Research Article

Psychological Science
21(1) 80–85

DOI: 10.1177/0956797609354074
http://pss.sagepub.com

Consolidation of Episodic Memories During Sleep: Long-Term Effects of Retrieval Practice

Mihály Racsmány[1,2], Martin A. Conway[3], and Gyula Demeter[4]
[1]Department of Cognitive Science, Budapest University of Technology and Economics; [2]Institute of Psychology, University of Szeged; [3]The Leeds Memory Group, Institute of Psychological Sciences, University of Leeds; and [4]Research Group on Cognitive Science, Hungarian Academy of Sciences and Budapest University of Technology and Economics

Abstract

Two experiments investigated the long-term effects of retrieval practice. In the retrieval-practice procedure, selected items from a previously studied list are repeatedly recalled. The typical retrieval-practice effects are considerably enhanced memory for practiced items accompanied by low levels of recall, relative to baseline, for previously studied items that are associated with the practiced items but were not themselves practiced. The two experiments demonstrated that the former effect persisted over 12 hr; the latter effect also persisted over 12 hr, but only if a period of nocturnal sleep occurred during the retention interval. We propose that consolidation processes occurring during sleep, and possibly featuring some form of off-line rehearsal, mediate these long-term effects of retrieval practice.

Keywords

retrieval practice, episodic memory, sleep, consolidation, rehearsal

Received 1/20/09; Revision accepted 4/29/09

It has long been thought that sleep plays a crucial role in the consolidation of recently formed memories. Current evidence shows that retention of procedural knowledge can be enhanced by a period of sleep (Stickgold & Walker, 2005), as can retention of motor skills (Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002). In a recent programmatic series of studies, Gaskell, Dumay, and their coworkers demonstrated that sleep is critical to the retention of new vocabulary, and in particular to the integration of newly acquired words into the lexicon (see Dumay & Gaskell, 2007, for a review). Furthermore, it has been observed that relatively few episodic memories formed during a day are retained the following day, which suggests that only a minority of episodic memories are selected for enduring retention (Conway, 2009; Williams, Conway, & Baddeley, 2008). According to one view, consolidation processes operating during sleep mediate these effects. Reactivation of the medial temporal lobe memory system, and especially hippocampal circuits, may be the locus of sleep-mediated consolidation (Wilson & McNaughton, 1994). Other brain areas have been implicated too, and it seems that networks in medial prefrontal cortex, operating at faster processing rates during sleep than during awake periods, rapidly and repeatedly replay processing sequences featured in the immediately preceding awake period (Euston, Tatsuno, & McNaughton, 2007). This mechanism may consist of a sequence of speeded off-line rehearsal, and possibly it is these intense bursts of rehearsal that lead to the consolidation of recent experience in long-term memory.

Consolidation processes operating during a nocturnal sleep cycle should influence the retention of recently formed episodic memories, and we explored this idea in two experiments using the retrieval-practice procedure (Anderson, Bjork, & Bjork, 1994; Racsmány & Conway, 2006). In this procedure, participants first study a list of words and then selectively practice recalling a subset of the list. Memory is then tested, typically by cued recall. The retrieval-practice procedure is particularly suited to exploring the consolidation of episodic memories, as it is thought that the study phase gives rise to the formation of an episodic memory of learning the study list and that the later practice phase gives rise to a pattern of activation and inhibition over the contents of the episodic memory. It is this pattern of activation and inhibition that mediates later access to memory content and that gives rise to the characteristic pattern of recall seen on the memory test (Racsmány & Conway, 2006). By this view, retrieval practice should give rise to long-term patterns of activation and inhibition that are strengthened by consolidation during sleep. The sole previous

Corresponding Author:
Mihály Racsmány, Department of Cognitive Science, Budapest University of Technology and Economics, Stoczek u. 2., Budapest, 1111, Hungary
E-mail: racsmany@cogsci.bme.hu

study of long-term retrieval-practice effects indicates that this may indeed be the case (MacLeod & Macrae, 2001), although we acknowledge that other researchers consider the effects of retrieval practice to be more likely short-term than long-term (Saunders & MacLeod, 2002; but see Anderson, 2001), and at least one current model (Norman, Newman, & Detre, 2007) proposes that REM sleep may "reset" inhibitory patterns. The retrieval-practice procedure was well suited for our study because it easily allows a period of sleep or equivalent period of wakefulness to be interposed between practice and test.

