BRIEF REPORT

Effects of Sleep Deprivation on Procedural Errors

Michelle E. Stepan, Kimberly M. Fenn, and Erik M. Altmann
Michigan State University

In a large sample (N = 234), we tested effects of 24-hr of sleep deprivation on error rates in a procedural task that requires memory maintenance of task-relevant information. In the evening, participants completed the task under double-blind conditions and then either stayed awake in the lab overnight or slept at home. In the morning, participants completed the task again. Sleep-deprived participants were more likely to suffer a general breakdown in ability (or willingness) to meet a modest accuracy criterion they had met the night before. Among sleep-deprived participants who could still perform the task, error rates were elevated, and errors reflecting memory failures increased with time-on-task. The results suggest that sleep-deprived individuals should not perform procedural tasks associated with interruptions and costly errors—or, if they must, they should perform such tasks only for short periods.

Keywords: sleep deprivation, procedural error, memory maintenance

Supplemental materials: http://dx.doi.org/10.1037/xge0000495.supp

Sleep deprivation is known to affect attention and vigilance (e.g., Lim & Dinges, 2008; Norton, 1970; Wimmer, Hoffman, Bonato, & Moffitt, 1992), but questions remain about its effects on higher-order cognitive processes (see Killgore, 2010 for review). Some studies found that sleep deprivation impaired working memory (Chee & Choo, 2004; Chee et al., 2006; Choo, Lee, Venkatraman, Sheu, & Chee, 2005; Durmer & Dinges, 2005; Habeck et al., 2004; Mu et al., 2005) or executive function (Durmer & Dinges, 2005; Nilsson et al., 2005), whereas others found no effect on higher-order processes (Binks, Waters, & Hurry, 1999; Nilsson et al., 2005; Tucker, Whitney, Belenky, Hinson, & Van Dongen, 2010; Wimmer et al., 1992). Some studies even found divergent effects using the same task (Chee & Choo, 2004; Chee et al., 2006). These equivocal findings may be a function of relatively small sample sizes.

Here we studied effects of sleep deprivation on error rates in a procedural task, focusing on the role of memory maintenance. In our task, participants perform a set of steps in a prescribed sequence. Every few trials—a trial being a performed step—the participant is interrupted for about 20 s. After the interruption, the participant must resume the interrupted sequence with the correct step. To know which step is correct, the participant must remember the last step performed before the interruption. Maintaining a memory for that step is made difficult by interference from the interrupting task.

The theoretical question we address is whether sleep deprivation affects memory processes and, if so, whether the effects are mediated by global cognitive impairments or impairments of memory-related processes specifically. We isolate effects on memory processes by comparing two trial types that differ only in the need for memory maintenance. A postinterruption trial follows an interruption, whereas a baseline trial follows another trial. The processing for the two trial types is identical except that a postinterruption trial requires recall of the step performed before the interruption, and therefore depends on memory maintenance during the interruption. Both trial types should show effects of global factors, such as increased fatigue or wake state instability (Doran, Van Dongen, & Dinges, 2001; Lim & Dinges, 2008). Only postinterruption trials should show effects specific to memory maintenance. Deprivation studies involving higher level tasks (e.g., Caldwell & Ramspott, 1998; Hack, Choi, Vijayapalan, Davies, & Stradling, 2001; Thomas et al., 2000) typically do not include measures such as our baseline trials that allow effects on higher-order processes to be isolated like this. We can also measure time-on-task effects specifically on memory maintenance because there are many postinterruption and baseline trials per session. Time-on-task effects after reduced sleep are well documented in vigilance tasks (e.g., Doran et al., 2001; Van Dongen, Maislin, Mullington, & Dinges, 2003) but have not been widely studied for higher-order cognitive processes. Finally, we aimed for conclusive results by testing a large sample. To our knowledge, ours is the largest
study of the effect of sleep deprivation on higher-order cognition.

**Method**

**Participants**

Informed consent was obtained from all participants and the study was approved by Michigan State University’s Institutional Review Board. Participants were Michigan State University undergraduates who were native English speakers, had never been diagnosed with a memory or sleep disorder, were not color blind, and had no strong morning or evening preference (scores of 42 to 58 on the Morningness–Eveningness Questionnaire; Horne & Östberg, 1976) or major sleep disturbances (scores of 0 to 10 on the sleep disturbance section of the Pittsburgh Sleep Quality Index; Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). On the night before the study, they slept at least 6 hrs and woke by 9:00 a.m. They also refrained from napping prior to their arrival in the laboratory. They were permitted no caffeine, alcohol, or drugs for 24 hr prior to the study.

