Report

Neuron

The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei

Highlights

- Cognitive complexity causes alterations in the low-dimensional whole-brain dynamics
- Deviations from the manifold dissociate correct from incorrect performance
- Low-dimensional trajectories relate to activity in medial thalamic nuclei

Authors

James Macquarie Shine,
Luke James Hearne,
Michael Breakspear, ...,
Russell Alan Poldrack,
Jason Brett Mattingley, Luca Cocchi

Correspondence

mac.shine@sydney.edu.au

In Brief

Shine et al. demonstrate that cognitive complexity reconfigures the low-dimensional state space of the human brain. The low-dimensional trajectories of whole-brain activity dissociate correct and error trials and relate to activity within the medial and posterior thalamic nuclei.

Shine et al., 2019, Neuron *104*, 1–7
December 4, 2019
https://doi.org/10.1016/j.neuron.2019.09.002

CellPress

Please cite this article in press as: Shine et al., The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei, Neuron (2019), https://doi.org/10.1016/j.neuron.2019.09.002

Neuron
Report

The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei

James Macquarie Shine,[1,11,*] Luke James Hearne,[2] Michael Breakspear,[3,4] Kai Hwang,[5] Eli Justin Müller,[1] Olaf Sporns,[6] Russell Alan Poldrack,[7] Jason Brett Mattingley,[8,9,10] and Luca Cocchi[4]
[1]The University of Sydney, Sydney, NSW 2050, Australia
[2]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ 08854, USA
[3]Hunter Medical Research Institute, University of Newcastle, Newcastle, NSW 2305, Australia
[4]QIMR Berghofer, Brisbane, QLD 4006, Australia
[5]Department of Psychological and Brain Sciences and The Iowa Neuroscience Institute, The University of Iowa, Iowa City, IA 52242, USA
[6]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA
[7]Department of Psychology, Stanford University, Stanford, CA 94305, USA
[8]Queensland Brain Institute, The University of Queensland, St. Lucia, QLD 4072, Australia
[9]School of Psychology, The University of Queensland, St. Lucia, QLD 4072, Australia
[10]Canadian Institute for Advanced Research, Toronto, ON M5G 1M1, Canada
[11]Lead Contact
*Correspondence: mac.shine@sydney.edu.au
https://doi.org/10.1016/j.neuron.2019.09.002

SUMMARY

Cognitive activity emerges from large-scale neuronal dynamics that are constrained to a low-dimensional manifold. How this low-dimensional manifold scales with cognitive complexity, and which brain regions regulate this process, are not well understood. We addressed this issue by analyzing sub-second high-field fMRI data acquired during performance of a task that systematically varied the complexity of cognitive reasoning. We show that task performance reconfigures the low-dimensional manifold and that deviations from these patterns relate to performance errors. We further demonstrate that individual differences in thalamic activity relate to reconfigurations of the low-dimensional architecture during task engagement.

INTRODUCTION

The human brain is a complex adaptive network in which coordinated activity induces dependencies between a diverse set of specialized regions that define its current functional state. A common characteristic of complex systems is that the degrees of freedom traversed by its dynamics are substantially lower than the number of components that comprise the system. This means that the functional principles of a neural system are typically opaque when analyzing individual elements in isolation. However, there are often distributed patterns that explain substantial variation in the data, such that the system demonstrates low dimensionality at the level of large-scale neural populations (Cunningham and Yu, 2014; Breakspear, 2017).

Knowledge of the brain's low-dimensional architecture affords a deeper understanding of the core principles underpinning whole-brain patterns of neural activity (Chialvo, 2010), along with an ability to dynamically track complex system activity as a function of cognitive performance (Shine et al., 2019). Additionally, this dimensionality reduction allows behavioral correlates to be appropriately interpreted in terms of the low-dimensional brain activity that underpins them (Huys et al., 2014; Kerkman et al., 2018).

In previous work, we used a combination of principal-component analysis and methods from dynamical systems theory to characterize low-dimensional state space dynamics in blood-oxygen-level-dependent (BOLD) data collected during the performance of a range of diverse cognitive tasks, including cognitive, mnemonic, emotional, social, and motor processing (Shine et al., 2019; Figure 1A). We showed that large-scale brain activity during execution of diverse cognitive functions could be approximated by dynamics on a relatively simple orbit contained within a low-dimensional phase space (or "embedding space"). The principal component (PC) that explained the most variance across the tasks comprised an integrative core of regions across frontal, parietal, and occipital cortices tasks (Figure 1A; Shine et al., 2019). The first PC also tracked with whole-brain network connectivity patterns and overlapped with known heterogeneity in regional neuromodulatory receptor gradients (Shine et al., 2019) and areas with heightened network controllability (Gu et al., 2015).

The remaining PCs describing trajectories that engaged combinations of cognitive tasks: PC_{2-4} were associated with combinations of different cognitive challenges and PC_5 was associated with switches into and out of unique task states (Shine et al., 2019). These findings suggest that state space dynamics defined by the first few PCs capture meaningful neurobiological processes that support distinct cognitive functions. However, the nature and diversity of the tasks used in the original study

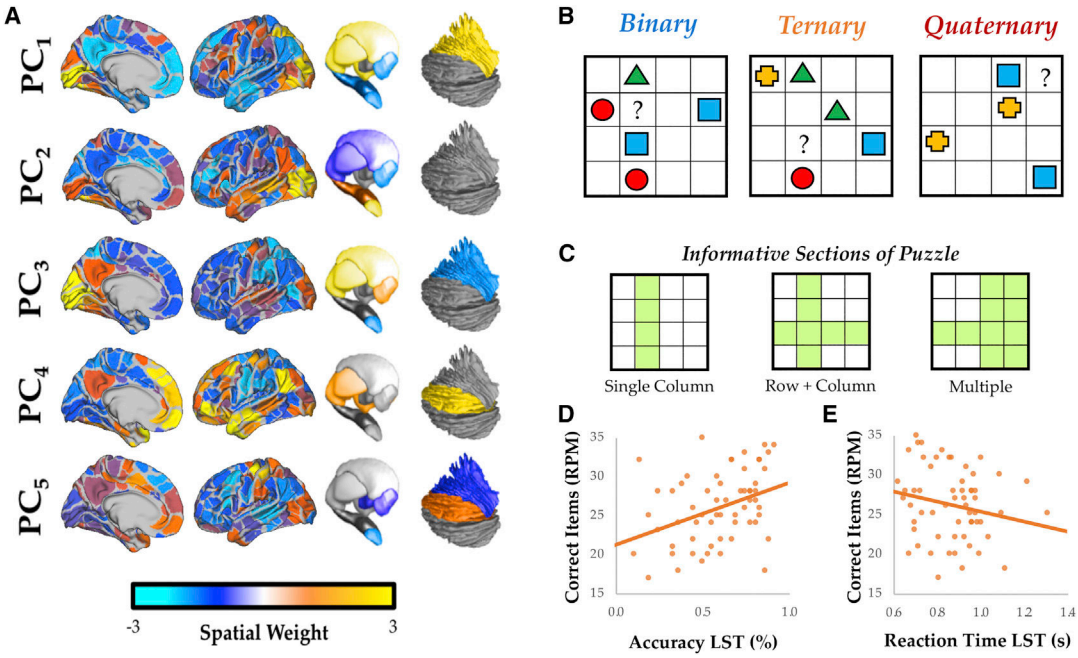

Figure 1. Low-Dimensional Embedding and the Latin Squares Task

(A) The top five spatial principal components (PCs) that explain the majority of variance across multiple cognitive tasks (Shine et al., 2019).

