Modeling the Visualization of Personal Experiences during Imagination in Medial Temporal DMN

Andrew J. Anderson, andanderson@mcw.edu
Depts. Neurology, Biomedical Engineering, Neurosurgery, Medical College of Wisconsin, USA

Abstract

Imagination enables the human brain to relive personal experiences by mentally simulating visual scenes. This ability has been linked to the Medial Temporal subsystem of the brain's Default Mode Network (MT-DMN). However, the representational codes associated with visualizing personal experiences have been understudied, in part because quantitatively modeling freeform imagination is challenging. To target this, we scanned fifty peoples' brain activity with fMRI as they reimagined their personal experience of twenty diverse natural scenarios (e.g. wedding/funeral/driving). To model the visualization of personal experiences, we deployed image-generation AI models to depict participants' verbal self-reports of their mental images (made outside the scanner) and image recognition models to re-represent the synthetic image features in a more abstract visual form, invariant to view and scale. A Representational Similarity Analysis suggested that MT-DMN selectively reflected visual model structure, when controlling for semantic features derived from language models. This effect was not observed in other brain networks which were more sensitive to the language model. This finding helps characterize the neural bases of imagination, and earmarks image AI models as valuable tools for neurally decoding imagination.

Keywords: Imagination, fMRI, Stable Diffusion, VGG, RSA, Language Model.

Main

Component processes of imagination and self-generated thought have been linked to three different subsystems of the brains' default mode network (DMN). Medial Temporal DMN has been implicated as the "mind's eye" underpinning the mental construction of visual scenes, Core DMN is associated with self-referential cognition, and Frontotemporal DMN is associated with abstract semantic, and social cognition (Andrews-Hanna & Grilli 2021, Andrews-Hanna et al. 2014).

To test the hypothesis that MT-DMN selectively simulates visual information structure during imagination, fifty neurotypical adults were recruited to undergo an fMRI experiment illustrated in **Figure 1 Step 1**. Before scanning, 20 generic scenario cues (e.g. dancing, exercising, wedding) were read to each participant, who vividly imagined themselves personally experiencing each scenario (Anderson et al. 2020). Participants provided a brief verbal description of each mental image. Participants then underwent fMRI as they re-imagined the same scenarios in random order on written prompt. fMRI preprocessing produced a single fMRI volume for each mental image per participant.

To test fMRI data for visual information structure, Stable Diffusion (Podell et al. 2023) was deployed to synthesize five images corresponding to each mental image scenario description (20*5 images per person). Because it was unreasonable to think that Stable Diffusion would synthesize the finer details of mental images, such as the spatial configuration of actors/objects, we transformed the synthetic images to more abstract visual representations reflecting the visual identity of object categories via the image classification model VGG 16 (Simonyan & Zisserman, 2015). We extracted embedding vectors from the final layer (Fc8), and pointwise averaged embeddings across the

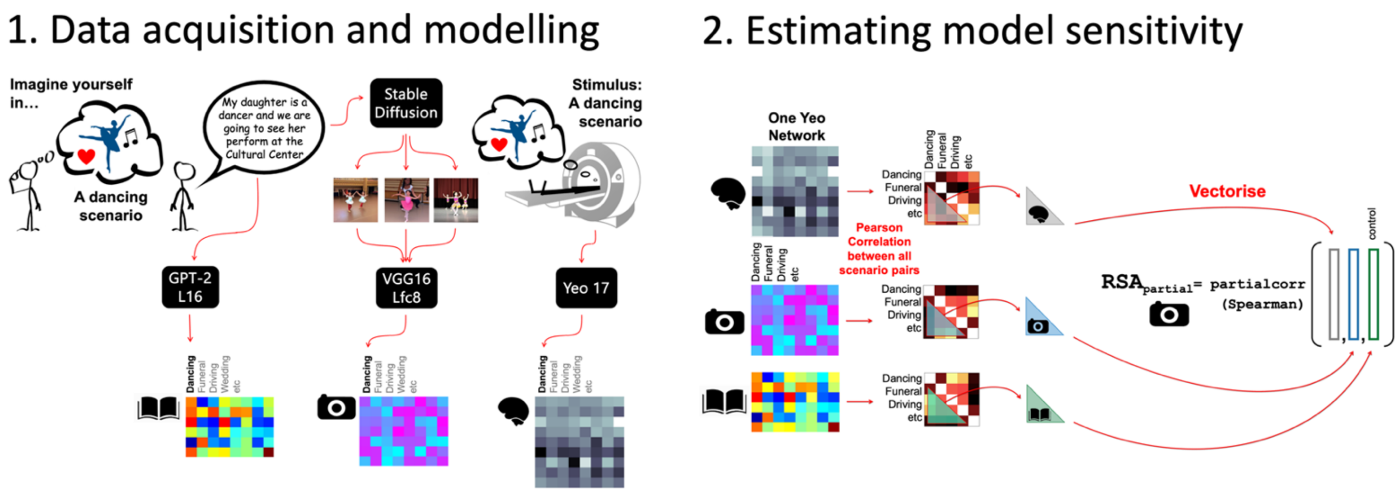

Figure 1. Step1 Data and Models. Step 2 Analysis to expose selective sensitivity to model structure.

five synthesized images vectors corresponding to each participant's description of one mental image, following related work (Anderson et al. 2017).

