

Sleep on it, but only if it is difficult: Effects of sleep on problem solving

Ut Na Sio · Padraic Monaghan · Tom Ormerod

© Psychonomic Society, Inc. 2012

Abstract Previous research has shown that performance on problem solving improves over a period of sleep, as compared with wakefulness. However, these studies have not determined whether sleep is beneficial for problem solving or whether sleep merely mitigates against interference due to an interruption to solution attempts. Sleep-dependent improvements have been described in terms of spreading activation, which raises the prediction that an effect of sleep should be greater for problems requiring a broader solution search. We presented participants with a set of remote-associate tasks that varied in difficulty as a function of the strength of the stimuli–answer associations. After a period of sleep, wake, or no delay, participants reattempted previously unsolved problems. The sleep group solved a greater number of difficult problems than did the other groups, but no difference was found for easy problems. We conclude that sleep facilitates problem solving, most likely via spreading activation, but this has its primary effect for harder problems.

Keywords Insight · Problem solving · Sleep · Learning · Task complexity

Many scientific discoveries and masterpieces in art and literature were inspired by dreams, suggesting the importance of sleep in solution discovery (Cartwright, 1974). One

famous example, for instance, is Loewi's discovery of the chemical transmission of nerve impulses in his dream (Stickgold & Walker, 2004). Recent empirical studies have supported the anecdotal evidence that sleep has a profound facilitatory effect across a range of different types of problems (Cai, Mednick, Harrison, Kanady, & Mednick, 2009; Wagner, Gais, Haider, Verleger, & Born, 2004). Furthermore, Kuriyama, Stickgold, and Walker (2004) discovered that sleep has a differential effect for easier versus harder motor skill tasks: Sleep was most beneficial for greater complexity. However, it is unclear whether such distinctions in difficulty apply to more complex cognitive tasks such as problem solving. In this study, we examined the effect of sleep in terms of how task characteristics govern the effect of sleep on problem solving.

A growing body of work suggests that sleep has an effect on associations among concepts in processing and memory (e.g., Cai et al., 2009), facilitating restructuring of information (Payne, 2011; Payne et al., 2009; Stickgold, Scott, Rittenhouse, & Hobson, 1999), which is a key aspect of problem solving (e.g., Ohlsson, 1992, 2011). Using the DRM paradigm, Payne et al. reported sleep-dependent consolidation of false memories, which has been interpreted in terms of activation spreading from representations of presented words to related concepts during sleep. Stickgold et al. tested the effects of semantic priming among weakly and strongly related prime–target pairs when participants were awakened from different stages of sleep. Stickgold et al. found that on waking from rapid eye movement (REM) sleep, weakly related word pairs produced more semantic priming than did strongly related pairs, as compared with waking from non-rapid eye movement sleep or being awake. However, these three groups did not differ in the strength of priming between strongly related words. This result was interpreted in terms of spreading activation

U. N. Sio · P. Monaghan · T. Ormerod
Centre for Research in Human Development and Learning,
Department of Psychology, Lancaster University,
Lancaster LA1 4YF, UK

P. Monaghan (✉)
Department of Psychology, Fylde College, Lancaster University,
Lancaster LA1 4YF, UK
e-mail: p.monaghan@lancaster.ac.uk

among concepts facilitated by sleep, indicating that, during REM sleep, activation of the presented stimuli spread widely to more remotely associated concepts, rather than activation being confined only to close associates.

Spreading activation has been proposed as one of the primary cognitive mechanisms underlying insight (Ohlsson, 1992, 2011) and has been discussed in particular in relation to solving problems such as remote-associate tests (RATs). RAT problems require finding a word that is related to three given words (e.g., *lick*, *sprinkle*, *mine*; answer: *salt*). For RAT problems, activation is conceived to pass across a semantic associative network (Collins & Loftus, 1975) between stimulus and target words, with intermediary associates also becoming activated (Mednick, 1962), and when activation of the target word exceeds a threshold, it becomes available as a solution to the problem. Such associative networks with spreading activation are now being implemented in computational models (Hills, Jones, & Todd, 2012; Kenett, Kenett, Ben-Jacob, & Faust, 2011). Cai et al. (2009) found that when participants were implicitly presented with the solution to previously unsolved RAT problems in an unrelated lexical decision task, their solution performance improved after a short nap, as compared with a similar period of wakefulness, an effect they discuss as consequent upon sleep-enhanced spreading activation.

RAT problems vary greatly in terms of how much spreading activation is required between the stimulus and target words, which can be measured in terms of how closely associated the stimulus and target words are. If sleep-dependent improvement on RAT problems is a result of a boost to spreading activation during sleep, we should see a larger effect of sleep for problems where activation has to spread more widely across the semantic network between the stimulus and the target words.