(B) Latin Squares Task with three levels of problem complexity: binary ("*B*"; blue); ternary ("*T*"; yellow); and quaternary ("*Q*"; red)—the goal of the task is to determine the shape denoted by the question mark ("?").

(C) Informative sections of each puzzle, indicated by green shading. *B* trials required looking across as single row or column, *T* trials required searching over a row and a column, and *Q* trials required searching over multiple rows and columns (eye movements were similar across conditions). Reasoning complexity increases linearly from *B* to *T* to *Q*.

(D) Pearson's correlation between accuracy on the LST and scores on Raven's Progressive Matrices (RPM): r = 0.395; p = 0.016.

(E) Pearson's correlation between mean reaction time on LST and scores on the RPM: r = −0.265; p = 0.003.

precluded analyses of the role of the complexity of the cognitive processes on the associated dimensionality of neural dynamics. In addition, the limited spatial resolution of 3T fMRI limited the ability to disambiguate the role of specific subcortical nuclei, such as the thalamus, which play crucial roles in cognitive processing (Bell and Shine, 2016; Hwang et al., 2017; Sherman, 2004; Hearne et al., 2017).

To resolve the underlying dynamics associated with cognitive complexity with increased precision, we acquired 7T BOLD data at high spatial and temporal resolution while 60 participants performed a cognitively challenging task (the Latin Squares Task [LST]; Figure S1A) across different levels of complexity (Birney et al., 2006; Hearne et al., 2017, 2019). Similar to the popular game "Sudoku" (Ercsey-Ravasz and Toroczkai, 2012), the LST requires participants to identify a latent organizational pattern among a set of visual shape elements (Figure 1B). Performance on the LST can be used to objectively quantify the ability of the brain to solve reasoning problems of increasing complexity (Birney et al., 2006; Halford et al., 2007). Solving the most challenging problems of the LST (a constrained combinatorial optimization problem) requires the on-line manipulation of a number of abstract associations (Halford et al., 2007; Rutishauser et al., 2018) and the exploration of a large proportion of the search space (Figure 1C). Accuracy on the LST decreases linearly as a function of task complexity (Hearne et al., 2017, 2019; Figure S1A) and LST performance correlates with performance on the Raven's Advanced Progressive Matrices (RPM) task (accuracy: r = 0.395, p = 0.016; Figure 1C; reaction time: r = 0.431, p = 0.008; Figure 1D), a standard measure of fluid intelligence (Gray et al., 2003).

By translating trial-locked hemodynamic responses from the LST into the low-dimensional embedding space, we were able to calculate the temporal trajectories of the top 5 PCs (tPC*) that were trained on data from multiple tasks (Shine et al., 2019) in order to determine whether the components were sensitive to alterations in complexity-driven modulations of cognitive load (Birney et al., 2006; Hearne et al., 2019). In this way, the original tasks from the Human Connectome Project (HCP) cohort were able to act in the manner of traditional "control" conditions, allowing us to trace state-space dynamics within a multitask embedding space. Based on our previous work (Hearne et al., 2017; Shine et al., 2019), we hypothesized that tPC^*_1, which tracks with network-level integration (Shine et al., 2019), should relate here to increases in cognitive complexity.

RESULTS

As predicted, modulation of cognitive complexity had a marked effect on low-dimensional flow within the state space (Figure 2A). The dimensionality of the brain state while completing the task

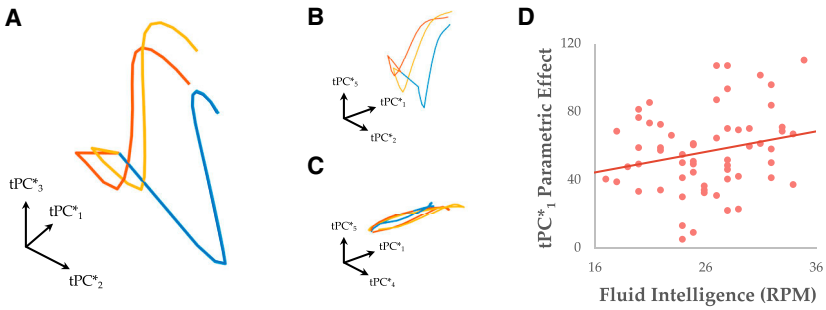

Figure 2. The Effect of Cognitive Complexity on Low-Dimensional Brain Dynamics

(A) State-space embedding of the trajectories (tPC*) of $PC_{1/2/3}$ as a function of increasing cognitive complexity on the LST task. Different lines depict the trajectory associated with the three different levels of task complexity: binary (low complexity), blue; ternary (medium complexity), yellow; and quaternary (high complexity), red.

(B) State-space embedding of tPC^*_1, tPC^*_2, and tPC^*_5, which shows reconfigurations as a function of cognitive complexity.

(C) State-space embedding of tPC^*_1, tPC^*_4, and tPC^*_5, which is relatively invariant with respect to cognitive complexity.

