Hippocampal versus cortical language processing: Similar selectivity profile, difference in tuning axes

Elizabeth Jiachen Lee (jelizlee@mit.edu)
Dept of Brain and Cognitive Sciences, Massachusetts Institute of Technology,
43 Vassar Street Cambridge, MA 02139 USA

Idan A. Blank (iblank@psych.ucla.edu)
Dept of Psychology, University of California Los Angeles,
Los Angeles, CA 90095 USA

Evelina Fedorenko* (evelina9@mit.edu)
Dept of Brain and Cognitive Sciences, Massachusetts Institute of Technology,
43 Vassar Street Cambridge, MA 02139 USA

Greta Tuckute* (gretatu@mit.edu)
Dept of Brain and Cognitive Sciences, Massachusetts Institute of Technology,
43 Vassar Street Cambridge, MA 02139 USA

Abstract:

The role of the medial temporal lobe structures, including the hippocampus, in language processing remains largely unknown. In a large-scale fMRI dataset (>600 participants), we identified a language-responsive region (LangHippoc) in the anterior left hippocampus. This region responds to meaningful linguistic inputs and is engaged in semantic processing but is not engaged during cognitively demanding spatial memory and arithmetic tasks. Critically, by performing an encoding-model-guided search procedure on another fMRI dataset of responses to 1,000 diverse sentences, we searched for sentences that maximally differentiate responses in LangHippoc and the cortical language network (LangCortex). We found that the tuning axes of LangCortex and LangHippoc can be teased apart: LangCortex is more modulated by linguistic processing difficulty, whereas LangHippoc shows a preference towards particular kinds of content: descriptions of places and objects.

Keywords: language; semantics; hippocampus; selectivity; large language models; encoding models

Introduction

The hippocampus is crucial in forming relations across time and space—a core component of human memory (Cohen & Eichenbaum, 1993). Language also depends on relational processes: mapping word forms onto units of meaning and keeping track of semantic relationships across multiple timescales. However, whether or how the hippocampus contributes to language processing above and beyond the cortical language regions (Fedorenko et al., 2024) remains largely unknown. Hippocampal damage (even bilaterally) does not lead to severe language impairments (Vargha-Khadem et al., 1997), but can disrupt some aspects of language, including resolution of ambiguous discourse referents (Duff et al., 2011; Kurczek et al., 2013). In this work, we first identify regions in the hippocampus that respond to meaningful language. Second, we evaluate the *selectivity* for language over other tasks. Finally, using a data-driven modeling approach, we investigate the sentence-level *tuning properties* of the hippocampal language region relative to cortical language regions.

Methods

Identifying and characterizing the hippocampal language area: 603 participants completed a reading-based language 'localizer' task contrasting sentences (S) and strings of nonwords (N) in fMRI (Fedorenko et al., 2010). To define the cortical language network (LangCortex), we extracted the top 10% most localizer-responsive voxels within 5 broad anatomical masks in the left hemisphere for each participant. To search for language-responsive areas in the hippocampus (LangHippoc), we followed the same procedure, except using an anatomical hippocampal mask. A subset of 517 participants completed a spatial working memory (WM) task and a subset of 83 participants completed an arithmetic task, each with a harder (H) and easier (E) condition (Malik-Moraleda, Ayyash et al., 2022). We also defined a control region (ControlHippoc) in the hippocampus using the top 10% voxels that are most responsive to the hard condition of the spatial WM task relative to fixation. Finally, a subset of 25 participants completed a semantic plausibility judgment task (Sem) on sentences and pictures, with a perceptual judgment task (Perc) control condition (Ivanova et al., 2021).

Discovering the encoding axes of the hippocampal vs. cortical language areas: To investigate fine-grained response tuning, we used a condition-rich dataset of responses to n=1,000 diverse sentences from 8 participants (Tuckute et al., 2024). LangCortex, LangHippoc, and ControlHippoc were defined using the same procedure as above. We additionally extracted responses from another control region, the right temporoparietal junction (rTPJ), a region implicated in mental state attribution (Saxe & Powell, 2006); this region was defined using a probabilistic atlas based on a Theory of Mind localizer (Dodell-Feder et al., 2010). We fitted a GPT2-XL-based encoding model on the 1,000 sentences (ridge regression) and searched across ~1.8M sentences to find sentences predicted to maximize the *difference* between LangHippoc and LangCortex.