Experiment 1

In the retrieval-practice procedure, exemplars from various categories are first studied. After the study phase, selected items from selected categories are then repeatedly recalled, typically three times, in response to cues consisting of a category name plus word fragment. For example, if "fruit-orange" is a studied item, "fruit-o_____?" might be a retrieval-practice cue. The three phases of study, practice, and test usually are separated only by the few minutes required to give the instructions for each phase. The design yields three types of items: items that have been practiced (Rp+), items that have not themselves been practiced but that originate from a category for which another item has been practiced (Rp–), and items from categories for which no items have been practiced (Nrp).

The typical finding is that memory for Rp+ items is highest, memory for Nrp items is at an intermediate level, and memory for Rp– items is poorest. This pattern is taken to indicate strong activation of Rp+ items resulting from retrieval practice making these items highly accessible to recall, weaker activation of Nrp items, and inhibition of Rp– items (Anderson & Spellman, 1995; Bjork, Bjork, & Anderson, 1998; Racsmány & Conway, 2006; Storm, Bjork, Bjork, & Nestojko, 2006). According to this explanation, practice recalling an item from a previously studied set of category exemplars induces inhibition of exemplars that are not practiced and that could potentially compete with and disrupt recall of the cued items (cf. Anderson & Levy, 2007). Thus, studying "apple," "pear," and "orange" and then repeatedly practicing recall of only "orange" induces inhibition of "apple" and "pear." The net result is that memory for "apple" and "pear" (Rp– items) is hurt, whereas memory for "orange" (an Rp+ item) is enhanced. Other interpretations of these effects of retrieval practice have emphasized the role of interference rather than inhibition (e.g., Camp, Pecher, & Schmidt, 2007; see also Mensink & Raaijmakers, 1988).

In our first experiment, participants were assigned to two groups: a sleep group and a no-sleep group. The sleep group studied and practiced the items in the evening; the following morning, some 12 hr later and after their usual period of nocturnal sleep, their memory for the items was tested. The no-sleep group studied and practiced the items in the morning; 12 hr later, in the evening, their memory was tested.[1] We expected that the no-sleep group would not show the typical retrieval-practice effect and instead would simply show forgetting of the items. In contrast, and assuming that consolidation can enhance retention, we expected the sleep group to show the usual retrieval-practice pattern.

Method

Participants. Sixty-four undergraduate Hungarian students from the Budapest University of Technology and Economics (32 females, 32 males) participated in return for partial credit in an introductory psychology course. Their ages ranged from 19 to 26 years. There were 32 participants each in the sleep and no-sleep groups (16 females and 16 males randomly assigned within gender to each group). Note that all participants in the sleep group were tested a minimum of 1 hr after awakening.

Materials. Following Anderson et al. (1994), we used 10 categories, 2 of which were fillers. Each target and filler category consisted of 6 exemplars. Exemplars were moderate- to high-frequency words drawn from two Hungarian word-frequency norms (Füredi & Kelemen, 1989; Kónya & Pintér, 1985). For each subject, 4 target categories were practiced and 4 were nonpracticed; across subjects, each target category was equally often practiced and nonpracticed. The practiced and nonpracticed exemplars from practiced categories were counterbalanced over participants. In sum, in each learning session, participants learned 60 exemplars from 10 categories (2 of which were fillers), practiced 18 exemplars (including 6 fillers) from 6 categories (including the 2 filler categories), and finally tried to recall 60 exemplars (including 12 fillers) from the original 10 categories. During both practice and final cued recall, items from filler categories were always in the first and last positions in order to avoid the confounding effect of category position.

