Influence of a Sleeping Versus Waking Retention Interval on Spatial, Visual, and Auditory Memory Performance

We investigated the influence of a retention interval spent sleeping or waking on participants' performance in spatial, auditory, and visual tasks. Using Jenkins and Dallenbach's research (1924) as a paradigm, we replicated and extended the original study using a 2 x 3 mixed design with repeated measures. The 2 independent variables were the activity during the retention interval (i.e., sleeping or waking) and the 3 types of memory tasks (i.e., spatial, auditory, and visual). Fifty-seven undergraduate students participated in 2 sessions. Results indicate that a retention interval spent sleeping had a beneficial effect on auditory memory performance. We did not find a significant effect for visual and spatial memory performance, but attribute this to ceiling effects within the experimental design.

Memory retention has been the subject of much psychological experimentation and investigation. Researchers have investigated both how memories are formed from a given stimulus and how those memories are then retained. However, it was only in the early 20th century that memory was paired with sleeping for investigation and study. Researchers began to focus on how sleep, both its length and quality, affects memory formation and retention. Evidence has emerged that suggests an intricate link between memory consolidation and sleep length and quality.

Memory research proposes that activities immediately after learning have a large impact on the retention of the learned material. Sleep is believed to facilitate memory consolidation because it prevents interference from novel stimuli (Benson & Feinberg, 1977). However, sleep investigators propose that interference alone cannot explain the beneficial effects of sleep on memory consolidation: the intrinsic characteristics of sleep appear to help to consolidate and retain memories (Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000).

Using Ebbinghaus' memory research as their foundation (as cited in Jenkins & Dallenbach, 1924), Jenkins and Dallenbach (1924) compared the rate of forgetting during sleep and waking. Jenkins and Dallenbach asked their two participants to learn a series of nonsense syllables to the point of mastery and then asked them to recall those syllables at intervals of waking ranging from 1, 2, 4, or 8 hours and randomly varied. Jenkins and Dallenbach's study demonstrated that recall was twice as effective after sleeping intervals when compared to recall after waking intervals. Their research also indicated that this difference between waking and sleeping intervals became increasingly pronounced as the interval length increased. The average number of reproductions for both 4-hour and 8-hour recall intervals of waking was significantly lower than responses at 2-hour waking intervals, whereas responses after intervals of sleep were maintained at 2, 4, and 8 hours.
hours. This study supports the hypothesis that sleep improves memory retention compared to waking.

Benson and Feinberg (1977) extended Jenkins and Dallenbach’s research to focus on time intervals and the forgetting curve, and they challenged Jenkins and Dallenbach’s proposal that sleeping merely prevents interference and thus improves retention through insulation. Using a paired associates list as their learning task, Benson and Feinberg systematically replicated Jenkins and Dallenbach’s study. Their sample size was significantly larger (N = 60). Their retention intervals were set at 8, 16, and 24 hours. Because both groups were allowed to sleep after the 8-hour interval, the 16- and 24-hour intervals allowed researchers to test the enhancing effect of sleep on memory. If the Jenkins and Dallenbach conclusions were correct, all participants would show improved recall when tested at 16- and 24-hour intervals. In fact, Benson and Feinberg’s results confirmed Jenkins and Dallenbach’s findings: sleep onset shortly after learning verbal materials had a beneficial effect on recall in the 8- to 24-hour period following learning.

Koulack (1997) raised the question of how circadian rhythms affect memory consolidation. Koulack proposed that circadian rhythms increase a participant’s vigilance and consequently play a critical role in aiding memory consolidation. Koulack showed 40 male participants a series of 40 words individually and at regular intervals. Koulack then showed his participants 80 words, 40 new and 40 that they had previously seen. Participants were instructed to indicate whether the word on the screen was “new” or “old.” Koulack’s results showed that afternoon learners had higher scores than morning learners. Koulack’s research also showed that participants who slept had higher scores than those who remained awake. Koulack’s study therefore confirmed Jenkins and Dallenbach’s (1924) early research, reinforcing the idea that sleeping improves memory consolidation.