Of an initial sample of 269 participants, 20 were lost to attrition, two to caffeine consumption, one to technical problems, and 12 to failure to maintain criterion accuracy (discussed in the Materials section) in Session 1. Data from the remaining 234 participants (18 to 25 years old, M_age = 19.02, SD = 1.19; 138 women) were submitted to analysis. Participants were run in groups of up to 11, across two testing rooms, and were given course credit as compensation.

**Materials**

Here we summarize key elements of the procedural task; detailed descriptions are presented in Altmann, Trafton, and Hambrick (2014, 2017). Participants performed a sequence of seven steps in a loop. The correct order of the steps is specified by an acronym, UNRAVEL. Each letter of UNRAVEL identifies a different two-alternative forced-choice decision rule to apply to a randomly generated stimulus. Figure 1 shows two sample stimuli, the seven decision rules, and the correct response to each stimulus according to each rule to illustrate the various decisions. The stimulus contains no information about the step to perform next, so participants must remember where they are in the sequence. When participants reach the “L” step they start over with the “U” step.

There were four blocks of trials per session. Each block contained 66 trials on average (SD = 12) and was interrupted 10 times at random points by a typing task in which participants had to correctly transcribe two letter strings, each consisting of the 14

---

**Figure 1.** Above: Example of two randomly generated stimuli from the UNRAVEL task (where each letter of UNRAVEL identifies a different two-alternative forced-choice decision rule to apply to a randomly generated stimulus). Below: The UNRAVEL rules that correspond to each step (letter) in the UNRAVEL acronym, and the correct keyboard responses for each rule based on the two stimuli above. The letters in boldface type represent the possible response options for each rule. Adapted from “Momentary interruptions can derail the train of thought,” by E. M. Altmann, J. G. Trafton, and D. Z. Hambrick, 2014, *Journal of Experimental Psychology: General, 143,* pp. 215–226. Copyright 2014 by the American Psychological Association. See the online article for the color version of this figure.
UNRAVEL responses (see Figure 1) presented in random order. After the second string, participants had to pick up where they left off in the UNRAVEL sequence. An interruption lasted about 20 s ($M = 22.47, SD = 6.47$). A session took about 35 min ($M = 36.13, SD = 9.25$).

We measured accuracy in terms of sequence errors and nonsequence errors. A sequence error occurred if the participant performed the incorrect sequence of the step performed immediately prior. For example, if steps “N,” “R,” “V,” and “E” were performed in succession, “V” would be a sequence error because that step should have been “A.” However, “E” would be correct because it follows “V.” A nonsequence error occurred if the participant performed the correct step but made the incorrect response for that step given the stimulus (because all 14 responses are unique, we could infer the intended step from any response). A trial was coded as correct if there was neither a sequence nor a nonsequence error.

If fewer than 70% of trials in a block were correct, the participant was instructed to be more accurate at the end of the block. We coded a session as a failure if accuracy was below 70% on two or more blocks, on grounds that the participant did not follow the instruction to be more accurate. We excluded participants who failed Session 1 from all analyses because we could not be sure they understood the task.

**Procedure**

Participants were recruited for a study on sleep deprivation and told they would either remain awake all night or be permitted to sleep. They arrived at 10:00 p.m. for Session 1 and completed sleepiness and mood assessments, as discussed in the online supplemental material. Next, participants completed the UNRAVEL task and a battery of cognitive and personality measures that were part of a separate study. After completing all tasks (~2 hrs), participants were randomly assigned to conditions. Researchers and participants were blind to condition until all evening testing was finished. Participants randomly assigned to the sleep group ($n = 112$) were given a Charge 2 activity monitor (Fibit Inc., San Francisco, CA) to track their sleep, and then given a ride home. Participants assigned to the deprivation group ($n = 122$) stayed awake overnight in the laboratory.

Deprivation participants were monitored throughout the night by two trained research assistants (see the online supplemental material for further information). They were permitted to read, do homework, watch TV/movies, play games, or engage in other quiet activities but were not permitted to engage in activities that activated the autonomic nervous system. We provided water and snacks and allowed participants to consume anything that did not contain caffeine or alcohol. Every 2 hrs, participants completed sleepiness and mood assessments. Participants were sleep-deprived for approximately 24 hr at the start of Session 2.