In this study, we compared the effect of sleep on different RAT problems varying in terms of problem difficulty. The difficulty of RAT problems has been determined to be a function of how distant the associations are between stimulus and target words (Mednick, 1962). When difficult RAT problems are solved, in which the words and the answer are remotely associated, expanding the search of the problem space via spreading activation is needed in order to reach the correct associates. For easy RAT problems, in which the presented words and answer are closely associated, successful solution requires only a narrow search within the problem space (Ball & Stevens, 2009). Analogous to the research on sleep primarily affecting complex rather than simple motor tasks (Kuriyama et al., 2004), we predict that sleep will most affect more difficult RATs,

which require activation of weaker associates for solution discovery. However, we predict that performance on easy RATs should not be significantly enhanced after sleep, because sleep does not particularly facilitate access to strong associations, as compared with wakefulness (Stickgold et al., 1999).

According to research on problem solving, a delay even without sleep, between the initial and subsequent attempts in problem solving, can still facilitate performance (Sio & Ormerod, 2009). It is suggested that this “incubation” interval provides time for the forgetting of inappropriate solution concepts that mislead individuals during initial attempts (Smith, 1995; Smith & Blankenship, 1991), as well as a gradual spread of activation toward previously ignored but relevant memory items (Bowers, Regehr, Balthazard, & Parker, 1990; Smith, 1995; Smith & Blankenship, 1991; Yaniv & Meyer, 1987). To ensure that any observed postsleep performance improvement is sleep rather than time dependent, the majority of past studies on the effect of sleep for task performance have compared posttraining performance between a period of sleep and a comparable period of wakefulness (Stickgold, 2005). However, in some circumstances, a pause in problem solving may not be beneficial (Sio & Ormerod, 2009). Problem solving has been characterized as a search for a solution within a problem space (Newell & Simon, 1972). An uninterrupted effort allows individuals to perform a comprehensive search within the problem space. A discontinued effort (e.g., interruption filled by other activities) may impair problem-solving performance by not allowing participants to continue a focused search to completion (Gall & Mendelsohn, 1967; Olton & Johnson, 1976; Sio, 2010; Wiley, 1998). We cannot therefore conclude that there are advantageous effects of sleep for problem solving unless we also compare performance with a condition where participants continue with the problem without long interruption. In our study, we include a comparison between participants who continued immediately with the problem solution and participants who had a period of incubation or sleep between first and second exposures to the problems.

Method

Participants

Twenty-seven male and 34 female students from Lancaster University were paid to participate in this experiment. Mean age of the participants was 20.5 years ($SD = 2.3$), all had normal or corrected-to-normal vision, and all spoke British English as their first language.

Materials

Thirty problems were taken from the set of RAT problems tested by Bowden and Jung-Beeman (2003b).¹ All were determined to be appropriate for British participants by two first-language British-English-speaking judges. The distinction between easy and difficult problems was determined on the basis of performance of the group of participants in the study from the first testing session. This was because norms from previous studies of solution rate of RAT problems were not correlated with performance on the RAT problems by our group of participants (see the Results section for more details). The problems were thus divided according to a median split into those that the participants as a whole found easier to solve and those that were more difficult. Thus, problems that the participant group as a whole solved less often were assigned to the difficult group of items, and those problems that the participant group solved more often were assigned to the easy group of items.

To test the reliability of these solution rates, we conducted an additional experiment to measure performance on the same set of RATs. Twenty-five Lancaster University students (13 male, 12 female) with mean age 21.8 years ($SD = 5.8$) were recruited from the same population as the main study. The procedure was the same as that for the first session of the main experiment, described below, except that participants attempted all the RATs in a single session. The solution rate data in this additional study were highly correlated with performance in the main study, $r(27) = .888$, $p < .001$, indicating that the easy/difficult distinction was consistent across the two samples of our population.

We observed that measuring ease of RAT problems is best done on the population currently being assessed, rather than relying on solution rate norms for participants with different backgrounds tested under slightly different conditions. We found that there was no correlation between performance on the RAT problems for our British English participants in the first testing session and Bowden and Jung-Beeman's (2003b) RAT solution rates based on American English participants, $r(27) = -.083$, $p = .668$. Similarly, the additional 25 participants sampled from the same population that attempted to solve all the RATs also demonstrated a lack of correlation with the Bowden and Jung-Beeman (2003b) solution rates, $r(27) = -.180$, $p = .349$.

A further 12 RATs were taken from the Bowden and Jung-Beeman (2003b) set for presentation as novel items in the second test session, and again all were judged to be suitable for British-English-speaking participants.

We measured length of sleep between test sessions for the groups that slept during the study using an *Actisleep* activity

monitor (ActiGraph, Pensacola, FL), the data from which were analyzed using the Sadeh, Sharkey, and Carskadon (1994) algorithm to determine sleep length.

Design and procedure

Participants were tested two times on the RAT problems in one of three conditions that varied the time interval between the first and second tests and whether participants slept or stayed awake between tests and controlled for time of day of testing (see Table 1). For the control group, there was no delay between the first and second testing sessions, so the second test followed immediately after the first session was completed. The control group comprised two subgroups that varied the time of day of testing (testing was at 9 a.m. or at 9 p.m.), to ensure that across the whole group performance was not confounded with time of day. For the incubation group, participants were tested at 9 a.m. and retested 12 h later (9 p.m. the same day). For the sleep group, the first and second test sessions were separated by an intervening night of sleep. The sleep group comprised two subgroups which, as with the control group, controlled for effects of time of day of testing (one subgroup was tested at 9 p.m. and 9 a.m. the next day, the other subgroup was tested at 9 p.m. and 9 p.m. the following day).