Ficca, Lombardo, Rossi, and Salzarulo (2000) addressed the role of REM sleep in memory processes. Ficca et al.’s research compared the effects of disorganized sleep cycles and sleep discontinuity on the recall of verbal material in young adults. These researchers hypothesized that regular occurrences of NREM-REM cycles are critical to the retention of verbal material learned just prior to sleep onset. The study’s 12 participants were given a verbal recall task using 20 unrelated word pairs. Ficca et al. found that “sleep efficiency” (i.e., the quantitative amount of intranight wakefulness) did not significantly affect recall. However, their results suggested that the organizational quality of sleep, defined as regular cycles of NREM-REM sleep, is more important than continuity for memory tasks. Ficca et al. were careful to highlight that their findings are applicable only to verbal material. This disruption may not be present when other learning modalities are tested.

Stickgold et al. (2000) addressed how sleep benefits visual discrimination tasks. Using a variation of Karni and Sagi’s (1991) visual discrimination task (as cited in Stickgold et al., 2000), Stickgold et al. showed their participants a series of screens, each with slightly different images, with a 0.40-msec interval. Participants were then asked to report what image they saw on the screen. In conjunction with the learning task, Stickgold et al. analyzed their participants’ sleeping patterns. Stickgold et al. found that performance on a visual task improved if the participant had a minimum of 6 hours of sleep prior to retesting. Slow-wave sleep (SWS) and REM are critical in consolidating memory for a visual task. Stickgold et al. proposed that it is the sequence of events mediated by SWS and REM sleep that improves performance on a visual task. Although their test could not empirically determine how sequential events during SWS and REM facilitate memory consolidation, they proposed that SWS allows memories to transition from the hippocampus into the neocortex. REM sleep then facilitates the strengthening of these memories in the neocortex and the formation of new associative memories.

The present study merges these two notions to investigate how sleep and waking affect memory retention. In a replication and extension of previously conducted research (Jenkins & Dallenbach, 1924), we hypothesized that learning, followed by a sleep interval of at least 6 hours, is more effective in fostering memory retention than when learning is followed by a waking interval of the same length.

**Method**

**Participants**

Fifty-seven students enrolled in undergraduate psychology courses at a small, private, northern California university volunteered to participate. Participants ranged in age from 18 to 22 years. Volunteers were recruited via the psychology department’s research participation pool and through psychology faculty advisors. All participation was voluntary, and students received participation units in their introductory psychology class for their participation or extra credit in their psychology courses. As an incentive to increase interest in subject material, the eight highest-scoring participants were awarded $5 at the completion of our study. All participants were treated in accordance with the “Ethical Principles of Psychologists and Code of Conduct,” and confidentiality was maintained (American Psychological Association Ethics Committee,
1992). For the purpose of data analysis, only participants in the sleeping group who reported getting at least 6 hours of sleep were included. For participants in the waking group, only those who reported that they did not nap during their waking interval were included in the data analysis.

Twenty-eight participants' scores were used in the data analysis: 14 participants in the sleeping group and 14 participants in the waking group.

Materials
An adapted version of the Rey complex figure was used for the spatial learning task (from Kolb & Whishaw, 1996; see Appendix). Pilot testing indicated a modified version of this figure to be of appropriate difficulty for students to learn in a period of 3 minutes. The figure was enlarged to the size of 26 x 17½ in. (66 x 44¼ cm). Blank 8½ x 11 in. (22 x 28 cm) sheets of paper were used as practice materials. The response materials were sheets of standard, letter-sized paper with brief directions instructing students to replicate the figure as completely and accurately as possible from memory.

We used a previously published, paired associates list for our auditory task (see Searleman and Hermann, 1994). This list was appropriate because previously published research indicated it to be composed of semantically unrelated word pairs that were of sufficient familiarity to our pool of college-age participants. The response sheet, printed on standard, letter-sized paper, listed brief directions and had a printed line corresponding to each prompted response.

Visual stimuli consisted of 32 black-and-white photos of faces. Each photo was enlarged to 8½ x 14 in. (22 x 35.56 cm) sheets on high-quality card stock. The photographs were taken from the recruitment bulletin of a northern California professional art school. All photos showed only the individuals' faces, not including their hair or neckline. Faces varied in age, gender, and ethnicity. Facial expressions were varied as well. The response sheets for this task listed letters A and B, which directly correlated to each pair of faces presented. Directions printed atop the sheet instructed participants to circle the letter corresponding to the photo seen in the original 16 faces presented during the learning portion of the procedure.

