The role of sleep in consolidating semantic knowledge

Anna Schapiro[1], Ken Norman[1*], & Tim Rogers[2*]

*equal contribution

[1]Dept. of Psychology, Princeton Neuroscience Institute, Princeton University

[2]Dept. of Psychology, University of Wisconsin-Madison

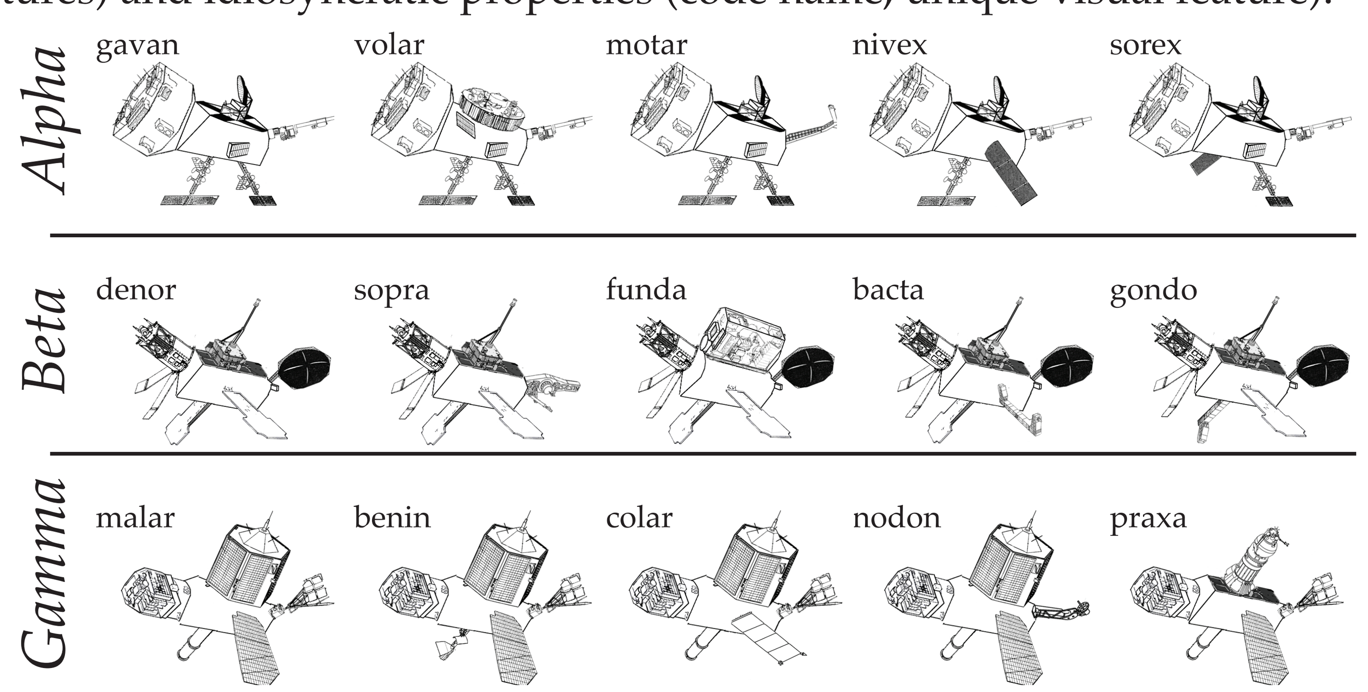
Introduction

Sleep is thought to be crucial for initial placement of new arbitrary, episodic information into cortical knowledge structures[1,2]. During slow wave sleep (SWS), the hippocampus replays memories of recent experiences, promoting consolidation of the memories in cortex[3].

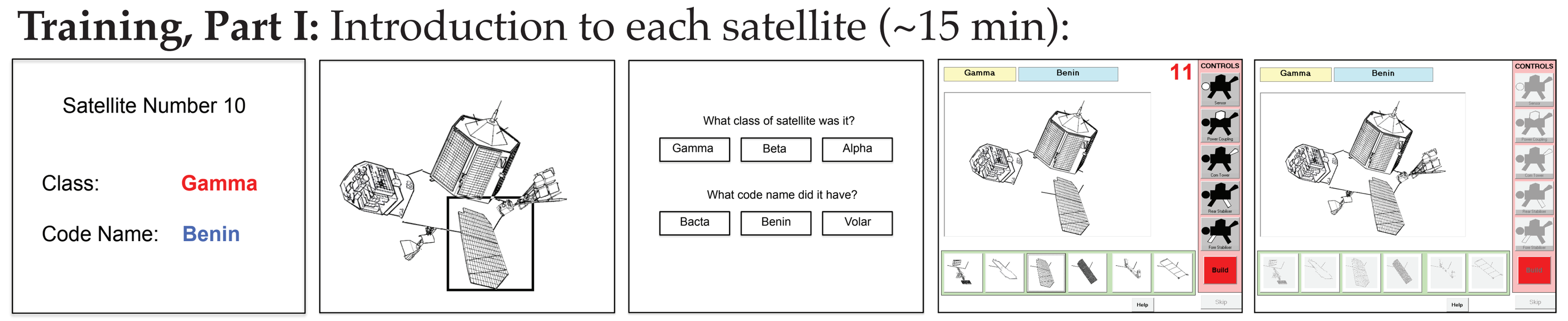
- **How does sleep impact the learning of new *structured* information?**
- **What computational mechanisms might underly changes in this structured knowledge during sleep vs. wake periods? In particular, how might the hippocampus and cortex interact to support the consolidation of semantic information?**

