Distributed Working Memory in a Computational Model of the Human Brain

Rares A. Dorcioman (r.a.dorcioman@uva.nl)
Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands

Mengli Feng (mengli.feng@student.uva.nl)
Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands

Abhirup Bandyopadhyay (abhirup.bandyopadhyay@ki.se)
Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands

Jorge F. Mejias (j.f.mejias@uva.nl)
Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands

Abstract

Working memory is a fundamental cognitive function that allows us to transiently store and manipulate relevant information in memory. While traditionally associated with localized prefrontal activity, recent electrophysiological and imaging studies reveal distributed activity across multiple brain regions. To uncover the mechanisms behind this distribution, we developed a detailed, data-constrained model of the human brain by integrating diverse large-scale datasets. Our model demonstrates that distributed working memory patterns emerge primarily through long-range synaptic projections rather than solely from local recurrent connectivity. We found that the network operates optimally near a critical region at the edge of a bifurcation, consistent with recent experimental and modeling findings, and explains approximately 60% of the observed variability among brain areas involved in working memory. Furthermore, simulations of task-specific conditions, such as verbal and spatial working memory, indicate that high agreement with experimental data is achieved only when higher cortical regions modulate the network or when recurrent connectivity is enhanced across multiple circuits. These results suggest that working memory performance is the product of an interplay between distributed projections and context-dependent modulation, offering new insights into the neural substrates of human cognition.

Keywords: working memory; dynamics; modeling; networks;

Introduction

Working memory is a flexible cognitive function that enables temporary storage and manipulation of sensory information, underpinning tasks like decision making, reasoning, and learning. Moreover, its dysfunction has been implicated in various psychotic disorders, such as schizophrenia (Forbes, Carrick, McIntosh, & Lawrie, 2009), hence uncovering its neural substrates could advance efforts to find better treatments.

Human neuroimaging, alongside systematic multi-region recordings in non-human primates (Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017; Leavitt, Mendoza-Halliday, & Martinez-Trujillo, 2017), and rodents (Voitov & Mrsic-Flogel, 2022), has revealed that working memory engages a distributed network of brain regions. Moreover, distinct working memory tasks appear to recruit different neural circuits, suggesting a more complex, regionally varied process (Volle et al., 2005; Cohen et al., 1997). Although computational models have begun to propose mechanisms for distributed working memory in non-human primates (Mejías & Wang, 2022) and rodents (Ding, Froudist-Walsh, Jaramillo, Jiang, & Wang, 2024), the underlying processes in the human brain remain largely unexplored. This emerging perspective calls for integrative approaches to better understand how diverse brain regions coordinate to support working memory.

Methods

To build a detailed, data-constrained computational model of the human brain and explore distributed working memory, we integrated three complementary datasets: (i) a parcellation-based structural connectome using the 'Schaefer 200' cortical parcellation (Jung, Eickhoff, & Popovych, 2022; Schaefer et al., 2018) (ii) intracortical myelin content measured via T1w:T2w maps to infer hierarchical relationships between areas (Glasser et al., 2016; Demirtaş et al., 2019), and (iii) a PET-derived brain map of NMDA receptor density (Hansen et al., 2022). This approach allowed us to constrain our model at multiple levels. The model comprises a network of 100 cortical areas from the left hemisphere, interconnected by projections whose strength is based on previous human connectivity data (Jung et al., 2022; Schaefer et al., 2018) and modulated by a global coupling parameter (G). Each area is represented by a simplified circuit model (Wong & Wang, 2006) featuring two excitatory populations (selective for visual stimuli A and B) and one inhibitory population (Figure 1A). We introduced directional connections using intracortical myelin data: feedforward projections preferentially excite target areas, while feedback projections slightly bias inhibitory populations (parameter α). Area-specific heterogeneity was introduced using NMDA receptor density data (Hansen et al., 2022) to modulate local recurrent excitation (Figure 1B).