Several questionnaires were used. Two previously published surveys were printed on standard, letter-sized paper. These were the Owl and Lark questionnaire (from Horne & Ostberg, 1976) and the Epworth Sleepiness Scale (from Johns, 1991). These questionnaires were chosen because previous research shows they are informal measures of circadian rhythms and sleep debt. In addition, a set of questions was designed to investigate how much participants slept between the two sessions, if they exercised, if they thought about the experiment between sessions, and if their sleep followed a regular pattern.

Design
The study entailed a 2 x 3 mixed design with repeated measures. The two independent variables were the activity during the retention interval (i.e., sleeping or waking) and the three types of memory tasks (i.e., spatial, auditory, and visual). Activity during the retention interval was a between-subjects factor that was manipulated by assigning one group of participants to sleep for the majority of time between the learning and testing of the three memory tasks, and one group of participants to remain awake for the majority of this interval. Participants selected which group to attend; however, they were not aware of the varied conditions between each group. Repeated measures were used to score each participant’s performance on each memory task during each session attended.

For all participants there was a 10-hour interval between the first learning session and the second recall session. Furthermore, final data analysis included only those participants who reported sleeping at least 6 hours subsequent to learning. Those in the waking condition spent the entire interval between the learning and testing awake; participants who napped between learning and recall were removed from data analysis. Participants reported their sleeping habits on questionnaires distributed at the conclusion of our final session.

For the memory tasks we administered a task involving spatial processing, one involving visual processing, and one involving auditory processing. This within-subject element of the design allowed us to compare memory performance on three types of tasks before and after the retention interval.

To counterbalance possible sequencing effects that might result from presenting each task to participants in a uniform order, we used block randomization. There were six possible testing sequences for our study. We randomly assigned these sequences in blocks of six. For example, because there were 12 groups in our first week of experimentation, each testing sequence appeared twice. We also kept the task order constant between the initial learning and final testing sessions within each particular group. Therefore, participants were tested in the same order in which they had previously learned the tasks.

The dependent variable in the present study was participants’ performance on three individual memory tasks. Researchers obtained these scores 10 hours after
participants initially learned the tasks. By allotting a designated number of points for the correct responses, we were able to quantify participants’ performance on each of the memory tasks. Therefore, a score of 100% signified that an individual obtained the full number of points possible within a particular task and thus correctly remembered all elements within it.

Many variables were held constant across both groups to ensure that any observed variations on task performance were due to the manipulation of the independent variable and not to any extraneous differences existing between groups. For example, the sex and number of experimenters administering the memory tasks at both learning and testing were kept constant. We also controlled the instructional format within groups, as experimenters of all participant groups read aloud prescribed instructional sheets that designated a time limit and clear directions for each of the memory tasks. Moreover, experimenters administered all tasks in an identical fashion, varying only the order in which each was presented. For example, experimenters showed the faces used to test visual memory in a fixed order within all groups, varying them in the same manner for both Session 1 and Session 2. All of these set conditions ensured that each participant received identical instructions, guidelines, and task arrangements. Thus every participant experienced equivalent experimental procedures.

Procedure

Experiments were conducted in groups of between 1 and 16 participants. All participants attended both sessions. During the first session participants were given three memory tasks: a spatial memory task, an auditory memory task, and a visual memory task. After the tasks were administered, experimenters tested participants’ performance on these tasks. During the second session, held 10 hours after the first session, participants were again tested on the three memory tasks in the same order. To minimize demand characteristics, participants were told that our study was investigating learning environments and memory. At no point during the testing procedure were participants aware of the sleeping manipulation.