Stimuli and design

Each "satellite" had properties shared with class (class name, shared visual features) and idiosyncratic properties (code name, unique visual feature):

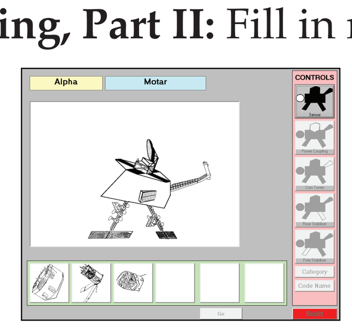


Training, Part I: Introduction to each satellite (~15 min):



Study names — Study images — Recall names — Reconstruct image — Receive feedback

Training, Part II: Fill in missing part or name with feedback (~30-45 minutes):

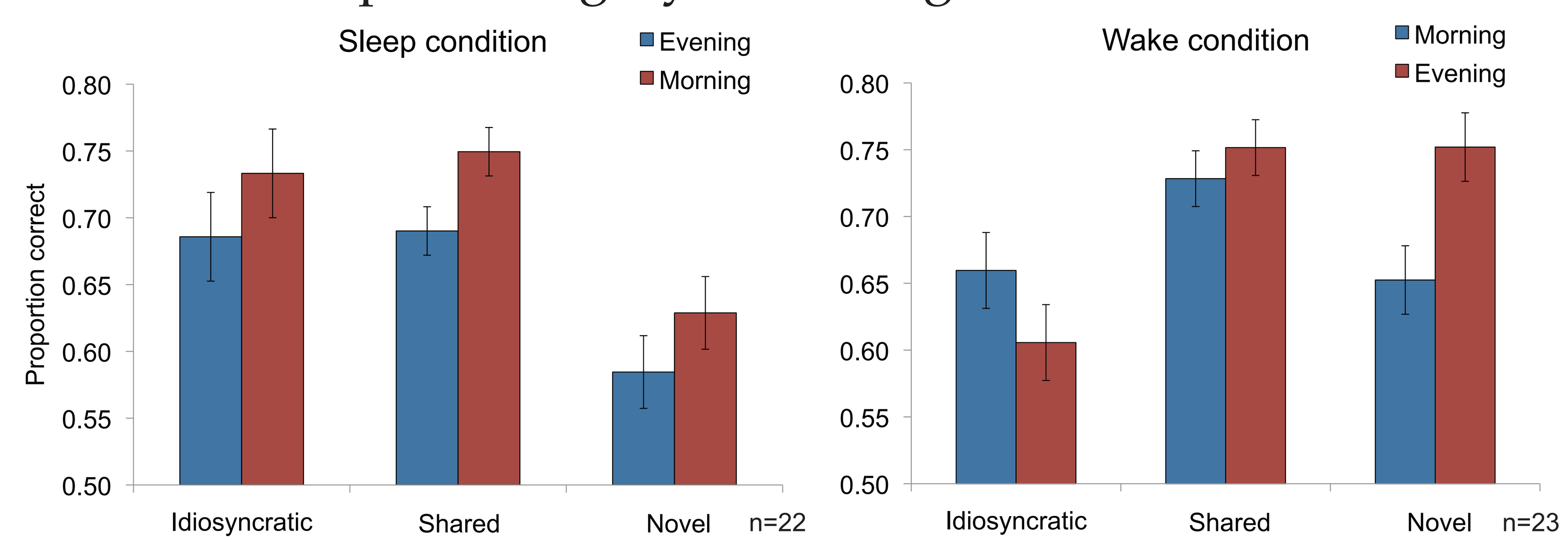


- Idiosyncratic features tested more frequently to match shared and idiosyncratic feature performance.
- Training stops when average performance at 66% correct.