(D) Scatterplot showing significant positive Pearson's correlation (r = 0.229; p = 0.023) between fluid intelligence scores on the RPM and engagement of tPC^*_1 during *Q* trials of the LST (the effect was selective for tPC^*_1; $p < 0.001$).

was inversely correlated with cognitive complexity: that is, increasing difficulty on the LST led to a greater recruitment of tPC^*_1, which in turn explained the most variance in our task (R^2 = 50.10% ± 5.7%). We observed a parametric *increase* in tPC^*_1 (frontal, parietal, occipital cortices, and striatum; 5.9–9.9 s; $p < 10^{-5}$; p value corrected for multiple comparisons across the trial; Maris and Oostenveld, 2007) and tPC^*_5 (motor and frontal cortices; 7.0–9.9 s; $p < 10^{-5}$) and a *decrease* in tPC^*_4 (midline frontoparietal cortices; 8.2–9.9 s; $p < 10^{-5}$; Figure S2) during the reasoning portion of the task. Increasing cognitive complexity also altered the temporal relationship between tPC^*_1 and both tPC^*_2 and tPC^*_3 (Figure 2A). The trajectories were positively correlated at low task complexity (i.e., *B* trials) but became progressively more anti-correlated at higher complexity (i.e., *Q* trials; all $p < 0.001$; Figures 2A and 2B for tPC^*_1, tPC^*_2, and tPC^*_5). Crucially, cognitive complexity did not impact all dimensions within the data (i.e., the combined trajectories of tPC^*_1, tPC^*_4, and tPC^*_5; Figure 2C), suggesting that cognitive complexity forces an increased separation of the relative flow between specific portions of the low-dimensional state space. More broadly, the findings suggest that the brain's capacity to recruit additional functionally specialized regions with increasing task complexity and to integrate the information returned by those regions (Hearne et al., 2017) is tied to its ability to engage (and disengage) an intrinsic low-dimensional spatial mode (Robinson et al., 2016).

Low-dimensional, state-space dynamics were also associated with individual differences in fluid intelligence. Individuals with better scores on the RPM task showed stronger recruitment of tPC^*_1 during the LST (0.0–8.2 s post-task onset; $r \geq 0.3$) and greater disengagement during rest (11.7–15.8; $r \leq -0.3$; Figure 2D). The relationship between tPC^*_1 engagement and RPM held at higher levels of complexity ($r[tPC^*_{1Q}, RPM] = 0.223$, p = 0.043; $r[tPC^*_{1T}, RPM] = 0.259$, p = 0.023); however, the relationship was not significant for *B* trials ($r[tPC^*_{1B}, RPM] = 0.115$; p = 0.190; Figure S2), likely due to ceiling effects (Figure S1B). Interestingly, the extent of impaired task performance as a function of complexity was inversely correlated with tPC^*_1 engagement across subjects (r = −0.233; p = 0.036; Figure S2). That is, individuals that performed the task correctly did not require as much reconfiguration of the manifold as individuals that struggled at higher levels of complexity. Together, these results suggest that low-dimensional, state-space trajectories are sensitive to both state (Figures 2 and S2) and trait behavioral effects (Figure S2).

Next, we determined whether alterations in low-dimensional, state-space dynamics were associated with performance errors on the LST. Based on our previous work (Shine et al., 2019), we hypothesized that task errors should reflect a failure to effectively engage the low-dimensional dynamic flows associated with the successful completion of the task. Examination of the low-dimensional, state-space trajectory across levels of complexity confirmed this hypothesis. Relative to correct trials, task errors were associated with a failure to engage tPC^*_1 (4.6–9.4 s; $p < 0.001$; non-parametric permutation test with correction for multiple comparisons; Maris and Oostenveld, 2007) and tPC^*_5 (8.2–9.4 s; $p < 0.001$) and an inability to appropriately disengage tPC^*_2 (1.7–5.3 s; $p < 0.001$; Figure 3A) relative to successful trials. In addition, there was greater variance as a function of task complexity, and this was more pronounced in error trials than correct trials (Figure S3; $p < 0.001$). To quantify this effect, we calculated the mean Euclidean distance between all *Q* trials (i.e., both correct and error trials) and the mean correct *Q* trajectory, which was created by collapsing across all correct *Q* trials across sessions and participants. Error trials were significantly more likely to deviate from the correct manifold than correct trials (Figure 3B; $p < 0.05$).

A putative benefit of the brain being organized as a low-dimensional system is the ability to utilize the degeneracy inherent within the CNS (Tononi et al., 1999) to facilitate effective behavioral performance. A direct prediction from this framework is that single-subject, low-dimensional, state-space dynamics should relate to individual differences in task performance. To test this hypothesis, we isolated the state-space trajectories for each subject and each level of complexity. We then computed the parametric effect of complexity for each trajectory and then calculated the Euclidean distance that the top 5 PC* trajectories for each subject deviated from the mean group-level correct and confident parametric manifold trajectory (by construction, high performing individuals contributed proportionally more to the resulting manifold than poor performers). Importantly, this approach identifies deviations of the trajectory of the entire brain state and not simply a single PC trajectory, which could potentially be corrected by compensations in other PC dimensions.

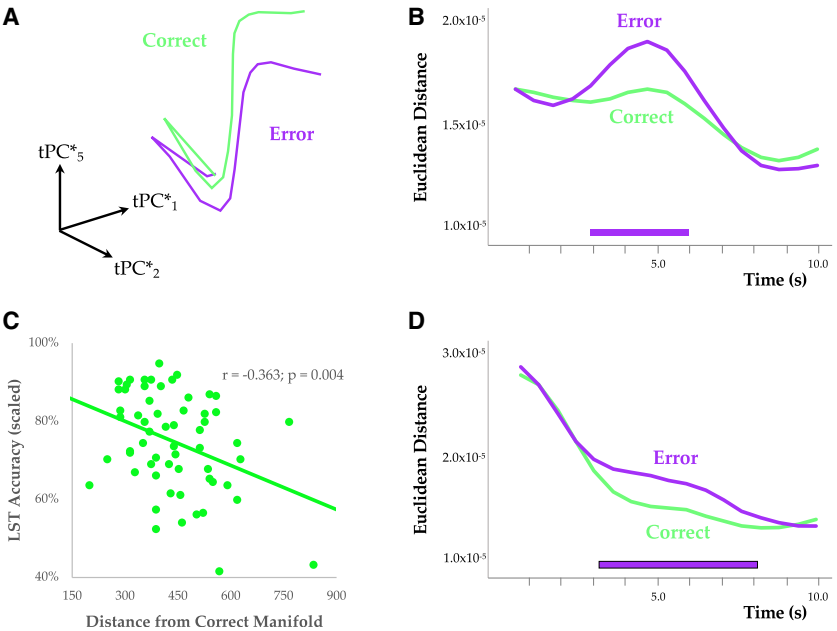

Figure 3. The Effect of Cognitive Errors on Low-Dimensional Brain Dynamics

(A) Mean trajectory of $tPC^*_{1/2/5}$ for correct (green) versus error (purple) quaternary trials.

(B) Mean Euclidean distance of correct (green) and error (purple) trials from the "correct" quaternary manifold—bar depicts $p < 0.05$. Time on the x axis is relative to the start of each trial (as estimated by the FIR model).

(C) Scatterplot comparing LST accuracy (parametrically scaled across levels of complexity) and the distance from the correct group-mean manifold (correct trials; parametrically across levels of complexity). Results showed an inverse Pearson's correlation ($r = -0.363$; $p = 0.004$) between the two measures, indicating that more accurate individuals traversed a state-space trajectory close to the group average trajectory, leading to correct and confident answers.