First session. For the procedural memory task, participants were shown an adapted version of the Rey complex figure for 3 minutes (Kolb & Whishaw, 1996; see Appendix). During these 3 minutes, participants were given the opportunity to practice drawing the figure on scratch paper. After the 3 minutes, the experimenter removed the scratch paper and Rey figure. Participants then drew the figure as accurately and completely as possible from memory on the paper provided. Once participants had completed the figure, the experimenter showed the Rey figure a second time to provide participants with the opportunity to see how accurately they had reproduced the figure. To score this task, the experimenter awarded 1 point for every line a participant drew correctly. One point was deducted for every incorrectly placed line (i.e., line in the incorrect orientation, or participant added lines that were not part of the figure shown). A total of 60 points could be earned on this task.

The auditory memory task consisted of 12 paired associates. During the learning phase of this task, the experimenter read 12 pairs of words to participants (e.g. aardvark, table). The experimenter told participants that the first word of each pair was the prime word, and the second word of each pair was the prime’s complement. During the testing phase, the experimenter read the 12 prime words (e.g. aardvark), and participants were asked to provide the complement word (e.g. table) on the testing sheet. After the testing phase, the experimenter read the 12 word pairs so participants could determine which, if any, incorrect word complements they had provided. To score this task, the experimenter awarded participants 1 point for every correct word complement they recalled for a possible score of 12 points.

To test visual memory, we gave participants a face recognition task. During the learning phase of this task, participants viewed pictures of 16 individual faces for a period of 5 seconds each. During the testing phase of this task, the experimenter showed participants eight pairs of faces again for a period of 5 seconds each. Each experimenter positioned herself where all participants could easily see the photograph. Each pair consisted of a face that participants saw during the showing of the original 16 pictures and a new face that participants had not yet seen. Participants were instructed that the picture the experimenter held in her right hand corresponded to the letter A on their testing sheet, whereas the picture held in the experimenter’s left hand corresponded to the letter B. Participants identified the face that they had seen during the learning phase of the task by circling either A or B on their testing sheet. After the testing phase of this task, the experimenter showed the original 16 faces for a second time. To score this task, the experimenter awarded 1 point for every correct face participants identified. A total of 8 points could be earned on this task.

Second session. Ten hours after the first session, participants returned to the testing site to complete the experiment, fill out three short questionnaires, and receive debriefing. Participants performed the memory tasks in the same order they received them in the first session. The purpose of the second testing ses-
sion was to determine participants’ recall and performance on the three memory tasks 10 hours after initial learning. To test recall for visual memory, participants completed the Rey figure they saw during the first session as accurately and completely as possible from memory. Participants were given 3 minutes to complete the task. Again, the experimenter awarded participants 1 point for every correctly drawn line and subtracted 1 point for every incorrectly drawn line.

To test recall for auditory memory, the experimenter read the same 12 prime words that participants heard in the first testing session. Upon hearing the prime word, experimenters prompted participants to provide the complement to the prime, as remembered on the space provided on the testing sheet.

The experimenter awarded participants 1 point for every correct word complement they recalled, for a possible 12 points. To test recall for visual memory, participants again viewed eight pairs of faces. Participants had seen eight of the faces during the first session (they were part of the original 16 faces shown), whereas eight of the faces were new to participants. Experimenters instructed participants to identify which face they had seen before by marking A or B on the testing sheet provided. Again, face A was held in the experimenter's right hand and face B was held in the experimenter's left hand. To score this task, the experimenter awarded 1 point for every correct face recognized, for a total of 8 possible points.

After participants completed the three tasks, they completed three short surveys. One survey asked participants about their lifestyle habits, including how many hours of exercise they engaged in on the day of testing, the times they went to bed and woke up, and how stable they considered their sleep habits to be. Experimenters also distributed the Epworth questionnaire (from Johns, 1991) out of 42 possible points. Participants who reported higher scores suffered more sleep debasement. We used the Owl and Lark questionnaire (from Horne & O'Keefe, 1976) to determine what time of day individual participants were most alert. We calculated responses as ratios out of 38 points possible. Those participants who scored close to zero (i.e., 8/38) were considered to be more alert during the nighttime. Those approaching one (i.e., 36/38) were determined to be more alert during the morning hours.

For the lifestyle habits questionnaire, we coded all responses numerically. The “thinking of study” question produced responses from 0 to 4, with 0 representing participants who did not think of the study at all between sessions and 4 being participants who reported thinking about the study many times during the retention interval between sessions. The sleep quality question produced numeric responses on a scale of 1 to 5, with 1 representing very disrupted and inconsistent sleep and 5 signifying a restful night's sleep. All responses pertaining to time were converted into minutes for statistical evaluation.