To control for non-visual semantic structure in the fMRI data, we also modeled participants' mental image descriptions with the language model GPT-2 (Radford et al. 2019). Embedding vectors from Layer 16/24 were extracted, following evidence that layers at 2/3 depth are a good choice for modelling fMRI data (Caucheteux, Gramfort & King, 2022).

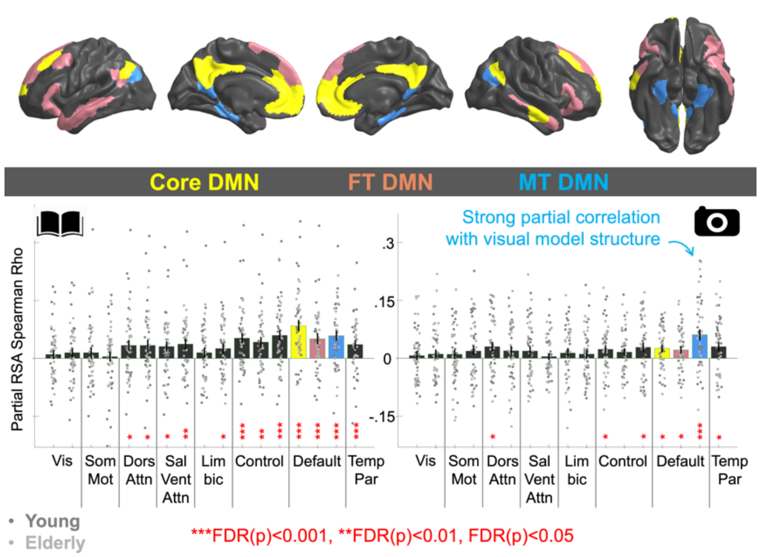

Figure 2. MT-DMN reflects the visual model

To test for visual information structure selective to MT-DMN, fMRI data was extracted from 17 pre-defined brain networks, according to the Yeo parcellation scheme (Yeo et al. 2011). These included MT/Core/FT-DMN. A partial correlation Representational Similarity Analysis (Kriegeskorte, Mur & Bandettini, 2008) was then used to compare fMRI data within each network to the visual model representations, controlling for the language model, and vice versa (**Figure 1 Step2**).

Signed ranks tests applied to the partial RSA coefficients (n=50) revealed that the visual model made a selectively strong contribution to explaining the representational structure of within MT-DMN, and much weaker contributions to all other networks (**Figure 2 Camera Icon**).

The language model made an equivalent contribution to explaining MT-DMN but made stronger contributions to explaining fMRI data in Core/FT-DMN and other networks (**Figure 2 Book Icon**).

Conclusions

This work supports the hypothesis that MT-DMN selectively contributes to visualizing mental scenes in imagination and demonstrates how AI image models can be deployed to help interpret the neural bases of imagination.

References

Anderson AJ, McDermott K, Rooks B, Heffner KL, Dodell-Feder D, Lin FV. (2020). Decoding individual identity from brain activity elicited in imagining common experiences. *Nature Communications*. 20;11(1):1-4.

Anderson AJ, Kiela D, Clark S, Poesio M. (2017). Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns." *Transactions of the Association for Computational Linguistics* 5 17-30.

Andrews-Hanna JR, Grilli MD (2021) Mapping the imaginative mind: charting new paths forward. *Curr Dir Psychol Sci* 30:82–89.

Andrews-Hanna JR, Smallwood J, Spreng RN (2014) The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann N Y Acad Sci* 1316:29–52.

Caucheteux C, Gramfort A, King J-R. (2022). Deep language algorithms predict semantic comprehension from brain activity." *Scientific reports* 12.1 16327.

Kriegeskorte N, Mur M, Bandettini PA. 2008. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*. 24;2:249.

Podell D, English Z, Lacey K, Blattmann A, Dockhorn T, Müller J, Penna J, Rombach R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. 2019. Language models are unsupervised multitask learners. OpenAI blog. 1(8):9.

Simonyan K., Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations, ICLR 2015

Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, et al. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol.* 106:1125–1165.