Test: Fill in missing part, code name, or class name of trained satellites, and missing part or class name of novel category exemplars, without feedback.

Sleep condition:	Training (8pm)	Test	12 hours (including sleep)	Test
Wake condition:	Training (8am)	Test	12 hours (wake)	Test

Effects of sleep on category knowledge

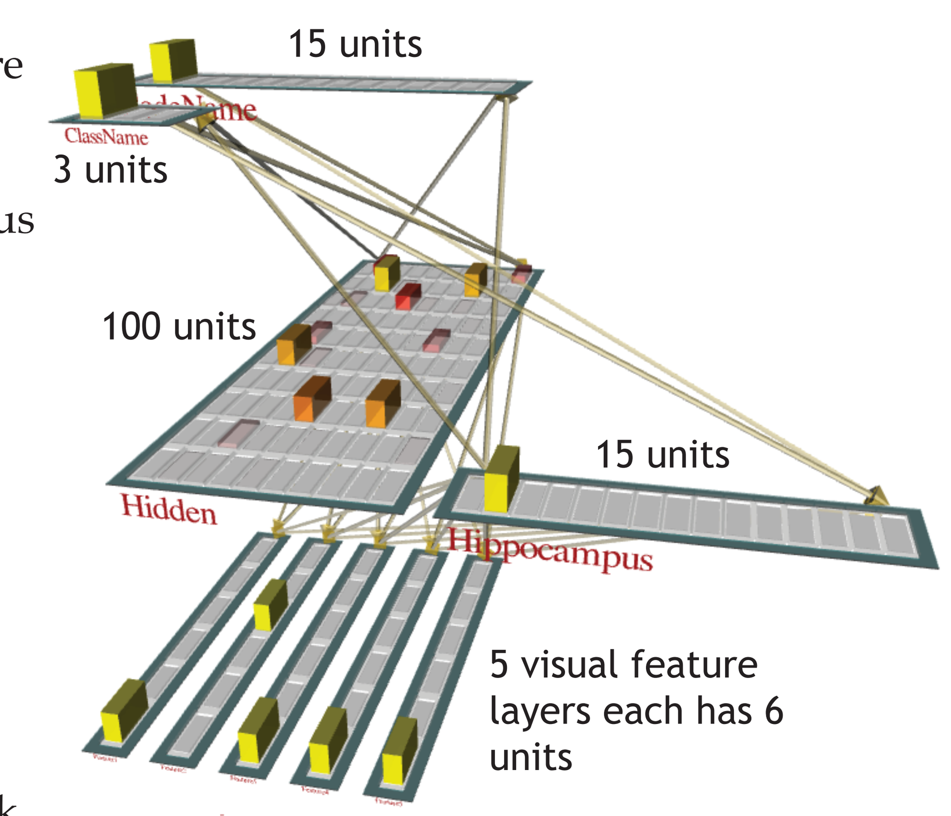


- Idiosyncratic feature memory better after sleep than wake (p=0.047).
- Interaction between change in idiosyncratic features for sleep vs. wake (p=0.025).
- Shared and novel items don't differ over sleep vs. wake; both improve (ps<0.01).