Results

Scoring Procedure

Measuring task performance. We collaborated to define procedural and scoring criteria for all tasks and questionnaires prior to scoring any data. We then divided the data by task and scored each task independently. Each researcher was assigned one specific task to score. To maintain continuity in scoring, only one researcher scored all tasks. Both portions of the auditory task were scored on a scale of 1 to 12, awarding 1 point for each correct response with 12 points possible. The visual task was similarly scored, with the total number of recognized faces in proportion to the total number of face pairs shown. Therefore, if a participant correctly recognized seven faces in the eight pairs, his or her total score was 7/8 for that visual task. The spatial task consisted of 28 lines, allowing for a total of 28 possible points. One point was awarded for each correctly demarcated line. We quantified the perimeter of the figure to be composed of four lines, each side representing one line. Four points were then awarded if the figure was correctly divided into eight right triangles of equal area. Participants earned the remaining 20 points if all other line segments in the interior of the figure were correctly placed. The circle and three dots in the second quadrant of the figure each represented 1 point. One point was deducted for intrusions and for lines drawn in an incorrect orientation. We calculated final scores for all tasks as proportions with the number correct over the total number of responses possible.

Scoring questionnaires. We scored responses to the Epworth questionnaire (from Johns, 1991) out of 24 possible points. Participants who reported higher scores suffered more sleep debasement. We used the Owl and Lark questionnaire (from Horne & O'Keefe, 1976) to determine what time of day individual participants were most alert. We calculated responses as ratios out of 38 points possible. Those participants who scored close to zero (i.e., 8/38) were considered to be more alert during the nighttime. Those approaching one (i.e., 36/38) were determined to be more alert during the morning hours.

For the lifestyle habits questionnaire, we coded all responses numerically. The “thinking of study” question produced responses from 0 to 4, with 0 representing participants who did not think of the study at all between sessions and 4 being participants who reported thinking about the study many times during the retention interval between sessions. The sleep quality question produced numeric responses on a scale of 1 to 5, with 1 representing very disrupted and inconsistent sleep and 5 signifying a restful night's sleep. All responses pertaining to time were converted into minutes for statistical evaluation.
retention interval. We eliminated participants in the waking group if they scored greater than 8 on the Epworth Sleep Debriefness scale. We eliminated participants in the sleeping group if they reported sleeping less than 360 minutes during the retention interval. After eliminating these participants, 14 participants in each group were used for statistical analysis.

**Data Analysis**

We used a multivariate analysis of covariance (MANCOVA) to determine the influence of retention interval activity on performance on each task at Time 2 while controlling for (a) task performance at Time 1 and (b) the intercorrelations in task performance across the three tasks. The mean performance at Time 2 for an auditory task was significantly greater for those participants who slept during the retention interval ($M = 0.71$) than for those who remained awake ($M = 0.43$), $F(1, 27) = 6.00, MSE = 0.013, p = .022$.

For the visual task, the MANCOVA demonstrated that a sleeping interval when compared with a waking interval did not affect participants’ ability to recognize a given stimulus correctly (Time 2) ($M = 0.94$), when controlling for performance at Time 1 and the intercorrelations in task performance across the three tasks, $F(1, 27) = 0.044, MSE = 0.022$. For the visual task, a comparison of mean scores at Time 1 and Time 2 shows that there was no significant difference in memory performance regardless of retention interval activities. Thus, there was no significant effect of the retention interval activity on performance at Time 2 for the visual task. See Table 1 for a summary of mean values for all tasks.

Finally, the MANCOVA indicated that a sleeping interval when compared with a waking interval did not affect participants’ ability to reconstruct the previously seen figure (Time 2) ($M = 0.94$) when controlling for spatial task performance at Time 1 and the intercorrelations in task performance across the three tasks, $F(1, 27) = 0.279, MSE = 0.019$. In comparing mean scores at Time 1 and Time 2, there was no significant difference in memory performance for the spatial task. There was no significant effect of the retention interval activity on performance at Time 2 for the visual task.