Results

Establishing responses to and selectivity for language/semantics in the hippocampus: First, we asked whether any regions of the hippocampus are responsive to language. Fig. 1A shows the probabilistic map of LangHippoc (yellow indicates higher overlap among participants, n=603), which shows a preference for meaningful language in the anterior portion of the hippocampus (β=.09, p<<.001). Next, we quantified language lateralization in the anatomical hippocampal region ($LI = \frac{(\#\ sig.\ vox.LH - \#\ sig.vox.RH)}{(\#\ sig.\ vox.LH + \#\ sig.\ vox.RH)}$; a voxel is significant if p<.01 for the S>N contrast). The hippocampus showed a left-hemisphere preference (*LI*=.47), similar to LangCortex (*LI*=.54). To test whether *any* part of the hippocampus would respond to language, we extracted responses from ControlHippoc, an equally sized control region located in the posterior hippocampus (Fig. 1B). ControlHippoc did not respond selectively to meaningful language (β=.006, p=.10), establishing that LangHippoc displays a preference for language not present throughout the entire hippocampus.

Second, we asked whether LangHippoc is selective for

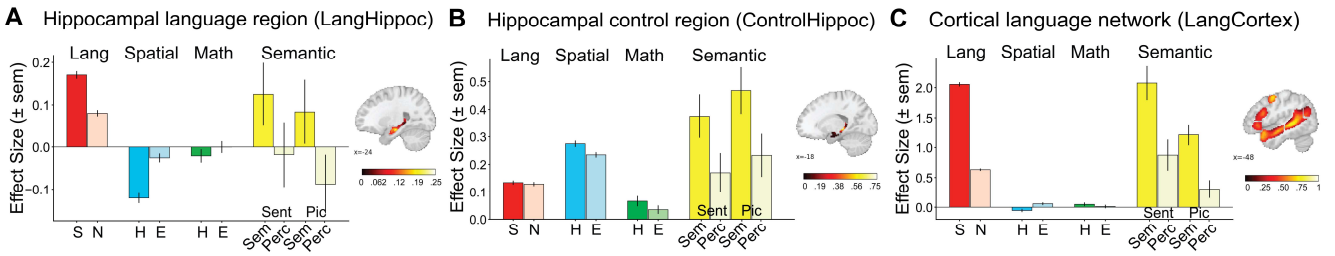

Figure 1: **(A)** Responses of the LangHippoc to each of the four tasks: language, spatial memory, math, and semantic judgement. The probabilistic map denotes the proportion of participants for whom each voxel was in the top 10% for the S>N contrast. **(B)** ControlHippoc **(C)** LangCortex.

language over non-linguistic tasks. We found that similar to LangCortex (Fedorenko et al., 2024), LangHippoc did not respond to a spatial WM task or an arithmetic task (β=-.10 & β=-.021, n.s.; Fig. 1A,1C) cf. ControlHippoc, which responded to both (β=.04 & β=.03, ps<.05; Fig. 1B). Finally, motivated by the role of the hippocampus in semantic memory (Manns et al., 2003), we examined LangHippoc's responses during a semantic judgment task on sentences and pictures (Methods) and found strong responses relative to a perceptual control task (β=.17, p=.05 (pictures) & β=.14, p=.04 (sentences)), although similar effects were observed in ControlHippoc (β=.24, p=.004 (pictures) & β=.21, p=.001 (sentences); Fig. 1A,1B). Across all conditions, the selectivity profile of LangHippoc but not ControlHippoc is similar to that of LangCortex (Fig. 1).

Teasing apart the tuning axes of LangHippoc vs. LangCortex: We performed an encoding model-guided search procedure (across ~1.8M sentences) to differentiate tuning properties of LangHippoc relative to LangCortex. We then characterized the resulting sentences using GPT2-XL-estimated surprisal, and 6 sentence properties regressed from ground-truth behavioral ratings of 2,000 sentences (Tuckute et al., 2024).

By driving LangHippoc and LangCortex independently, we found that both were driven by processing difficulty (quantified using surprisal), but in maximizing the response *difference*, we found that LangHippoc was tuned to semantic content relative to LangCortex, specifically, to content about places and physical objects (Fig. 2A; predicted mean effect size Δ=1.35). This pattern was similar for ControlHippoc but lower effect size (Δ=.17) (correlation between LangHippoc and ControlHippoc r=.48, cf. LangHippoc and LangCortex r=.30). To test whether the similarity between the hippocampal regions was meaningful (or whether any region outside LangCortex may show this profile), we applied the same approach to responses of LangHippoc and the rTPJ–a region implicated in mentalizing (Saxe & Kanwisher, 2003)–and showed that the rTPJ exhibits a distinct preference for sentences about mental state content (Fig. 2B; effect size Δ=1.66).