Procedure. Participants were randomly assigned to either the sleep or the no-sleep group. All participants completed a short questionnaire about the length and quality of their sleep period prior to the experiment. Those who had slept less than 4 hr or used sleeping pills were excluded from the experiment. The no-sleep group completed the sleep questionnaire only on the day of the experiment, answering the questions with reference to the previous night's sleep, whereas the sleep group completed the same sleep questionnaire on the day of the study phase and also on the day of the recall test, in each case answering the questions with reference to the previous night's sleep.

At 8 p.m., the sleep group completed the study phase followed by the practice phase; these participants returned to the laboratory for the surprise delayed recall test at 8 a.m. the following morning. Note that in all cases the test was given a minimum of 1 hr after awakening. The no-sleep group completed the study phase and practice phase at 8 a.m. and the surprise recall test at 8 p.m. on the same day. Neither group knew that they were returning to take a memory test; rather, all participants were led to believe that they were returning to take part in a new and unrelated experiment.

In the study phase, participants were instructed that category-exemplar pairs would be presented on a computer screen and that they should study the pairs in preparation for a later memory test. Each category-exemplar pair was presented in uppercase letters in the center of the screen for 5 s. Order of presentation was semirandomized; exemplars from the same category did not appear on consecutive trials. When participants had completed the study phase, the experimenter distributed retrieval-practice booklets. Participants believed that this second phase was the memory test. Each page in the booklet showed one of the category names studied previously and the first two letters of one member of that category, also studied previously. Participants were instructed to complete the exemplar fragment with one of the words they had studied earlier. They were informed that some of the exemplars might be tested more than once and that in those cases they should respond with the remembered item. Rp+ items were repeated three times. At the end of the retrieval-practice phase, the booklets were collected, and participants were sent home for 12 hr. When they returned to the laboratory, they were given cued-recall booklets, in which the name of one of the previously studied categories appeared at the top of each page. Participants were instructed to recall as many examples as they could for each category in the 10-min period allocated for this test. They were instructed to complete the pages in order and not to return to a previous category once they had turned the page in the recall booklet. Order of presentation of the target categories was counterbalanced across participants.

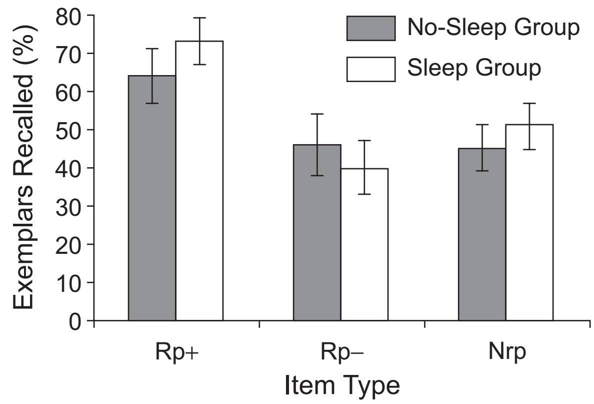

Fig. 1. Mean percentages of exemplars recalled by the sleep and no-sleep groups in Experiment 1. Results are shown separately for practiced items (Rp+), unpracticed items from practiced categories (Rp–), and unpracticed items from unpracticed categories (Nrp). Error bars show 95% confidence intervals.

Results

Planned comparisons revealed that the critical contrast of Nrp with Rp– items was reliable only in the sleep group, $t(1, 31) = –3.7$, $p_{rep} = .99$, $r = .55$ (for the no-sleep group, $t < 1$). Thus, the retrieval-practice effect was observed only in the sleep group (see Fig. 1 for mean percentages).[2] An independent t test revealed that there was no reliable difference between the two groups' recall of Rp+ items $t(62) = –1.12$; the long-term beneficial effect of selective practice (relative to baseline—i.e., Nrp items) was similar in the two groups. Debriefing interviews uncovered no evidence of conscious, intentional rehearsal in either group, and participants indicated that they were generally surprised by the delayed cued-recall test.

