

Should Androids Dream of Electric Sheep? Mechanisms for Sleep-dependent Memory Consolidation

George Kachergis¹, Roy de Kleijn², and Bernhard Hommel²

¹george.kachergis@nyu.edu

¹Psychology Department, New York University, NY, USA

²Institute for Psychology / LIBC, Leiden University, the Netherlands

Keywords: memory consolidation; sleep; dreaming

Humans spend almost a third of our lives asleep, with the most convincing explanation being that we otherwise suffer degradation of many cognitive and motoric skills. However, there is now also substantial empirical evidence that both declarative (i.e., facts and events—‘what’, ‘where’, ‘when’) and procedural (i.e., skills—‘how’) memory benefit from even short periods of sleep. Memory is typically described as three processes: 1) encoding: forming new traces from experience, 2) consolidation: integrating memories with prior knowledge and strengthening/crystallizing the trace, and 3) retrieval: task-dependent extraction of overall familiarity or recall of particular traces. Sleep is generally accepted to aid in consolidation, but under what circumstances it helps and by what mechanisms is not well understood.

Storage versus Processing

Machine learning algorithms can be classified as either incremental—allowing data to be added to the model instance by instance—or batch, requiring a (sometimes large) set of training instances before the model produces useful predictions. Incremental or online algorithms (e.g., naïve Bayes) clearly offer the advantage of being able to work (however poorly) with very little data, and can learn immediately when new data are acquired. Moreover, since instances are processed immediately, they do not need to be stored for later updating. One disadvantage is that online updating may require significant computational resources, perhaps at an inconvenient time. In contrast, batch (i.e., offline; e.g., support vector machines, decision trees) learning algorithms may need a large store of data and quite some time to build an initial useful model, and adding a single training instance may require iterating over the entire (and increasing) data store to update the model. A survey of learning algorithms will reveal the classic algorithmic tradeoff: one can store more, and process less upfront (but retrieval can be costly), or process more upfront and store less.

Another problem with many incremental algorithms is the potential to arrive at different learning outcomes based on the order the instances are encountered in. In many cases, such order effects are undesirable, but humans and animals show a variety of order effects (e.g., in associative learning: Kachergis (2012). Could sleep be a chance to mitigate the order effects brought on during online learning? A few batch-update models have been found to have roughly-equivalent incremental versions. For example, latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990, LSA) learns semantic similarities of words via the singular

value decomposition (SVD)—an expensive matrix operation—of a large word $\times$ document co-occurrence matrix. This large matrix—adults know over 70,000 unique words, and have read thousands of documents—must be kept in memory to be updated when a new document is read. Updating the model requires performing the SVD again, so it would be quite expensive to update knowledge every time a new document is read. It is more sensible to read a batch of documents—although, of course, this means that any new knowledge is not available in the model until the latest batch is incorporated.

Models that use batch updating require storing all of the instances in long-term memory, allowing the model to iterate over all episodes—even multiple times—to extract higher-level features (e.g., correlations of multiple features). On the other hand, incremental updating can reduce the need to store so much information, much of which may be redundant or already over-learned. We conclude that sleep might be a way to get the best of both worlds: incremental learning based on salient features for immediate use, in addition to storage of daily episodes—especially exciting or confusing memories—that can be replayed during sleep to make more thorough, careful updates to knowledge representations before further compressing the memories.

Sleep Characteristics, Effects, and Theory

Sleep in mammals and birds consists of cycles of four stages, proceeding from non-rapid eye movement (NREM) stages 1, 2, and 3 (also called slow-wave sleep), to rapid eye movement (REM) sleep. Human adults typically go through four or five cycles each night, reaching REM sleep every 90 minutes or so. More slow-wave sleep (SWS; NREM3) occurs early in the night, whereas more REM sleep occurs in the last few hours of a night’s sleep. Loss of NREM3 and REM sleep results in drastically increases in these stages the following night, suggesting they are of critical importance. From neural recordings of rats, it appears that memory replay during non-REM sleep occurs at a 10x speedup, whereas REM replay is roughly at the speed of the behavioral episode (Bendor & Wilson, 2012). Waking levels of acetylcholine (ACh) during REM sleep may support encoding of new declarative memories, whereas low ACh during SWS is thought to allow replay and transfer of hippocampal memories to the neocortex (Hasselmo, 1999).

Although implicit memory effects have also been found in sleep studies, we focus on declarative memory (i.e., semantic and episodic memory; facts and knowledge). Declarative memory is thought to be largely dependent on the hippocampus enabling sleep-based consolidation of memory. During

SWS, episodic information stored in the hippocampus is replayed and projected to brain regions in the neocortex, storing stable, permanent memories. This information flow reverses during later REM sleep, conceivably allowing the hippocampus to remove the unstable, short-term memories in order to make room for new memories to be stored there (Wamsley & Stickgold, 2011). Sleep has been shown to improve recall for nonsense syllables (Jenkins & Dallenbach, 1924) and for paired-associate word stimuli (Gais & Born, 2004).