Results

To simulate a generic working memory task, a one-second cue activated excitatory population A in early visual cortex, and during the delay period, activity spread to frontal, temporal, parietal, and occipital regions (Figure 1C, D). Rather than relying on local attractor dynamics (Wang, 1999; Compte, 2006), distributed working memory emerged via long-range interactions. The model produced a nonlinear activity profile: about 35 areas maintained only spontaneous firing, while roughly 40 areas sustained firing rates above 15 spikes per second (Figure 1E). This resulted in heterogeneous dynamics, with early sensory areas rapidly decaying after the cue, whereas regions such as 'Default PFC 10' and 'Default PFC 6' maintained persistent, stimulus-selective activity (Figure 1F).

Such activity patterns emerge when global coupling strength and feedback inhibition are finely balanced, positioning the system at a critical threshold between over-excitation and low network activity (Figure 2). The apparent gap in the parameter space represents states in which the activity is dominated by the population non-selective to the presented stimulus (Pop B), in a winner-takes-all regime.

To reproduce neural activity observed in task-specific experiments (e.g. verbal and spatial working memory), we had to either increase local self-excitation, hence inducing local bistability, or to introduce top-down modulation via a GO signal presented to higher cortical areas.

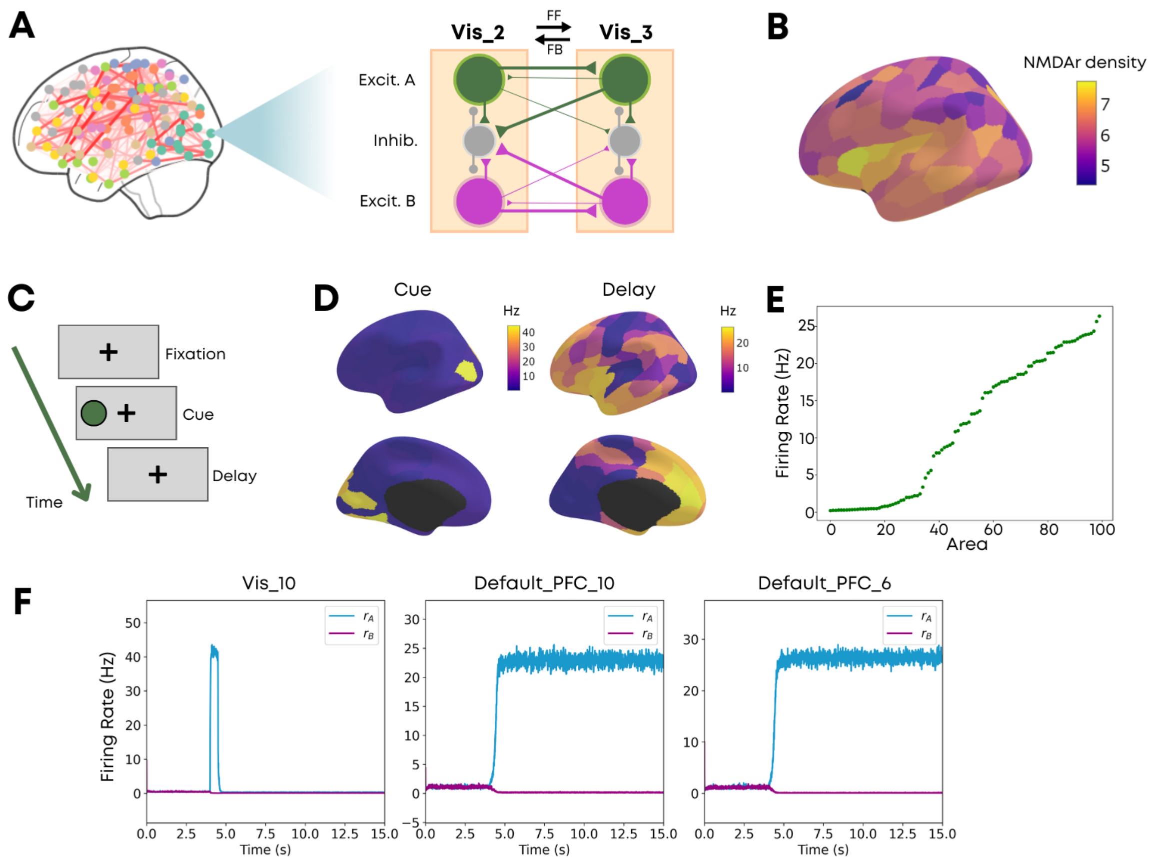

Figure 1: Distributed working memory patterns emerge in a human brain model. (A) Human brain connectome and local model used. (B) Maps of cortical NMDA receptor density. (C) Working memory task employed in our study. (D) Brain activity maps during task. (E) Firing rates for each area in the model during the delay period. (F) Temporal evolution of selective firing rates. A visual stimulus, for which the excitatory population A is selective, is received for 1 second.

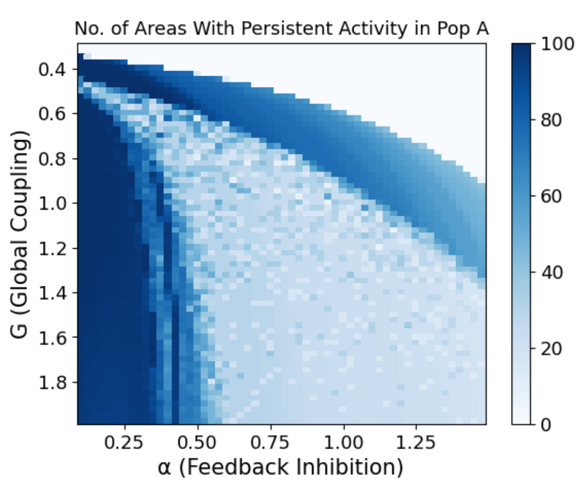

Figure 2: Trials-averaged number of areas showing persistent activity in the stimulus-selective population (Pop A) as a function of the global connection strength (G) and the level of feedback inhibition (α).

Discussion