Experiment 2

A problem with the retrieval-practice procedure is that although it may induce inhibition of Rp– items, performance on Rp– items must almost certainly also be impaired by output interference from Rp+ items. Given that we were primarily interested in the effects of sleep on memory performance in the retrieval-practice procedure, this was in some respects a secondary issue. Nevertheless, in order to reduce the potential effects of output interference, and also to further examine the effects of sleep on retrieval practice, we decided to run a replication of Experiment 1 in which output was more directly controlled. To achieve such control, we constructed a new study set in which the first letter of each word was unique within its category. At test, participants were cued with the category names and the first letters of studied items. Using these cues, we were able to control the order in which items were recalled. In addition, to control for potential time-of-day effects, we included a new control group who studied and practiced items at 8 a.m. and were then given the surprise recall test 1 hr later; we refer to this group as the morning no-sleep group. We reasoned that if the morning no-sleep group showed the retrieval-practice effect, then this effect might be attributable to the time of day of the test, rather than a period of sleep intervening between study and test.

Method

Participants. A new cohort of 96 undergraduate Hungarian students from the Budapest University of Technology and Economics (48 females, 48 males) participated in return for partial credit in an introductory psychology course. Their ages ranged from 20 to 28 years. There were 32 subjects in each of the three groups (16 females and 16 males randomly assigned within gender to each group). All participants in the sleep group and in the morning no-sleep group were tested a minimum of 1 hr after awakening.

Materials. Following Anderson et al. (1994), we used 10 categories, 2 of which were fillers. Each target and filler category consisted of 6 exemplars (as in Experiment 1). The exemplars were moderate- to high-frequency words drawn from two Hungarian word-frequency norms (Füredi & Kelemen, 1989; Kónya & Pintér, 1985). For each subject, 4 target categories were practiced and 4 were nonpracticed; across subjects, each target category was equally often practiced and nonpracticed. The practiced and nonpracticed exemplars from practiced categories were counterbalanced over participants. The study list contained 6 words from each of the 10 categories, for a total of 60 words. Within each category, every word had a unique initial letter, so that a category name and first letter could serve as a specific cue for each target word.

Procedure. The procedure was the same as in Experiment 1 with the exception of changes in the cued-recall test. In that test, the cues appeared on a computer screen one at a time for 5 s each. Each cue consisted of a category name together with the first letter of one of the studied exemplars (e.g., "Fruit – O______"). Items for each category were presented as a block; the cues for the three Rp– words were always presented first, in random order, and the cues for the Nrp and Rp+ items were then presented ungrouped and in random order. Participants wrote their responses in a response booklet.

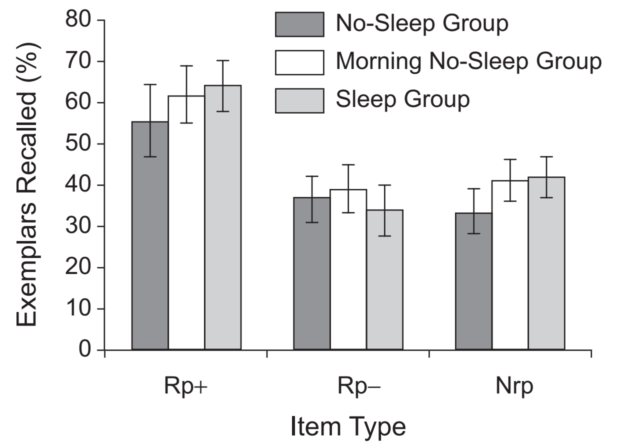

Fig. 2. Mean percentages of exemplars recalled by the sleep, no-sleep, and morning no-sleep groups in Experiment 2. Results are shown separately for practiced items (Rp+), unpracticed items from practiced categories (Rp–), and unpracticed items from unpracticed categories (Nrp). Error bars show 95% confidence intervals.