(D) Restoration to the correct manifold on trials that began a large distance from the correct manifold for correct (green) and error (purple) trials. Bars depict $p < 0.05$ from non-parametric permutation testing.

We then correlated the total deviation (in Euclidean distance) from the correct manifold with the performance accuracy for each subject. We observed a robust, inverse correlation ($r = -0.363$; $p = 0.004$; Figure 3C), suggesting that subjects with greater accuracy on the LST were more likely to follow a state-space trajectory closer to the group average correct trajectory. In other words, the brain states of poor performers on the LST were often found further away from the correct trajectory through state space.

Another hypothesized signature of low-dimensional neural organization is known as "restoration" (Hopfield, 1994), which refers to the capacity of a system to leverage its low-dimensional architecture in order to "restore" noisy deviations from the locations in state space that might otherwise lead to errors in performance. To test for restoration, we identified trials in which the trajectory began a substantial distance (i.e., >66th percentile; defined as above) from the correct manifold (calculated on the remaining trials). There was significantly greater restoration of the low-dimensional trajectory on trials that began "off" the manifold for correct trials (4.1–6.4 s; $p < 0.001$; Figure 3D), but not for error trials. That is, trials that began off the manifold but were restored were more likely to lead to correct performance on the task. Together, these results suggest that cognitive errors can be conceptualized as a failure to maintain an appropriate low-dimensional trajectory in state space.

A plausible candidate for shaping high-dimensional neural activity onto relatively low-dimensional manifolds is the thalamus (Hearne et al., 2017; Sherman, 2004; Figure 4A). Although traditionally viewed as a simple "relay" between specialized regions of the cerebral cortex, the thalamic nuclei also play a crucial role in arousal (Liu et al., 2018; Parvizi and Damasio, 2001; Stitt et al., 2018), memory (Jensen et al., 2002; Van der Werf et al., 2003), attention (Liu et al., 2012; Saalmann and Kastner, 2011), learning (Sarasso et al., 2018), and cognition (Halassa and Kastner, 2017; Hwang et al., 2017). Anatomically, numerous individual thalamic nuclei are known to project to multiple cortical regions (Behrens et al., 2003; Selemon and Goldman-Rakic, 1988). This diverging, few-to-many thalamocortical connectivity pattern led us to predict a strong temporal relationship between thalamic activity and low-dimensional state engagement across the cortex.

To test this hypothesis, we leveraged the high spatial specificity of the 7T fMRI data to extract time series from three distinct groups of thalamic nuclei (medial, posterior, and anterolateral; Figure 4A; see Table S1 for complete list, including those regions excluded due to spatial limitations) that are traditionally associated with separate cognitive functions (Niemann et al., 2000). Although the distinctions are not absolute (Jones, 2001), the anterolateral nuclear group (i.e., ventral anterior and lateral dorsal nuclei) are typically associated with sensorimotor function, the posterior nuclear group (i.e., the pulvinar) with spatial attention, and the medial nuclear group with higher cognitive functions, such as working memory and attentional selection (Golden et al., 2016; Jones, 2001; Zhou et al., 2016). We observed a linear parametric effect of cognitive complexity (i.e., a linear increase across *B*/*T*/*Q* conditions) in the posterior and medial thalamic groups, but not the anterolateral group. Interestingly, the medial and posterior trial-locked finite impulse response (FIR) time series each peaked prior to the engagement of tPC^*_1. However, there was not a consistent pattern observed across subjects ($t = 1.02$; $p = 0.312$), suggesting that the temporal precedence may instead be reflective of individual differences in thalamic and tPC^*_1 engagement across the task. Despite the lack of a consistent intra-subject temporal pattern, we did observe a positive relationship between the timing of the peak of medial and tPC^*_1 parametric effect (Spearman's rho = 0.315; $p = 0.014$), suggesting a stable correspondence across subjects. As such, these results are consistent with findings from invasive recordings of neural activity suggesting that thalamic activity drives activated, gamma-frequency "up" states in the cerebral cortex (Crunelli et al., 2015; McCormick et al., 2015), which in turn protect the cortical state from otherwise distracting thalamic input (Watson et al., 2008).

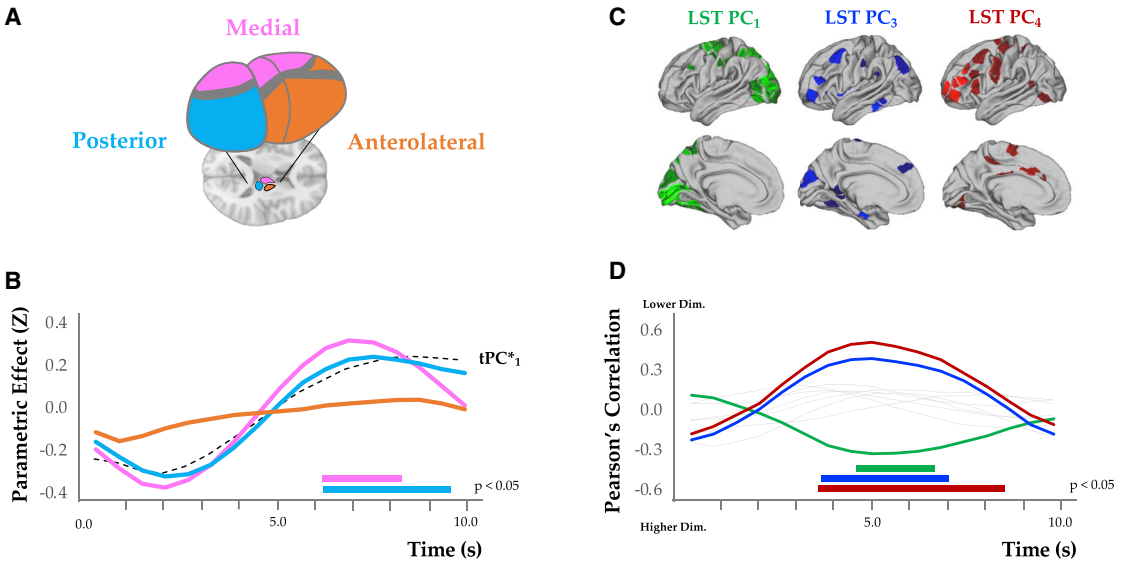

Figure 4. Relationship between the Thalamus and Low-Dimensional Architecture

(A) Three thalamic groups were defined: a medial group (pink), which comprised the parvocellular mediodorsal thalamic nucleus; a posterior group (blue), which comprised four subdivisions of the pulvinar nucleus; and a anterolateral group (orange), which comprised nuclei from the anterior, lateral, and ventral nuclei of the thalamus—left: axial slice through the brain (Montreal Neurological Institute [MNI] 1-mm template; Z = 10); right: 3-dimensional rendering of the approximate spatial locations of the three groups.

