Learning orthogonal working memory representations protects from interference in a dual task

A. Mahrach[1], X. Zhang[2,3], D. Li[2,3], C.T. Li[2,3], A. Compte[1]

[1] IDIBAPS, Barcelona, Spain
[2] Institute of Neuroscience, State Key Laboratory of Neuroscience, Chinese Academy of Sciences, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, China
[3] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China

Working memory (WM) is a cognitive function that allows for the short-term maintenance and manipulation of information when no longer accessible to the senses. It relies on temporarily storing stimulus features in the neuronal activity. However, the mechanisms protecting WM from task-irrelevant influences are unknown. Recent studies involving WM tasks have suggested that activity before and after distraction is decomposed into two orthogonal subspaces, one invariant across distraction, thus preserving WM from interferences. However, whether orthogonalization is a general mechanism for WM protection remains an open question, and the network mechanisms supporting it are unclear. Here, we investigated WM representations in calcium imaging data from the prelimbic cortex (PrL) in mice learning to perform a recent olfactory dual task. The task consists of an outer delayed paired-association task (DPA) combined with an inner Go-NoGo task. We studied how PrL reflected the process of learning to protect the representation of DPA samples against Go/NoGo distractor odors. As mice learned, we evaluated the overlap between PrL activity and the low-dimensional manifolds encoding sample/distractor odors. Early in training, activity before distraction positively overlapped with sample and distractor encoding subspaces, and distraction impaired generalization of the sample code. Later in training, non-distracted activity only overlapped with the sample subspace, and the sample code was more stable. We give a mechanistic account of how low-dimensional WM representations in PrL change with learning in an EI network model of recurrent spiking neurons with low-rank connectivity. Our model associates learning with (1) orthogonalization of sample and distractor WM subspaces and (2) orthogonalization of each subspace with corresponding irrelevant sensory information. We confirmed both predictions with photoinhibition of the Anterior Cingulate Cortex to PrL inputs. Altogether, our results suggest that rotations of WM representations in PrL play a fundamental role in preserving WM from interfering tasks.