Results

Planned comparisons found that the critical contrast of Nrp with Rp– items was reliable in the sleep group, $t(1, 28) = –2.43$, $p_{rep} = .95$, $r = .43$, but not in the no-sleep group and the morning no-sleep group, *t*s < 1.2. Thus, the retrieval-practice effect was observed only in the sleep group (see Fig. 2). A one-way independent analysis of variance found no reliable difference between groups on recall of Rp+ items, $F < 1.2$, showing that the long-term beneficial effect of selective practice was present to the same degree in all groups (see Fig. 2). The debriefing interviews again indicated that participants did not rehearse items in the retention interval, and that all participants were surprised by the memory test. In sum, the overall pattern of findings replicated the pattern observed in Experiment 1 and indicates that output interference and time-of-day differences had little, or possibly no, influence in the two experiments.

Discussion

One account of the effects of retrieval practice posits that they are mediated by an episodic memory of the study phase (Racsmány & Conway, 2006). According to this view, which we term the *episodic-inhibition hypothesis* to distinguish it from accounts focusing on other types and sources of inhibition in long-term memory, retrieval practice establishes a pattern of activation and inhibition over the contents or features of an episodic memory of the study phase. As the episodic memory is consolidated in long-term memory, the pattern of activation and inhibition, which determines the accessibility of the contents of the memory, stabilizes and becomes resistant to further change. One major mechanism of this process of consolidation is rehearsal. According to the episodic-inhibition hypothesis, as a memory is repeatedly retrieved and its contents are accessed, its durability in long-term memory increases, and the accessibility levels of its contents become fixed (Racsmány & Conway, 2006; Racsmány, Conway, Garab, & Nagymáté, 2008).

The present findings suggest that sleep is important to this process of consolidation, as indeed other researchers using different procedures have also observed (e.g., Drosopoulos, Wagner, & Born, 2005). The findings of Experiment 2 indicate that retrieval-practice effects begin to dissipate after a retention interval of just 1 hr in the absence of rehearsal. Interestingly, in a related experiment not reported here, we found that if there is rehearsal in the retention interval, then retrieval-practice effects can be maintained over at least 12 hr (with no period of sleep). We use the term *rehearsal* here in a slightly nonstandard way, as according to our episodic-inhibition view, rehearsal occurs when a memory is activated and the pattern of activation and inhibition over its contents is instantiated. Such rehearsal does not have to occur consciously or intentionally, although, of course, it might. We suggest that when rehearsal occurs in this way, it approximates what has been termed *elaborative rehearsal* (Craik & Lockhart, 1972), and it promotes the integration of the memory with other memories and knowledge structures in autobiographical memory (see Conway, 2009). It is perhaps the degree and nature of the integration that determines the durability of access to a memory and its contents. Clearly, other memories formed during the retention interval may reduce integration, prevent it, or interfere with it in some other way. It seems likely that the opportunity for interference by new memories was greater in our no-sleep than in our sleep groups, and, consequently, integration may have been attenuated in the no-sleep relative to the sleep groups. (Note that this would not have been the case if rehearsal had been intentionally undertaken during the retention interval.)

According to this reasoning, the greater degree of integration of memories in the sleep groups underlies the long-term retrieval-practice effects we observed in these groups. This integration is, perhaps, similar in kind to the integration of new words with the lexicon found to occur after periods of nocturnal sleep (Dumay & Gaskell, 2007).

These novel long-term, sleep-related, retrieval-practice effects lend some support to suggestions that spontaneously occurring retrieval practice in everyday cognition may mediate aspects of remembering and forgetting (e.g., Anderson, 2001). But we can now add to this idea the notion that consolidation and integration processes occurring during sleep are also important in maintaining access to memories and their contents. The present findings demonstrate that consolidation of recently formed episodic memories during sleep may be integral to the normal functioning of episodic memory.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interests with respect to their authorship and/or the publication of this article.

Funding

Financial support for this research was provided by OTKA (Hungarian National Science Foundation) Grants K68463 and IN77932. Martin A. Conway is supported by a Professorial Fellowship (RES-051-27-0127) from the Economic and Social Research Council of Great Britain. Mihály Racsmány is grantee of the Bolyai János Research Scholarship of the Hungarian Academy of Science.