(B) Parametric effect across complexity for each nuclear group—colored bars designate significant parametric effects ($p < 0.05$ from non-parametric permutation testing) and the dotted black (tPC^*_1) line depicts mean low-dimensional trajectories.

(C) Spatial maps of PCs 1, 3, and 4 fit to the LST task data—for clarity, only positive loadings for single hemisphere are shown (patterns were symmetrical across hemispheres).

(D) Pearson's correlation across participants between the parametric effect of complexity and the dimensionality of the BOLD data (defined as the % variance explained by the $PC_{1–10}$, estimated across 333 cortical parcels). Across participants, activity in the medial group was selectively associated with a decrease in the explained variance of PC_1 (green) and an increase in the explained variance of PC_3 (dark blue) and PC_4 (dark red). Colored bars designate a significant parametric effect ($p < 0.05$; based on non-parametric permutation testing). No other PCs showed a significant effect (gray lines).

These observations raise an interesting question: how does the brain collapse activity onto a low-dimensional manifold while retaining sufficient informational complexity to solve a challenging cognitive task, such as the LST? The thalamus has a unique topological architecture, which may help to strike an optimal balance between globally integrated and locally segregated cortical states (Bell and Shine, 2016; Hwang et al., 2017; Sherman, 2004). Previous work in vision science has suggested that reducing the dimensionality of neural signals should be balanced against the retention of the maximal amount of information (DiCarlo and Cox, 2007; Cayco-Gajic and Silver, 2019), an algorithmic capacity for which the thalamus is ideally suited. In keeping with this view, we predicted that the thalamus should promote integration by fostering connectivity between diverse cortical regions required to perform the LST (Hearne et al., 2017) while simultaneously ensuring that the low-dimensional architecture of the brain retains the capacity for the flexible comparison of different potential solutions during the reasoning portion of the task.

To test this hypothesis, we compared individual differences in thalamic recruitment with cortical dimensionality across our cohort of 60 participants. A new principal-component analysis (PCA) was fit to cortical FIR data at the individual subject level, and the percentage of variance explained by the first PC (fit on whole brain data while holding out thalamic time series) was used as an estimate of the dimensionality of the cortical state across the task (Garrett et al., 2013), where high values denote a relatively low-dimensional architecture (and vice versa). Individual recruitment of the medial thalamic group was inversely associated with the variance explained by the first PC across cognitive complexity (Figure 4C; 4.1–8.1 s; $p < 0.05$; Figure S4; STAR Methods). This first PC loaded onto superior frontal, superior parietal, and occipital cortical regions (green in Figure 4C) and was positively correlated with PC_1 from the HCP dataset ($r = 0.744$; $p = 2.2 \times 10^{-67}$). In contrast, individual thalamic recruitment was positively associated with the variance explained by the third and fourth PCs (Figure 4D), which loaded onto more lateral frontoparietal (blue) and frontopolar (red) cortical regions, respectively. In other words, the thalamus appears to play a crucial role in fostering low-dimensionality (e.g., recruiting tPC^*_1; Figure 4B) while preserving sufficient dimensionality to complete the task (e.g., inverse correlation with variance explained in the first PC and positive correlation with variance explained in $PC_{3/4}$; Figure 4C), particularly in lateral frontoparietal control regions.

DISCUSSION

Our findings contribute to a rapidly growing literature on the relationship between low-dimensional network dynamics and

the functional signature of whole-brain network activity (Cunningham and Yu, 2014; Mante et al., 2013). Specifically, we have shown that engagement of a low-dimensional manifold facilitates the amplification of higher-order cognitive processes necessary for integrating multiple elements in a complex relational reasoning task. Our framework also allows the extension of well-known concepts in cognitive neuroscience (such as the functional antagonism between the frontoparietal and default mode network; Fox et al., 2005) to the generation of novel mechanistic hypotheses on brain dynamics supporting cognition. Crucially, this reframing allows us to track the primary engagement and disengagement of the task-positive and default mode networks (i.e., with the activity within tPC^*_1) along with the coordinated temporal interactions present among higher latent dimensions. Importantly, these patterns, which also covary with cognitive complexity (see Figure 2), cannot be captured through the lens imposed by traditional voxel-wise analyses of brain imaging data. We were also able to show that not all degrees of freedom in the low-dimensional embedding are modulated by cognitive complexity (e.g., the trajectories between $tPC^*_{1/4/5}$ were relatively oblivious to complexity), suggesting that cognitive complexity leads to distinct reconfigurations in state space dynamics. Moreover, our results extend previous work in humans (Garrett et al., 2018) and non-human primates (Baxter, 2013) by demonstrating that medial thalamic nuclei play a key mediating role in the control of distributed cortical processing during high-level cognitive operations.

The current study opens a number of research directions for future work. First, whole-brain computational models should pay more heed to the role of thalamic nuclei than their typical focus on cortical dynamics (Breakspear, 2017). In particular, modulating large-scale brain dynamics through changes in modeled thalamic gain could offer important insights into the sculpting of the low-dimensional flow according the relational complexity seen here. Second, causally manipulating cortical brain activity (e.g., with pharmacology or stimulation) and investigating alterations in low-dimensional functional architecture will also elucidate the neural factors that instantiate balanced amplification in the brain (Stringer et al., 2016).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2019.09.002.

ACKNOWLEDGMENTS

J.M.S. was supported by a University of Sydney Robinson Fellowship and the National Health and Medical Research Council (1156536). L.C. was supported by the Australian National Health Medical Research Council (1099082 and 1138711).

AUTHOR CONTRIBUTIONS

L.J.H., J.B.M., and L.C. planned the study and collected the data. L.J.H. preprocessed the data. J.M.S. conducted the analysis and wrote the first draft of the manuscript. All other authors provided critical feedback and synthesis of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 2, 2019
Revised: July 14, 2019
Accepted: September 4, 2019
Published: October 22, 2019

REFERENCES

Baxter, M.G. (2013). Mediodorsal thalamus and cognition in non-human primates. Front. Syst. Neurosci. *7*, 38.

Behrens, T.E.J., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A.M., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., et al. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. *6*, 750–757.

Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage *37*, 90–101.

Bell, P.T., and Shine, J.M. (2016). Subcortical contributions to large-scale network communication. Neurosci. Biobehav. Rev. *71*, 313–322.

Birney, D.P., Halford, G.S., and Andrews, G. (2006). Measuring the influence of complexity on rational reasoning - The development of the latin square task. Educ. Psychol. Meas. *66*, 146–171.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. Nat. Neurosci. *20*, 340–352.