For declarative memory, there are two basic theories of how memory consolidation is improved during sleep: the active hypothesis states that consolidation depends on sleep, whereas the permissive hypothesis views consolidation as a time-dependent, interference-sensitive process that uses periods of low hippocampus input to process prior information (Mednick & Alaynick, 2010). Procedural memory is just generally thought to be ‘enhanced’ by sleep, but this idea is not universally accepted (Mednick & Alaynick, 2010). We will focus on proposing specific computational mechanisms for improving declarative memory, since the current models are more readily adapted to this task, and the empirical evidence indicating the necessity for this is strong.

Proposed Mechanisms

Our proposed modifications will be specified in terms of the REM (Retrieving Effectively from Memory) model from Shiffrin and Steyvers (1997), which is a multitrace memory model representing both episodic traces as well as lexical-semantic traces. Our first proposed modification is that the updating of the lexical-semantic (LS) features—which is typically not even simulated in REM—could take place during a sleep period, when episodic traces since the last sleep period are (randomly, or perhaps surprising or emotionally-charged ones) reactivated. That is, we assume that updating LS traces is tantamount to modifying the neocortical representations, which is best left for an offline period. Meanwhile, the hippocampal episodic traces may still be retrieved and used in various ways throughout the day. REM assumes that when the same stimuli appear multiple times in similar contexts, the old trace may be updated by filling in missing features from LS traces, instead of making a new trace (this differentiation process is how it accounts for the word frequency mirror effect and null list strength effect).

Retrieval in REM uses context features—reinstated by the probe, whatever its source (internal or external)—to activate a subset of long-term memory (e.g., to the studied list of items). For recognition, REM computes a likelihood ratio indicating how well a test cue (from the LS traces) match each episodic trace in the activated subset being considered. This likelihood ratio incorporates the base rate in the long-term, and the number of both the non-zero mismatching and matching features. Thus, the decision depends on not only the number of matching features, but also on how diagnostic the features are. Since small feature values will tend to be quite common and thus undiagnostic, whereas the more useful large feature values are rarely encountered, a potential mechanism for im-

proving memory would be to redistribute feature values. By choosing at least one (unique) high-valued, diagnostic feature for each trace (or group of highly-related traces), memory will be improved. This is clearly quite computationally expensive (which is why it should be offline), but a simple, greedy version might choose one of the common stimuli from the day’s traces, select one of its’ LS trace’s common features, and increment that feature value by one.

The SARKAE (Storing And Retrieving Knowledge And Events) model (Nelson & Shiffrin, 2013) develops REM further to explain how knowledge co-evolves along with episodic memory. Unlike REM’s traces, SARKAE’s traces represent not only feature values but counts of each feature value (e.g., “blue”) organized by feature types (e.g., “color”). Event traces contain a single feature value (with a count of 1) if the value is copied from the stimulus, but a feature count vector may instead be copied from memory. Knowledge traces are simply those event traces that have been reactivated and updated many times, and thus contain distributions of feature values. In the SARKAE framework, a straightforward role for sleep is to act as the cleanup period: event traces from the day are considered in turn, and their feature counts are either added to an existing knowledge trace (in cortex) if a similar one is found, or copied as a new knowledge trace.

The proposed mechanisms involve many comparisons and updates to long-term lexical-semantic traces stored in neocortex, making them more suitable for conducting during sleep. Note that while the complex version of redistributing diagnostic feature values would have to be done in batch, the simple greedy version (choosing a single feature to increment) is more batch-incremental. Although we specified these mechanisms in terms of the REM model, the same mechanisms could be used in related multitrace modeling frameworks such as SARKAE or MINERVA2.

References

- Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nat. Neuro.*, 15(10), 1439–1444.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6).
- Gais, S., & Born, J. (2004). Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory*, 11(6), 679–685.
- Hasselmo, M. E. (1999). Neuromodulation: Acetylcholine and memory consolidation. *Trends in Cognitive Sciences*, 3, 351–359.
- Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *American Jour. of Psych.*, 35(4), 605–612.
- Kachergis, G. (2012). Learning nouns with domain-general associative learning mechanisms. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proc. of cogsci 34* (p. 533-538). Austin, TX: Cognitive Science Society.
- Mednick, S. C., & Alaynick, W. A. (2010). Comparing models of sleep-dependent memory consolidation. *J. Exp. Clin. Med.*, 2(4), 156–164.
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, 120(2), 356–394.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4(2), 145–166.
- Wamsley, E. J., & Stickgold, R. (2011). Memory, sleep, and dreaming: Experiencing consolidation. *Sleep Med. Clin.*, 6(1), 97–108.