At 8:45 a.m. the following morning, sleep participants returned and all participants completed Session 2, which included mood and sleepiness assessments, UNRAVEL, and other cognitive tasks associated with the separate experiment. The experiment finished at 10:00 a.m., and deprivation participants were given a ride home.

**Results**

Deprivation participants failed Session 2 at a greater rate ($n = 18, 15\%$) than sleep participants ($n = 1, 1\%$). $\chi^2(1, N = 234) = 15.04, p < .001$. Thus, in some participants, sleep deprivation caused a breakdown in the ability, or perhaps the willingness, to perform the task as instructed, which they had managed to do the previous evening.

In our remaining analyses, we included only participants who passed Session 2 (sleep $n = 111$, deprivation $n = 104$). We examined sequence errors separately on postinterruption trials (trials immediately following an interruption) and baseline trials (trials following other trials). The two trial types involve identical processing except that postinterruption trials involve recall of the step performed before the interruption, and thus depend on memory maintenance.

Sequence errors are plotted in Figure 2. For postinterruption errors, an analysis of variance (ANOVA) with group (sleep, deprivation) as a between-subjects factor and session (1, 2) and block (1 to 4) as within-subjects factors showed a significant three-way interaction, $F(3, 639) = 3.73, p = .011, \eta^2 = .017$. To probe the interaction, we analyzed effects of group and block within each session. In Session 1, there was no main effect of group, $F(1, 213) = 1.51, p = .221, \eta^2 = .007$. There was a main effect of block, $F(3, 639) = 9.03, p < .001, \eta^2 = .041$, reflecting a practice effect, but no interaction, $F(3, 639) = 1.30, p = .275, \eta^2 = .006$. Thus, the two groups performed similarly on Session 1, as expected given double-blind administration of the task followed by random assignment to group.

For postinterruption errors in Session 2, there was a main effect of group, $F(1, 213) = 40.76, p < .001, \eta^2 = .161$, a main effect of block, $F(3, 639) = 13.04, p < .001, \eta^2 = .058$, and a significant two-way interaction, $F(3, 639) = 7.58, p < .001, \eta^2 = .034$. To probe the interaction, we analyzed effects of block within each group. The effect of block was not significant for sleep ($F < 1$), but was for deprivation, $F(3, 309) = 15.86, p < .001, \eta^2 = .133$. In the deprivation group, a polynomial trend analysis indicated a linear increase in errors across blocks, $F(1, 103) = 38.53, p < .001, \eta^2 = .272$, but no quadratic or cubic trend ($F < 1.12, ps > .29$). A pairwise comparison showed fewer postinterruption errors in Block 1 than Block 4, $t(103) = −6.19, p < .001, d = 0.607$, indicating worse performance at the end of the session than at the start. Thus, deprivation participants who passed Session 2 still made more postinterruption errors overall than sleep participants and progressively made more errors across Session 2.

For baseline errors, the Group $\times$ Block $\times$ Session ANOVA also showed a three-way interaction, $F(3, 639) = 3.05, p = .028, \eta^2 = .014$. To probe the interaction, we analyzed effects of group and block within each session. In Session 1, there was no main effect of group, $F < 1$. There was a main effect of block, $F(3, 639) = 12.70, p < .001, \eta^2 = .056$, reflecting a practice effect, and a marginal interaction, $F(3, 639) = 2.37, p = .070, \eta^2 = .011$, which we do not attempt to interpret. Thus, on this measure the groups again performed similarly in Session 1.

1 Nonsequence errors, and data and statistical tables for all analyses, are presented in the online supplemental material.
we analyzed the Block larger (line errors, with the effect on postinterruption errors substantially sleep-deprived participants made more postinterruption and baseline still impaired performance, in two ways (see Figure 2). First, before.

were instructed to achieve and were able to achieve the evening ability or unwilling to achieve a modest level of accuracy that they significant proportion of deprivation participants (15%) were unable or unwilling to achieve a modest level of accuracy that they achieved the evening performance (Lim & Dinges, 2008). If this mechanism had also been at work in our study, it should have affected overall performance and not just postinterruption errors. Baseline sequence errors increased with time-on-task, but the effect was not large and was similar for both groups, so we assume it was unrelated to sleep deprivation. Accordingly, wake state instability does not seem to have increased over the duration of the session and therefore does not seem to have caused the increase in postinterruption errors.