Cayco-Gajic, N.A., and Silver, R.A. (2019). Re-evaluating circuit mechanisms underlying pattern separation. Neuron *101*, 584–602.

Chialvo, D.R. (2010). Emergent complex neural dynamics. Nat. Phys. *6*, 744–750.

Crunelli, V., David, F., Lőrincz, M.L., and Hughes, S.W. (2015). The thalamocortical network as a single slow wave-generating unit. Curr. Opin. Neurobiol. *31*, 72–80.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509.

DiCarlo, J.J., and Cox, D.D. (2007). Untangling invariant object recognition. Trends Cogn. Sci. *11*, 333–341.

Ercsey-Ravasz, M., and Toroczkai, Z. (2012). The chaos within Sudoku. Sci. Rep. *2*, 725.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., and Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. USA *102*, 9673–9678.

Garrett, D.D., Samanez-Larkin, G.R., MacDonald, S.W., Lindenberger, U., McIntosh, A.R., and Grady, C.L. (2013). Moment-to-moment brain signal variability: a next frontier in human brain mapping? Neurosci. Biobehav. Rev. *37*, 610–624.

Garrett, D.D., Epp, S.M., Perry, A., and Lindenberger, U. (2018). Local temporal variability reflects functional integration in the human brain. Neuroimage *183*, 776–787.

Golden, E.C., Graff-Radford, J., Jones, D.T., and Benarroch, E.E. (2016). Mediodorsal nucleus and its multiple cognitive functions. Neurology *87*, 2161–2168.

Gordon, E.M., Laumann, T.O., Adeyemo, B., Huckins, J.F., Kelley, W.M., and Petersen, S.E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb. Cortex *26*, 288–303.

Gray, J.R., Chabris, C.F., and Braver, T.S. (2003). Neural mechanisms of general fluid intelligence. Nat. Neurosci. *6*, 316–322.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q.K., Yu, A.B., Kahn, A.E., Medaglia, J.D., Vettel, J.M., Miller, M.B., Grafton, S.T., and Bassett, D.S. (2015). Controllability of structural brain networks. Nat. Commun. *6*, 8414.

Halassa, M.M., and Kastner, S. (2017). Thalamic functions in distributed cognitive control. Nat. Neurosci. *20*, 1669–1679.

Halford, G.S., Cowan, N., and Andrews, G. (2007). Separating cognitive capacity from knowledge: a new hypothesis. Trends Cogn. Sci. *11*, 236–242.

Hearne, L.J., Cocchi, L., Zalesky, A., and Mattingley, J.B. (2017). Reconfiguration of brain network architectures between resting-state and complexity-dependent cognitive reasoning. J. Neurosci. *37*, 8399–8411.

Hearne, L.J., Birney, D.P., Cocchi, L., and Mattingley, J.B. (2019). The Latin Square Task as a measure of relational reasoning: a replication and assessment of reliability. Eur. J. Psychol. Assess. Published online May 10, 2019. https://doi.org/10.1027/1015-5759/a000520.

Hopfield, J.J. (1994). Physics, computation, and why biology looks so different. J. Theor. Biol. *171*, 53–60.

Huys, R., Perdikis, D., and Jirsa, V.K. (2014). Functional architectures and structured flows on manifolds: a dynamical framework for motor behavior. Psychol. Rev. *121*, 302–336.

Hwang, K., Bertolero, M.A., Liu, W.B., and D'Esposito, M. (2017). The human thalamus is an integrative hub for functional brain networks. J. Neurosci. *37*, 5594–5607.

Jensen, O., Gelfand, J., Kounios, J., and Lisman, J.E. (2002). Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex *12*, 877–882.

Jones, E.G. (2001). The thalamic matrix and thalamocortical synchrony. Trends Neurosci. *24*, 595–601.

Kerkman, J.N., Daffertshofer, A., Gollo, L.L., Breakspear, M., and Boonstra, T.W. (2018). Network structure of the human musculoskeletal system shapes neural interactions on multiple time scales. Sci. Adv. *4*, eaat0497.

Liu, Z., de Zwart, J.A., Yao, B., van Gelderen, P., Kuo, L.W., and Duyn, J.H. (2012). Finding thalamic BOLD correlates to posterior alpha EEG. Neuroimage *63*, 1060–1069.

Liu, Z., de Zwart, J.A., Schölvinck, M.L., Chang, C., Ye, F.Q., Leopold, D.A., and Duyn, J.H. (2018). Subcortical evidence for a contribution of arousal to fMRI studies of brain activity. Nat. Commun *9*, 395.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

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

McCormick, D.A., McGinley, M.J., and Salkoff, D.B. (2015). Brain state dependent activity in the cortex and thalamus. Curr. Opin. Neurobiol. *31*, 133–140.

Niemann, K., Mennicken, V.R., Jeanmonod, D., and Morel, A. (2000). The Morel stereotactic atlas of the human thalamus: atlas-to-MR registration of internally consistent canonical model. Neuroimage *12*, 601–616.

Parvizi, J., and Damasio, A. (2001). Consciousness and the brainstem. Cognition *79*, 135–160.

Robinson, P.A., Zhao, X., Aquino, K.M., Griffiths, J.D., Sarkar, S., and Mehta-Pandejee, G. (2016). Eigenmodes of brain activity: Neural field theory predictions and comparison with experiment. Neuroimage *142*, 79–98.

Rutishauser, U., Slotine, J.-J., and Douglas, R.J. (2018). Solving constraint-satisfaction problems with distributed neocortical-like neuronal networks. Neural Comput. *30*, 1359–1393.

Saalmann, Y.B., and Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. Neuron *71*, 209–223.

Sarasso, E., Agosta, F., Temporiti, F., Adamo, P., Piccolo, F., Copetti, M., Gatti, R., and Filippi, M. (2018). Brain motor functional changes after somatosensory discrimination training. Brain Imaging Behav. *12*, 1011–1021.

Selemon, L.D., and Goldman-Rakic, P.S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. J. Neurosci. *8*, 4049–4068.

Sherman, S.M. (2004). Interneurons and triadic circuitry of the thalamus. Trends Neurosci. *27*, 670–675.

Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K.A., Shine, R., Koyejo, O., Sporns, O., and Poldrack, R.A. (2019). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat. Neurosci. *22*, 289–296.

Stitt, I., Zhou, Z.C., Radtke-Schuller, S., and Fröhlich, F. (2018). Arousal dependent modulation of thalamo-cortical functional interaction. Nat. Commun. *9*, 2455.

Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., and Lesica, N.A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. eLife *5*, 91.