Our model allows, for the first time, to integrate biologically plausible full-brain dynamics with mechanistic functionalities linked to a cognitive function in humans. It suggests that working memory depends on long-range projections, complemented by context-dependent modulatory signals and task-specific neural circuits. The present study also provides predictions that could be tested experimentally to investigate the neurobiology of working memory. We foresee that this new type of models will play an important role in our understanding of distributed cognitive processes in humans.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101137289 (Virtual Brain Twin Project).

References

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. , *21*(2), 111–124.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. , *386*(6625), 604–608.

Compte, A. (2006). Computational and in vitro studies of persistent activity: edging towards cellular and synaptic mechanisms of working memory. , *139*(1), 135–151.

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., ... Murray, J. D. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. , *101*(6), 1181–1194.e13.

Ding, X., Froudist-Walsh, S., Jaramillo, J., Jiang, J., & Wang, X.-J. (2024). Cell type-specific connectome predicts distributed working memory activity in the mouse brain. , *13*, e85442.

Forbes, N. F., Carrick, L. A., McIntosh, A. M., & Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. , *39*(6), 889–905.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. , *536*(7615), 171–178.

Hansen, J. Y., Shafiei, G., Markello, R. D., Smart, K., Cox, S. M. L., Nørgaard, M., ... Misic, B. (2022). Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. , *25*(11), 1569–1581.

Jung, K., Eickhoff, S. B., & Popovych, O. V. (2022). *Parcellation-based structural and resting-state functional whole-brain connectomes of 1000brains cohort (v1.1).* EBRAINS.

Leavitt, M. L., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2017). Sustained activity encoding working memories: Not fully distributed. , *40*(6), 328–346.

Mejías, J. F., & Wang, X.-J. (2022). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. , *11*, e72136.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., ... Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. , *28*(9), 3095–3114.

Voitov, I., & Mrsic-Flogel, T. D. (2022). Cortical feedback loops bind distributed representations of working memory. , *608*(7922), 381–389.

Volle, E., Pochon, J. B., Lehéricy, S., Pillon, B., Dubois, B., & Levy, R. (2005). Specific cerebral networks for maintenance and response organization within working memory as evidenced by the 'double delay/double response' paradigm. , *15*(7), 1064–1074.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. , *19*(21), 9587–9603.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. , *26*(4), 1314–1328.