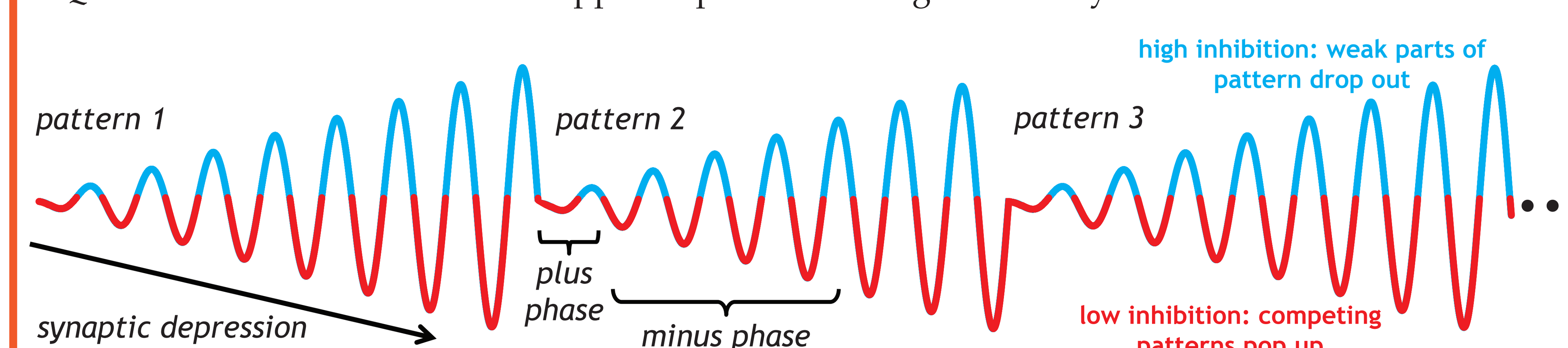
A computational account

- The brain receives minimal input and no feedback from the environment during sleep, making useful learning a computational challenge.
- Model has layers representing satellite features, a cortical hidden layer, and a hippocampus layer, where each unit connects to all features of one satellite.
- Training: all features except one clamped during minus phase and remaining feature added during plus phase.

$$\Delta \text{weight} = \text{learning rate} \cdot \left(\text{average plus phase coactivity} - \text{average minus phase coactivity} \right)$$

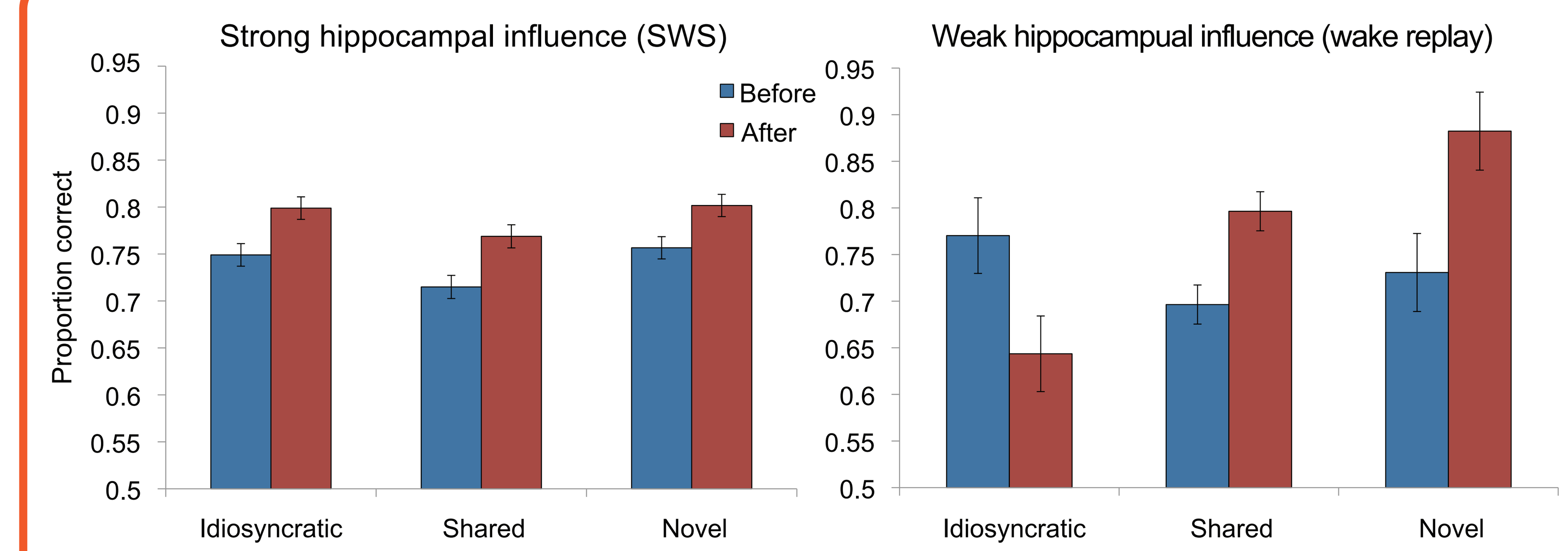


- Sleep begins by setting unit activations randomly; network falls into nearby attractor.
- Synaptic depression causes transition to next attractor. Also prevents repeatedly visiting the same attractors[4].
- Inhibitory oscillations distort pattern to reveal weak parts of memories and competing memories.
- Plus phase corresponds to period of high stability in activation pattern; slight drop in stability triggers minus phase, which continues until further drop below threshold.
- Same learning as above: modify distorted versions of pattern to look more like clean pattern.
- SWS: Hippocampus drives dynamics. Oscillations dominated by low inhibition.
- Quiet wake[5]: Weak influence of hippocampus. Less stringent stability criteria.