Tononi, G., Sporns, O., and Edelman, G.M. (1999). Measures of degeneracy and redundancy in biological networks. Proc. Natl. Acad. Sci. USA *96*, 3257–3262.

Van der Werf, Y.D., Scheltens, P., Lindeboom, J., Witter, M.P., Uylings, H.B.M., and Jolles, J. (2003). Deficits of memory, executive functioning and attention following infarction in the thalamus; a study of 22 cases with localised lesions. Neuropsychologia *41*, 1330–1344.

Watson, B.O., MacLean, J.N., and Yuste, R. (2008). UP states protect ongoing cortical activity from thalamic inputs. PLoS ONE *3*, e3971.

Zhou, H., Schafer, R.J., and Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. Neuron *89*, 209–220.

Please cite this article in press as: Shine et al., The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei, Neuron (2019), https://doi.org/10.1016/j.neuron.2019.09.002

Q3 Q4 STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Trajectories of top 5 cortical PCs	7T MRI scanner	beta_score_top5.mat
Time series of 3 thalamic groups	7T MRI scanner	thal_score1.mat

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, James M. Shine (mac.shine@sydney.edu.au).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Sixty-five healthy, right-handed adult participants (18-33 years) were recruited, of whom 60 were included in the final analysis (28 females). Participants provided informed written consent to participate in the study. The behavioral and neuroimaging data were reported in an earlier study (Hearne et al., 2017), but that study focused on a different question to the one addressed here. The research was approved by The University of Queensland Human Research Ethics Committee.

METHOD DETAILS

Summary

High spatiotemporal resolution 7T BOLD data were collected while participants performed a cognitively-challenging task (the Latin Squares Task, LST) across different levels of complexity (Birney et al., 2006; Hearne et al., 2017). By translating trial-locked haemodynamic responses from the LST into a previously-defined low-dimensional embedding space (Shine et al., 2019), we were able to track the temporal trajectories of the top-5 PCs (tPC*) as a function of cognitive complexity. Significance testing involved the comparison of each tPC* trajectory across cognitive complexity using ANOVAs and subsequent pairwise t tests at each TR. We subsequently estimated trial-locked haemodynamic time series from pre-defined thalamic nuclear groups, and then contrasted these time series with the trajectories of the low-dimensional states.

Latin Squares Task

Participants completed a modified version of the LST (Hearne et al., 2017), which is a nonverbal relational reasoning paradigm in which reasoning complexity is parametrically varied with minimal working-memory demands. 144 unique LST items were presented across three MR sessions (16 blocks), with 36 items in each relational complexity condition pseudo-randomly presented across four different cognitive complexities (Figure 1B). Binary ('*B*') problems required integration of information across a single row or column; Ternary ('*T*') problems involved integration across a single row and column; Quaternary ('*Q*') problems required integration of information across multiple rows and columns; Null trials (which were not analyzed here) involved presentation of an LST grid, but with an asterisk (*) instead of a question mark. After a 4,000 ± 586 msec cross-hair, each puzzle was presented for 5,000 msec. After a 1,000 ± 586 msec delay, participants were given 2,000 msec to manually designate a response using a four-button keypad. Participants were instructed to solve for the target according to the following rule: Each shape can only occur once in every row and once in every column (similar to the game 'Sudoku'). Motor responses were counterbalanced across individuals, such that equal numbers of participants had the same shape-response mapping. Five-point Likert-scale confidence ratings were collected (Hearne et al., 2017), and only trials with high confidence (trials rated 4 or 5 on the scale) were used in the final analysis (for both Correct and Error trials). Although the LST required a search over a greater proportion of the task grid as a function of Complexity, previous work suggests that eye movements were constrained to the same relative space across levels of complexity (Hearne et al., 2017).

Neuroimaging acquisition

1250 (~12 minutes) whole-brain 7T task-fMRI echo planar images were acquired using a multiband sequence (acceleration factor = 5) in the each of the three runs of the task (2mm^3 voxels; 586ms TR; 23ms TE; 40° flip angle; 208mm FOV; 55 slices). Structural images were also collected to assist functional data preprocessing (MP2RAGE sequence – 0.75mm^3 voxels 4300ms TR; 3.44ms TE; 256 slices).

Data pre-processing

Imaging data were preprocessed using custom MATLAB scripts. DICOM (Digital Images and Communications in Medicine) images were first converted to Nifti format and realigned. T1 images were reoriented, skull-stripped (FSL BET), and co-registered to the Nifti functional images using statistical parametric mapping (SPM8) functions. Segmentation and the DARTEL algorithm were used to improve the estimation of non-neural signal in subject space and the spatial normalization. From each gray matter voxel, the following signals were regressed: linear trends, signals from the six head-motion parameters (three translation, three rotation) and their temporal derivatives, white matter, and CSF (estimated from single-subject masks of white matter and CSF). The aCompCor method (Behzadi et al., 2007) was used to regress out residual signal unrelated to neural activity (i.e., five principal components derived from noise regions-of-interest in which the time series data were unlikely to be modulated by neural activity). Participants with head displacement > 3 mm in > 5% of volumes in any one scan were excluded (n = 5). No temporal filtering was applied to the data. Following pre-processing, the mean time series was extracted from 333 pre-defined cortical regions-of-interest (ROI) using the Gordon atlas (Gordon et al., 2016).

Finite Impulse Response

For each participant/session, a unique design matrix was implemented which modeled each trial (grouped according to Complexity [i.e., *B*, *T* or *Q*; and Response [i.e., Correct or Error]) as a Finite Impulse Response (i.e., with unique regressors for each TR within a trial), repeated across instances of each trial type. This approach allowed us to model the response of each region without specifying a canonical haemodynamic response function across all regions and hence, augmented individual regional differences in haemodynamic responses prior to converting the data into low-dimensional trajectories. Based on the predicted peak response for the active task component of the LST (which we estimated by convolving a 5,000 msec boxcar function with a canonical HRF; TR = 586 msec), only the first 17 TRs (~9.9 s) of each trial type were investigated (Hearne et al., 2017).