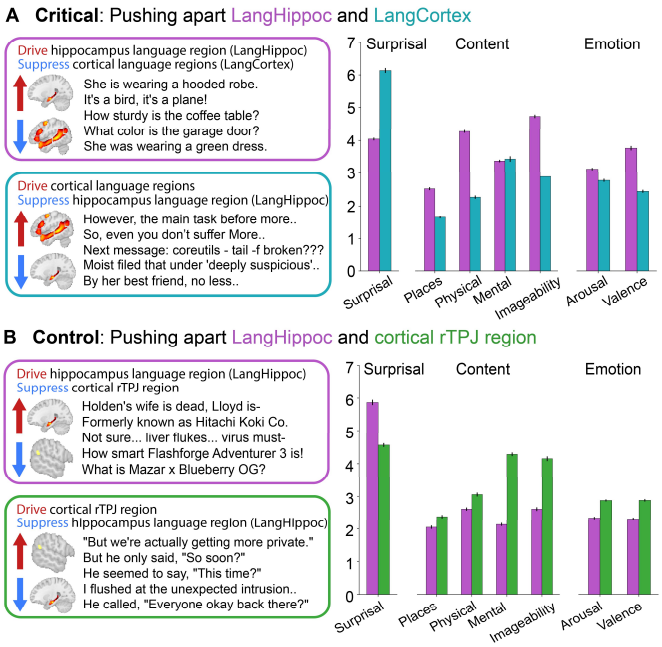

Figure 2: **(Left)** Sample sentences obtained from pushing apart two different regions. **(Right)** Average behavioral ratings for the top 250 sentences identified to maximally activate one region against another. **(A)** LangHippoc vs. LangCortex. **(B)** LangHippoc vs. cortical rTPJ.

Conclusion

We established a left-lateralized response to language in the hippocampus and characterized the selectivity of this region. Through an encoding-based analysis we then showed that although both LangHippoc and LangCortex are driven by processing difficulty, the tuning properties of LangHippoc are distinguishable from that of LangCortex – LangHippoc is more responsive to imageable content (places, objects) relative to LangCortex.

Acknowledgments

G.T. was supported by the Amazon Fellowship from the Science Hub, the International Doctoral Fellowship from the AAUW, and the K. Lisa Yang ICoN Graduate Fellowship. E.F. was supported by National Institutes of Health award U01-NS121471 and by research funds from the McGovern Institute for Brain Research, the Department of Brain and Cognitive Sciences, the Simons Center for the Social Brain, and the MIT Quest for Intelligence.

References

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system.* The MIT Press.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage, 55(2), 705-712.*

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The use of definite references signals declarative memory: Evidence from patients with hippocampal amnesia. *Psychological Science, 22*(5), 666–673.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S. & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104, 1177–1194.

Fedorenko, E., Ivanova, A.A. & Regev, T.I. (2024). The language network as a natural kind within the broader landscape of the human brain. Nat. Rev. Neurosci.

Ivanova, A. A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R., & Fedorenko, E. (2021). The language network is recruited but not required for nonverbal event semantics. Neurobiology of Language, 2(2), 176–201

Kurczek J, Brown-Schmidt S, Duff M. Hippocampal contributions to language: evidence of referential processing deficits in amnesia. J Exp Psychol Gen. 2013 Nov;142(4):1346-54. doi: 10.1037/a0034026. Epub 2013 Aug 12. Erratum in: J Exp Psychol Gen. 2019 Oct;148(10):1827.

Malik-Moraleda, S., Ayyash, D., Gallée, J. *et al* (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nat Neurosci* 25, 1014–1019.

Manns, Joseph R., Ramona O. Hopkins, and Larry R. Squire. "Semantic memory and the human hippocampus." Neuron 38.1 (2003): 127-133.

Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. Psychological Science, 17(8), 692-699.

Tuckute, G., Sathe, A., Srikant, S. *et al* (2024). Driving and suppressing the human language network using large language models. *Nat Hum Behav* 8, 544–561.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science, 277*(5324), 376–380.