Models of laterally-biased cortical representations determine spatial working memory biases

Low-dimensional neural representations are thought to capture much of the interactions between brain regions, although it is still unclear how information is passed between them. A "communication subspace" has been proposed as a way in which brain regions interact. Neural network models can replicate this low-dimensional variability, but they have so far been agnostic to the visual hemifields and the role of the hemispheric structure in working memory. Including this lateralization is critical, as it is a major feature of the visual cortex, with the effect weakening along the cortical hierarchy, suggesting a dynamic importance of lateralization during cognitive processing. Here, we address this gap by developing two types of attractor network models and comparing them with behavioural data. Both models consist of two subcircuits, one for each hemisphere, and each with a more detailed representation of the contralateral hemifield, which when taken together represent the entire visual field. However, how the contralateral mapping is performed differs between the two models: In the "dense model", each subcircuit has more neurons representing the contralateral hemifield, with neuron-dependent connectivity scaling to keep the circuit functional. The "mapping model" has an internal space in which connectivity operates in a homogeneous manner, but the mapping of the stimulus space onto the internal space is laterally biased in a hemisphere-specific manner. For each model, we run multiple simulations with different stimulus locations and extract behavioural biases depending on where the stimulus was presented for each model. We find that both models replicate systematic biases to some extent, though further work is needed to discern both model types. This is a first approach at including lateralisation in visual working memory models to understand working memory biases.