Low-dimensional state space embedding

The low-dimensional signature of a task is often related to relatively idiosyncratic aspects of the particular task, and hence can remain relatively insensitive to the components that recur across tasks (Shine et al., 2019). To avoid this issue, we estimated the trajectory of a set of five pre-defined principal components from a previous study that estimated the low-dimensional functional architecture of the brain across a diverse set of cognitive tasks (Shine et al., 2019). Specifically, the trajectory of each PC was estimated by calculating the weighted mean of the correspondence between the regional FIR β for each time point across all conditions and the spatial eigenvector associated with the top 5 PCs from our original study (Shine et al., 2019). Importantly, we chose to fit the PCs from our original study (Shine et al., 2019) to this data so as to avoid circularity – that is, if we defined the PCs in this dataset (which was designed to manipulate cognitive complexity), then by definition, we would be likely discover components that accordingly split variance in the data caused by this parametric manipulation. Instead, by projecting these data into an independent space defined by other tasks, we were able to create a stronger validation of the complexity hypothesis. To confirm that these trajectories were distinct from those estimated directly from a PCA of the LST, we conducted a PCA on group-level data concatenated across sessions/trials (1^{st} 5 PCs explained 86.95% ± 13.3 of the variance). While the first spatial PC was similar between the two approaches ($r_{PC1} = 0.76$), the subsequent PCs were substantially less similar (mean $r_{PC2-5} = 0.17$).

Cognitive Modulation of the Manifold

Low-dimensional embeddings were visualized by projecting the mean state space trajectory for correct *B*, *T* and *Q* trials across participants onto a 3D scatterplot (e.g., Figure 2A). Complexity effects were estimated by parametrically modulating the trajectories (i.e., *B*1 + T*2 + Q*3*) and comparing the data to a permuted null model in which regional signatures of the PC loadings were randomly shuffled prior to estimating each tPC* trajectory (5,000 iterations; $p < 0.05$). Significance testing involved the comparison of each tPC* trajectory across cognitive complexity using ANOVAs and subsequent pairwise t tests at each TR. tPC* trajectories from Correct and Error trials from *T* and *Q* trials were pooled and compared statistically using paired t tests at each TR (there were insufficient Errors in the *B* trials; Figure S1B).

To estimate whether the trajectory of correct and error trials deviated as a function of cognitive complexity, we calculated the mean Euclidean distance (absolute value of deviation between tPC* trajectories at each TR) between each trajectory. This approach allowed us to determine whether individual high-dimensional trajectories deviated from the mean correct manifold in ways that may have been lost by examining each tPC* trajectory in turn. We also estimated the standard deviation of each tPC* trajectory across correct and error trials and compared them statistically using paired t tests (with correction from multiple comparisons).

To determine whether individual differences in performance accuracy on the LST related to specific low-dimensional trajectories across state space, we separated the top 5 PC* trial-locked time series for each subject at each level of complexity and calculated the parametric effect of complexity at each point of the task. We next computed the Euclidean distance of the top 5 PC* trajectories from the mean correct parametric trajectory for each subject (defined as the group mean, correct and confident trajectory at each level of complexity). We correlated the sum of the deviations (in Euclidean distance) from the mean correct manifold over the whole trial with the performance accuracy for each subject using a Pearson's correlation. Finally, we calculated the parametric effect of performance

decrements as a function of cognitive complexity by multiplying the drop in accuracy from *T* to *Q* trials (*t2) and adding this to the drop in accuracy from *B* to *T* trials. We then correlated this value, across subjects, with the engagement of tPC^*_1 (parametrically across levels of complexity).

To test the prediction that low-dimensional trajectories should demonstrate evidence of 'restoration' (Hopfield, 1994), we identified the trajectories associated with the top 33.3% of deviations from the manifold in the first 5 TRs for both Correct and Error trials (separately) and then used a series of independent samples t tests to determine whether the trajectory of the subsequently correct trials more closely mapped the correct manifold over the course of trials. For all tests, statistical significance was defined using non-parametric permutation tests. In addition, we corrected for multiple comparisons using an approach that accounts for temporal smoothness in the data (Maris and Oostenveld, 2007).

Raven's Advanced Progressive Matrices

Outside of the MRI scanner, 37 participants completed the Raven's Advanced Progressive Matrices (RPM; 40 min time limit), which is a standard and widely used measure of fluid intelligence (Gray et al., 2003). Accuracy on the RPM was then compared with accuracy (Figure 1D) and reaction time (Figure 1E) on the LST, and with the low-dimensional state-space trajectories at the subject level using a series of Pearson's correlations (one for each trial time point). We also correlated performance on the RPM with the engagement of tPC^*_1 (Area Under the Curve) across levels of complexity (Figure S2).

Thalamic Time Series

Mean regional time series were extracted from a set of 31 pre-defined thalamic nuclei from the Morel atlas template (Niemann et al., 2000). Based on the limited spatial resolution of BOLD data (i.e., 2 mm^3 isotropic voxels), we further collapsed 17 of these regions into three distinct groups: Medial group, comprising the parvocellular mediodorsal nucleus; Posterior group, comprising four pulvinar nuclei (lateral, inferior, anterior and medial) and Anterolateral group, comprising twelve regions from anterior, lateral and ventral thalamus (Figures 4A and S4). Due to the aforementioned resolution limit, we were unable to model the lateral and medial geniculate nuclei, or the intralaminar nuclei, with sufficient detail, and they were therefore excluded from further analysis. Following time series extraction, an FIR analysis was conducted for each regional time series, and the data were then plotted and compared with tPC* trajectories (Figure 4B). Complexity effects were estimated by parametrically modulating the trajectories (i.e., *B*1 + T*2 + Q*3*) and comparing the data to a permuted null model in which regional signatures of the PC loadings were randomly shuffled prior to estimating each tPC* trajectory (5,000 iterations; $p < 0.05$).

To test the prediction that thalamic recruitment preserves relative cortical dimensionality, we ran a separate PCA for each individual on the cortical GLM fits across *B*, *T* and *Q* trials after first re-scaling the BOLD data to have zero mean and unit variance. We then used a series of Pearson's correlations to compare the percentage of variance explained by the first PC (60 participant vector) with the parametric effect of task complexity at each time point for the Medial, Posterior and Anterolateral thalamic groups in turn (60 participants x 3 thalamic groups x 17 TRs). Statistical analysis was conducted using a non-parametric permutation approach, in which we scrambled the first PC 5,000 times while keeping the thalamic time series constant and re-ran the analysis to populate a null distribution. Values greater than (or less than) the 97.5^{th} (2.5^{th}) percentile were deemed significant.

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance testing involved the comparison of each tPC* trajectory across cognitive complexity using ANOVAs and subsequent pairwise t tests at each TR. tPC* trajectories from Correct and Error trials from *T* and *Q* trials were pooled and compared statistically using paired t tests at each TR (there were insufficient Errors in the *B* trials; Figure S1B). We also estimated the standard deviation of each tPC* trajectory across correct and error trials and compared them statistically using paired t tests (with correction from multiple comparisons). For all tests, statistical significance was defined using non-parametric permutation tests. In addition, we corrected for multiple comparisons using an approach that accounts for temporal smoothness in the data (Maris and Oostenveld, 2007).

DATA AND CODE AVAILABILITY

Data and code are openly available at https://github.com/macshine/lst_statespace.