions via resting-state network topology. *Nature Communications*, *8*(1), 1027. https://doi.org/10.1038/s41467-017-01000-w

Jones, E. G., Wise, S. P., & Coulter, J. D. (1979). Differential thalamic relationships of sensory-motor and parietal cortical fields in monkeys. *Journal of Comparative Neurology*, *183*(4), 833–881. https://doi.org/10.1002/cne.901830410

Ketz, N. A., Jensen, O., & O'Reilly, R. C. (2015). Thalamic pathways underlying prefrontal cortex–medial temporal lobe oscillatory interactions. *Trends in Neurosciences*, *38*(1), 3–12. https://doi.org/10.1016/j.tins.2014.09.007

King, M., Shahshahani, L., Ivry, R. B., & Diedrichsen, J. (2023). A task-general connectivity model reveals variation in convergence of cortical inputs to functional regions of the cerebellum. *eLife*, *12*, e81511. https://doi.org/10.7554/eLife.81511

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, *302*(5648), 1181–1185. https://doi.org/10.1126/science.1088545

Krauth, A., Blanc, R., Poveda, A., Jeanmonod, D., Morel, A., & Székely, G. (2010). A mean three-dimensional atlas of the human thalamus: Generation from multiple histological data. *NeuroImage*, *49*(3), 2053–2062. https://doi.org/10.1016/j.neuroimage.2009.10.042

Mink, J. W. (1996). THE BASAL GANGLIA: FOCUSED SELECTION AND INHIBITION OF COMPETING MOTOR PROGRAMS. *Progress in Neurobiology*, *50*(4), 381–425. https://doi.org/10.1016/S0301-0082(96)00042-1

Nee, D. E., & Brown, J. W. (2013). Dissociable Frontal–Striatal and Frontal–Parietal Networks Involved in Updating Hierarchical Contexts in Working Memory. *Cerebral Cortex*, *23*(9), 2146–2158. https://doi.org/10.1093/cercor/bhs194

O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends in Neurosciences*, *33*(8), 355–361. https://doi.org/10.1016/j.tins.2010.05.002

O'Reilly, R. C., & Frank, M. J. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal Cortex and Dynamic Categorization Tasks: Representational Organization and Neuromodulatory Control. *Cerebral Cortex*, *12*(3), 246–257. https://doi.org/10.1093/cercor/12.3.246

Parent, A., & Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews*, *20*(1), 91–127. https://doi.org/10.1016/0165-0173(94)00007-C

Rac-Lubashevsky, R., & Frank, M. J. (2021). Analogous computations in working memory input, output and motor gating: Electrophysiological and computational modeling evidence. *PLOS Computational Biology*, *17*(6), e1008971. https://doi.org/10.1371/journal.pcbi.1008971

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature Neuroscience*, *21*(12), 1753–1763. https://doi.org/10.1038/s41593-018-0269-z

Robinson, D. L., & Petersen, S. E. (1992). The pulvinar and visual salience. *Trends in Neurosciences*, *15*(4), 127–132. https://doi.org/10.1016/0166-2236(92)90354-B

Root, D. H., Melendez, R. I., Zaborszky, L., & Napier, T. C. (2015). The ventral pallidum: Subregion-specific functional anatomy and roles in motivated behaviors. *Progress in Neurobiology*, *130*, 29–70. https://doi.org/10.1016/j.pneurobio.2015.03.005

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., & Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature*, *545*(7653), 219–223. https://doi.org/10.1038/nature22073

Schumacher, E. H., & Hazeltine, E. (2016). Hierarchical Task Representation: Task Files and Response Selection. *Current Directions in Psychological Science*, *25*(6), 449–454. https://doi.org/10.1177/0963721416665085

Shibata, H., & Naito, J. (2005). Organization of anterior cingulate and frontal cortical projections to the anterior and laterodorsal thalamic nuclei in the rat. *Brain Research*, *1059*(1), 93–103. https://doi.org/10.1016/j.brainres.2005.08.025

Shine, J. M., Lewis, L. D., Garrett, D. D., & Hwang, K. (2023). The impact of the human thalamus on brain-wide information processing. *Nature Reviews Neuroscience*, *24*(7), Article 7. https://doi.org/10.1038/s41583-023-00701-0

Sidibé, M., Bevan, M. D., Bolam, J. P., & Smith, Y. (1997). Efferent connections of the internal globus pallidus in the squirrel monkey: I. topography and synaptic organization of the pallidothalamic projection. *Journal of Comparative Neurology*, *382*(3), 323–347. https://doi.org/10.1002/(SICI)1096-9861(19970609)382:3<323::AID-CNE3>3.0.CO;2-5

Sun, C., Wang, Y., Cui, R., Wu, C., Li, X., Bao, Y., & Wang, Y. (2018). Human Thalamic-Prefrontal Peduncle Connectivity Revealed by Diffusion Spectrum Imaging Fiber Tracking. *Frontiers in Neuroanatomy*, *12*. https://www.frontiersin.org/article/10.3389/fnana.2018.00024

Voytek, B., Kayser, A. S., Badre, D., Fegen, D., Chang, E. F., Crone, N. E., Parvizi, J., Knight, R. T., & D'Esposito, M. (2015). Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. *Nature Neuroscience*, *18*(9), Article 9. https://doi.org/10.1038/nn.4071

Yeterian, E. H., & Pandya, D. N. (1985). Corticothalamic connections of the posterior parietal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *237*(3), 408–426. https://doi.org/10.1002/cne.902370309

Young, W. S., Alheid, G. F., & Heimer, L. (1984). The ventral pallidal projection to the mediodorsal thalamus: A study with fluorescent retrograde tracers and immunohistofluorescence. *The Journal of Neuroscience*, *4*(6).

Zikopoulos, B., & Barbas, H. (2006). Prefrontal Projections to the Thalamic Reticular Nucleus form a Unique Circuit for Attentional Mechanisms. *Journal of Neuroscience*, *26*(28), 7348–7361. https://doi.org/10.1523/JNEUROSCI.5511-05.2006