Instead, the underlying mechanism seems to be related specifically to memory maintenance. Memory maintenance is a complex process with many component mechanisms. As a possible clue to which component was affected, we note that the increase in postinterruption errors across blocks in Session 2 correlates with the number of experienced interruptions. This suggests that the underlying mechanism might be a buildup of proactive interference in episodic memory for past performance. In one theory of cognitive control, specifically this kind of buildup is a primary operational constraint on a system that has to be goal-directed yet be able to change goals (Altmann & Gray, 2008). In terms of the present results, the implication is that sleep deprivation impairs the mechanisms that manage the buildup of proactive interference. This account is broadly consistent with findings linking sleep depriva
tion to prefrontal cortex functioning (Chee & Choo, 2004; Choo et al., 2005; Drummond, Brown, Salamat, & Gillin, 2004; Drum-

For baseline errors in Session 2, there were main effects of group, $F(1, 213) = 6.96, p = .009, \eta^2 = .032$, and block, $F(3, 639) = 5.22, p = .001, \eta^2 = .024$, but no interaction ($F < 1$). The block effect reflects an increase in errors across the session. This increase is relatively small, and did not differ by group, so our interpretation is that the underlying mechanism that caused the increase in errors is not the one that caused the increase in postinterruption errors for the deprivation group. To assess this interpretation statistically, we compared the block effect for postinterruption and baseline errors in the same analysis. A Group $\times$ Block $\times$ Error Type (postinterruption, baseline) ANOVA for Session 2 showed a significant three-way interaction, $F(3, 639) = 6.85, p < .001, \eta^2 = .031$. To probe the interaction, we analyzed the Block $\times$ Error Type interaction separately for each group. The interaction was not significant for sleep ($F < 1$), but was significant for deprivation, $F(3, 309) = 12.62, p < .001, \eta^2 = .109$. This analysis supports the inference that sleep deprivation had an effect specifically on postinterruption errors, causing them to become more frequent as Session 2 progressed.

**Discussion**

In a large sample, we investigated the effects of sleep deprivation on error rates in a procedural task that requires maintenance of task-related information during interruptions. In the evening, sleep and deprivation groups performed similarly. The next morning, a significant proportion of deprivation participants (15%) were unable or unwilling to achieve a modest level of accuracy that they were instructed to achieve and were able to achieve the evening before.

Among those who achieved criterion accuracy, sleep deprivation still impaired performance, in two ways (see Figure 2). First, sleep-deprived participants made more postinterruption and baseline errors, with the effect on postinterruption errors substantially larger ($\eta^2 = .161$ vs. $\eta^2 = .032$). Second, sleep-deprived participants made progressively more postinterruption errors across Session 2. The difference in error rates between the first and last block of the session was over half a standard deviation ($d = 0.607$). Baseline errors, which are like postinterruption errors except for maintenance failures, did not show the same pattern. Thus, memory maintenance processes became more impaired as Session 2 progressed.

This time-on-task effect resembles that found in sleep-deprived performance on vigilance tasks, but the underlying mechanism seems to be different. In the vigilance context, worsening performance over time has been linked to increasing wake state instability as the homeostatic sleep drive increases (Doran et al., 2001; Lim & Dinges, 2008). If this mechanism had also been at work in our study, it should have affected overall performance and not just postinterruption errors. Baseline sequence errors increased with time-on-task, but the effect was not large and was similar for both groups, so we assume it was unrelated to sleep deprivation. Accordingly, wake state instability does not seem to have increased over the duration of the session and therefore does not seem to have caused the increase in postinterruption errors.

Figure 2. UNRAVEL (where each letter of UNRAVEL identifies a different two-alternative forced-choice decision rule to apply to a randomly generated stimulus) sequence errors: Percentage of postinterruption (solid lines) and baseline (dashed lines) errors made across the four blocks of UNRAVEL in Session 1 (left side) and Session 2 (right side) for the sleep and deprivation groups. Error bars are standard error of the mean. See the online article for the color version of this figure.


Received May 14, 2018
Revision received July 10, 2018
Accepted July 18, 2018