Notes

1. A very important design feature is that the memory test is unexpected. In a separate experiment not reported here, we found that long-term retrieval-practice effects can occur if participants rehearse the retrieval-practice items during the retention interval.
2. An alternative interpretation might focus on diurnal effects, such as the awakening cortisol response (ACR), which is thought to influence memory. However, as the ACR peaks and then begins to decline within 30 to 45 min following sleep (Clow, Thorn, Evans, & Hucklebridge, 2004), and all participants were tested at least 1 hr after awakening (and most were tested 90 to 120 min postsleep), it seems unlikely that the ACR could have directly influenced memory performance in the sleep group. Moreover, although cortisol levels begin to fall toward the onset of sleep and are at their lowest levels in the first 3 to 4 hr of sleep, all participants in the no-sleep group were tested several hours prior to sleep, and there is no reason to suppose that their cortisol levels had changed systematically at this point in the sleep/wake cycle. Thus, the sleep and no-sleep groups most likely had highly similar diurnal cortisol levels.

References

Anderson, M.C. (2001). Active forgetting: Evidence for functional inhibition as a source of memory failure. *Journal of Aggression, Maltreatment and Trauma*, *4*, 185–210.

Anderson, M.C., Bjork, E.L., & Bjork, R.A. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063–1087.

Anderson, M.C., & Levy, B.J. (2007). Theoretical issues in inhibition: Insights from research on human memory. In D.S. Gorfrein & C.M. MacLeod (Eds.), *Inhibition in cognition* (pp. 81–103). Washington, DC: American Psychological Association.

Anderson, M.C., & Spellman, B.A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review*, *102*, 68–100.

Bjork, E.L., Bjork, R.A., & Anderson, M.C. (1998). Varieties of goal-directed forgetting. In J.M. Golding & C.M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103–139). Mahwah, NJ: Erlbaum.

Camp, G., Pecher, D., & Schmidt, H.G. (2007). No retrieval-induced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 950–958.

Clow, A., Thorn, L., Evans, P., & Hucklebridge, F. (2004). The awakening cortisol response: Methodological issues and significance. *Stress*, *7*, 29–37.

Conway, M.A. (2009). Episodic memories. *Neuropsychologia*, *47*, 2305–2313.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Drosopoulos, S., Wagner, U., & Born, J. (2005). Sleep enhances explicit recollection in recognition memory. *Learning & Memory*, *12*, 44–51.

Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*, 35–39.

Euston, D.R., Tatsuno, M., & McNaughton, B.L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, *318*, 1147–1150.

Füredi, M., & Kelemen, J. (1989). *A mai magyar nyelv szépprózai gyakorisági szótára* [A frequency dictionary of the literary language of Hungarian]. Budapest, Hungary: Akadémiai Kiadó.

Kónya, A., & Pintér, G. (1985). Kategória norma a verbális emlékezet vizsgálatához [Category norms for verbal memory research]. *Hungarian Psychological Review*, *2*, 93–111.

MacLeod, M.D., & Macrae, C.N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, *12*, 148–152.

Mensink, G.J.M., & Raaijmakers, J.W. (1988). A model of interference and forgetting. *Psychological Review*, *95*, 434–455.

Norman, K.A., Newman, E.L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, *114*, 887–953.

Racsmány, M., & Conway, M.A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 44–57.

Racsmány, M., Conway, M.A., Garab, E.A., & Nagymáté, G. (2008). Memory awareness following episodic inhibition. *Quarterly Journal of Experimental Psychology*, *61*, 525–534.

Saunders, J., & MacLeod, M.D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied*, *8*, 127–142.

Stickgold, R., & Walker, M.P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences*, *28*, 408–415.

Storm, B.C., Bjork, E.L., Bjork, R.A., & Nestojko, J.F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review*, *13*, 1023–1027.

Walker, M.P., Brakefield, T., Morgan, A., Hobson, J.A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, *35*, 205–211.

Williams, H.L., Conway, M.A., & Baddeley, A.D. (2008). The boundaries of episodic memories. In T.F. Shipley & J.M. Zacks (Eds.), *Understanding events: How humans see, represent, and act on events* (pp. 589–617). New York: Oxford University Press.

Wilson, M.A., & McNaughton, B.L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*, 676–679.