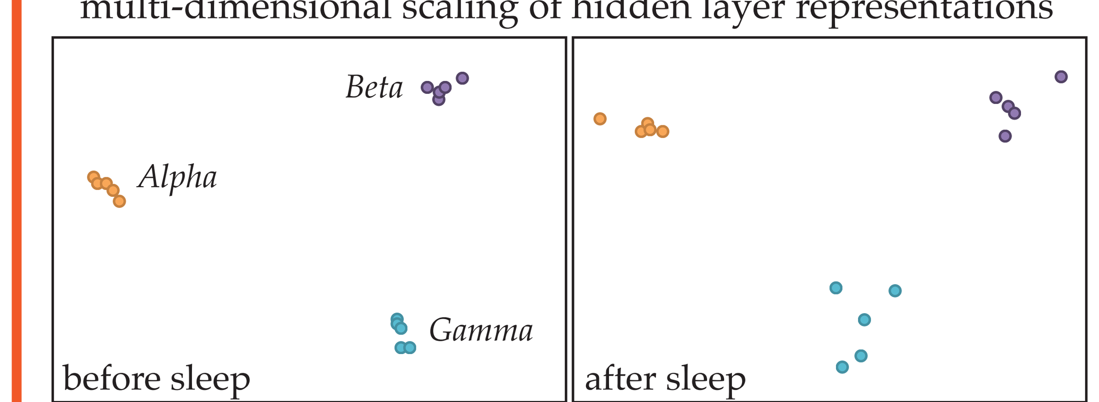
In collaboration with Seth Herd and Randy O'Reilly

Simulation results

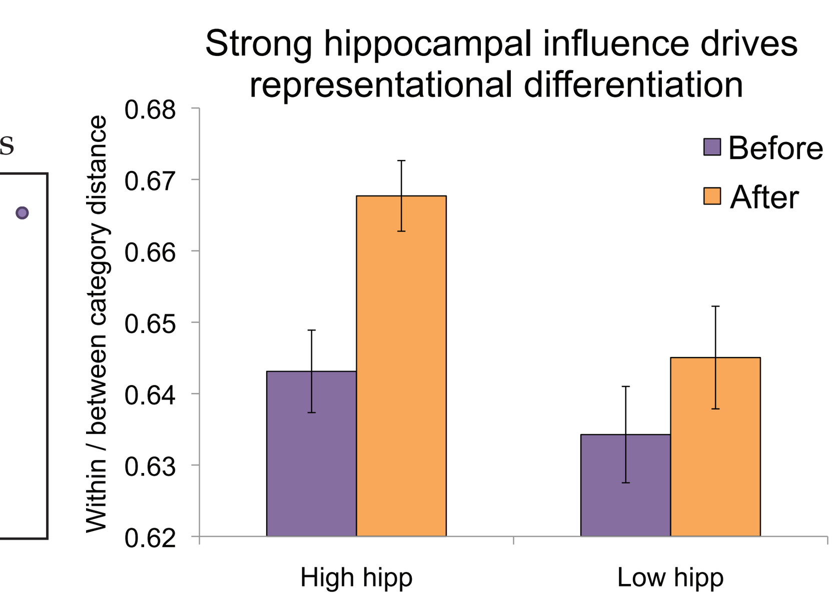


- Offline learning with strong hippocampal influence results in improved performance for all object features as well as improved generalization.
- Offline learning with weak hippocampal influence helps shared features and generalization, but hurts memory for idiosyncratic properties.
- All changes significant (ps<0.01)

multi-dimensional scaling of hidden layer representations



Strong hippocampal influence drives representational differentiation



Summary and discussion

- **Sleep uniquely enhances memory for idiosyncratic properties of category exemplars. Sleep and wake periods both improve shared properties and generalization ability.**
- **These effects can be simulated using a neural network model that learns autonomously during offline periods based on just-formed attractors.**
- **Different kinds of offline learning may be characterized by varying degrees of hippocampal influence (more during SWS and less during wake).**
- Ongoing and future directions: Running a nap version of paradigm in collaboration with Sara Mednick to directly assess contributions of sleep stages with PSG, and running fMRI version to test model's predictions about representational changes over different kinds of offline learning periods.

References and funding

1. Norman, K.A. & O'Reilly, R.C. (2003). Psychol. Rev., 110, 611-646.
2. McClelland, J.L., McNaughton, B.L., & O'Reilly, R.C. (1995). Psychol. Rev., 102, 419-457.
3. Born, J. (2010). World J. Biol. Psychiatry, 11 (Suppl. 1), 16-21.
4. Norman, K.A., Newman, E.L., & Perotte, A.J. (2005). Neural Networks, 18, 1212-1228.
5. Foster, D.J., & Wilson, M.A. (2006). Nature, 440, 680-683.

NSF GRFP Grant DGE-0646086 to A.S. and NIH Grant R01MH069456 to K.N.