1-005. Creative memory semantization through adversarial dreaming

Nicolas Deperrois[1,2] NICOLAS.DEPERROIS@UNIBE.CH
Jakob Jordan[3,4] JAKOB.JORDAN@UNIBE.CH
Mihai Petrovici[5,6] MIHAI.PETROVICI@UNIBE.CH
Walter Senn[1] WALTER.SENN@UNIBE.CH

[1]Universitat Bern
[2]Institut fur Physiologie
[3]University of Bern
[4]Department of Physiology
[5]University of Bern and Heidelberg University
[6]Department of Physiology and Kirchhoff-Institute for Physics

Classical theories of memory consolidation emphasize the importance of sleep-mediated replay as a key mechanism to extract semantic information from episodic memories. However, the creative nature of dreams suggest that efficient memory semantization may go beyond merely replaying previous experiences. Here, we suggest that rapid-eye-movement (REM) sleep reorganizes memory by randomly combining episodic memories and creating entirely new visual experiences. Non-REM (NREM) sleep, in contrast, is responsible for making internal representations robust via the perturbed replay of encoded memories. We support this hypothesis by implementing a cortical architecture with separate, hierarchically organized forward and backward pathways, loosely inspired by generative adversarial networks (GANs). During wakefulness, episodic memories are stored in hippocampus. During REM sleep, these memories are randomly combined in high-level areas to generate new, hypothetical activity patterns in low-level areas. While feedforward pathways learn to distinguish these internal dreams from externally driven inputs, feedback pathways adversarially learn to generate more realistic activity patterns. During NREM sleep, episodic memories generate perturbed activity in early sensory areas. Reconstructing the encoded memory from this activity improves the robustness of semantic memory to potential sensory perturbations. Our cortical architecture, trained on standard benchmark datasets, develops rich latent representations in an unsupervised fashion. Using these, a linear classifier achieves competitive recognition performance, on par with unsupervised machine learning methods. By systematically evaluating the quality of the learned representations, our results demonstrate the complementary function of NREM and REM sleep. The suggested architecture and learning paradigm are amenable to cortical microcircuit implementation in terms of different classes of layer 2/3 pyramidal neurons (Sacramento et al., 2018). Our biologically-inspired framework highlights the importance of creative dreaming for the successful semantization of memories, providing a new, functional perspective on the link between sleep, dreaming, hippocampal replay and memory consolidation.