A retention interval spent sleeping resulted in superior cued recall for the paired associates task but not a visual or spatial task when compared to a retention interval spent awake. Mean performance at Time 2 of the sleeping group was significantly greater than mean performance at Time 2 of the waking group on the auditory task. No such differences were found for the spatial or visual tasks.

Statistical analysis thus demonstrated that a retention interval spent sleeping when compared to an interval spent awake was beneficial to cued memory recall for an auditory task, but did not significantly affect performance for mean recall on a spatial task or mean recognition for a visual task.

**Discussion**

The results partially confirm the prediction that sleeping does have a positive effect on memory retention. That is, for the auditory task, results indicate that sleeping after a period of learning yields higher task performance at Time 2 when compared to an interval spent awake. However, the results do not support the same hypothesis when applied to the visual and spatial tasks. Performance at Time 2 on the visual and spatial tasks was not significantly affected by a sleeping interval.

Results of the current study support Jenkins and Dallenbach’s (1924) early findings: auditory task performance at Time 2 was greater after a period sleeping than after a period spent awake. However, whereas the Jenkins and Dallenbach paradigm utilized nonsense syllables, our auditory task involved the use of paired associates. Our results are, however, inconsistent with Koullack’s (1997) findings. Using word recognition as his learning task, Koullack found sleep to facilitate memory performance when compared to waking. Whereas Koullack’s design involved a recognition task similar to our visual task, the present findings do not directly coincide with his results. Perhaps the current visual and spatial tasks may have produced a ceiling effect in performance. That is, the tasks may not have been sufficiently challenging; participants did not show a marked difference in task performance at Time 1 and Time 2. In future studies, pilot testing to measure the difficulty of tasks would be helpful in preventing the ceiling effect we believe was observed.

Similar to Koullack (1997), Stickgold et al. (2000) investigated how a sleeping interval affects visual memory performance. Stickgold et al. found that performance on a visual task improves if the participant...
has a minimum of 6 hours of sleep prior to retesting. We attribute this discontinuity between the results produced by the present study and those of the Stickgold et al. study to potential ceiling effects for the visual task. Although researchers ensured that participants in the sleeping group slept a minimum of 6 hours during the retention interval, tasks may not have been sufficiently challenging to allow sleep to show a significant effect.

For future research, pilot studies should be used to ensure sufficient task difficulty and thus eliminate the possibility of a ceiling effect. For example, future researchers may want to use the original Rey complex figure (Kolb & Whishaw, 1996) rather than our simplified version. Also, the visual task could be made more difficult by showing participating more faces and reducing the individual differences between faces. For example, if all faces were Caucasian women between the ages of 20 and 30 years, participants would likely have a more difficult time distinguishing between faces they had seen before and new faces.

The experimental design could be further strengthened with a larger sample size. Because the majority of our participants were derived from a university participation pool, it was difficult to obtain a large and diverse population sample. A larger participation pool may have increased the noticeable effect of our manipulation and would have allowed our findings to be generalizable to a larger population. Also, because participants partook in the experiment on a voluntary basis and did not have adequate incentive to make scheduling sacrifices to accommodate our experimental time-table, random assignment could not be utilized in our study. Because there were many potential individual differences that could have impacted our results, random assignment should be used in future research to strengthen the experimental design. Financial compensation for participation could have offered participants a greater incentive to attend sessions, despite their inconvenient times.

Results support our initial hypothesis that sleeping, when compared to waking, has a positive effect on memory performance for an auditory task. Because we did not find significant results for the spatial or visual tasks, it may be concluded that the nature of the task is an important element to consider when studying retention level activity.

Theoretical implications of the results of our study lend support to the notion that sleep plays a critical role in memory processing. For example, these findings lend support to the hypothesis that a significant time spent asleep aids in memory consolidation. Furthermore, for this specific participant population, namely college students, these findings seem to imply that 6 hours of sleep is critical to processing newly learned auditory information, perhaps newly presented lecture materials. A night spent cramming before an exam without sufficient sleep may do little to instill the necessary information. Rather, it seems that adequate sleep on a regular basis might be more beneficial to memory retention and recall.

References

APPENDIX