TITLE:
The Neural Traces of Working Memory Forgetting.

AUTHORS:
Tíffany Oña-Jodar[1], Genís Prat[1,2], Chengyu Li[3], Josep Dalmau[4,5], Albert Compte[1,*], Jaime de la Rocha[1*]

1. Brain Circuits and Behavior Lab, IDIBAPS, Barcelona
2. University of Pittsburgh, Department of Bioengineering, Pittsburgh, USA
3. Chinese Academy of Sciences, Shanghai Center for Brain Science and Brain-Inspired Technology, Shanghai, China
4. Catalan Institution for Research and Advanced Studies (ICREA)
5. Faculty of medicine University of Barcelona, Hospital Clinic de Barcelona, Barcelona

(*equal contribution)
Keywords: working memory, persistent activity, attractor states, lapses, sequential effects, double-well model

Abstract: Working memory (WM) is central for cognition and is impaired in many brain disorders. Electrophysiology studies suggest that persistent selective activity in cortical areas may be the basis of WM maintenance and attractor states have been proposed as the underlying network mechanism conferring properties such as stability and capacity. Although attractor states present strong constraints to the limits of WM, a direct assessment of memory errors under this theoretical framework is still scarce. To investigate what makes WM fail, we used a two-alternative delayed-response task in which mice listen to a lateralized auditory stimulus and, after a variable delay (duration D=0-10 s), they lick the associated lateral port. Response accuracy decreased with delay showing that there were forgetting errors. Mice also showed non-memory errors or lapses, which occurred independently of delay duration. We modeled this behavior by fitting a two-state hidden Markov Model that stochastically switches between (1) an engaged WM state where memory-guided responses are dictated by the dynamics of a double well attractor model, and (2) a disengaged state which elicits lapse responses mostly determined by previous choices. Population recordings in the anterolateral motor cortex (ALM) showed that persistent selective activity encoded the impending choice during the mnemonic period in engaged correct trials but the encoding was significantly weaker in disengaged correct trials. Furthermore, errors in long-delay engaged trials showed a choice encoding reversal along the delay, the neural signature of a switch between the correct and the error choice attractor. Our results show that mice performance in a delayed-response task is limited by both lapses and forgetting errors possibly caused by fluctuation-driven switches between attractor memory states.