In the first test session, participants were asked to solve a series of RAT problems displayed individually on a computer screen for 1 min. Participants were able to enter their answer at any point. If their response was correct, the next problem was presented; otherwise, they were informed that their answer was incorrect, and they continued to attempt to solve the problem. At the end of 1 min, participants were given a final opportunity to make a response; otherwise, they pressed a key indicating "no response," and the next problem was presented. The first testing session was completed once 8 problems were unsolved or once all 30 RAT problems had been

Table 1 Time of day of the testing sessions in each condition

| Group | Time | | | |
|---------------------------------------|--------------|--------------|----------|--------|
| | 9 a.m. | 9 p.m. | 9 a.m. | 9 p.m. |
| Control, morning test ($n = 11$) | 1st/2nd test | | | |
| Control, evening test ($n = 12$) | | 1st/2nd test | | |
| Incubation ($n = 22$) | 1st test | 2nd test | | |
| Sleep, morning test ($n = 8$) | | 1st test | 2nd test | |
| Sleep, evening test ($n = 8$) | | 1st test | 2nd test | |

¹ One of the RAT problems contained a typographic error, so it was subsequently excluded from the analysis.

presented. Since the problems were presented in random order to participants, it was possible to determine the relative ease or difficulty of the problems for the whole group that were unsolved by each individual. In the second test session, participants were given the unsolved problems from the first session, randomly intermixed with 12 novel RAT problems.

For the incubation and sleep groups, participants were instructed to keep to their normal sleep–wake routine and to abstain from taking naps, alcohol, and caffeine during the course of the study. Participants in the sleep group were issued with a sleep monitor during the intersession overnight sleep in order to measure the length of their sleep.

Results

In all analyses, performance was measured in terms of the proportion of correct responses. In order to assess time-of-day effects, as well as potential effects of interval between tests for the sleep group, performance between the control morning and evening test subgroups and that between the sleep morning and evening test subgroups were compared. There were no significant main effects for the two control groups on accuracy of RAT solutions on the first (morning test, $M = .419$, $SD = .200$; evening test, $M = .413$, $SD = .183$) or on the second (morning test, $M = .318$, $SD = .148$; evening test, $M = .278$, $SD = .120$) test session, both $F_s < 1$. Comparisons between the two sleep subgroups were also not significantly different for first (morning test, $M = .492$, $SD = .122$; evening test, $M = .509$, $SD = .092$) and second (morning test, $M = .302$, $SD = .117$; evening test, $M = .281$, $SD = .160$) test sessions, both $F_s < 1$. Although the subgroups comprised small numbers of participants (see Table 1), the accuracies were very similar across groups, suggesting that time of day was not a factor influencing performance. Consequently, the two control subgroups and the two sleep subgroups were merged into a single control group and a single sleep group, and further analyses compared the three main groups: control, incubation, and sleep. Although the sleep group had the opportunity for incubation as well as sleep, the sleep group enabled us to test the additional effect of sleep for this group, as compared with the incubation-only group. However, the contribution of sleep and longer incubation in the morning sleep group (24 h between sessions) did not influence performance distinctly, as compared with the evening sleep group (12 h between sessions) who had the opportunity for a shorter incubation period, suggesting that differences in incubation did not affect performance over and above the effect of sleep.

In order to test that participants in the three groups did not differ in their initial ability to solve RATs, accuracy of RAT problem performance for the first test was compared in a one-way ANOVA and resulted in no significant difference, $F(2, 58) = 1.521$, $p = .237$, $\eta_p^2 = .050$, indicating that the groups were initially balanced in terms of performance on solving the RAT problems on first exposure (see Fig. 1, left panel).

Similarly, we tested whether each group differed in terms of their overall performance on the new items in the second test. A one-way ANOVA revealed no significant differences, $F(2, 58) < 1$, indicating that the groups did not differ in terms of practice effects for the second test (see Fig. 1, right panel). There was, however, a significant difference between overall performance in the first test ($M = .445$, $SD = .154$) and the performance on the new items in the second test ($M = .286$, $SD = .143$), $p < .001$. The set of RATs presented as new items in the second test were more difficult than the set of 30 RATs that constituted the first test; however, the important feature of the new items is that the groups did not differ in their performance for these items. A detailed examination on task difficulty for the sets of RATs is presented in the next section.

Task difficulty

We divided the set of RAT problems presented at the first test into difficult and easy sets, on the basis of a median split of solution rate during the first test; then ease of solution in the first test session was used to assess performance in the second session. The RAT problems classified as difficult had a mean solution rate of 31 % ($SD = 19.4$). The problems classified as easy had a mean solution rate of 68 % ($SD = 10.2$). Since

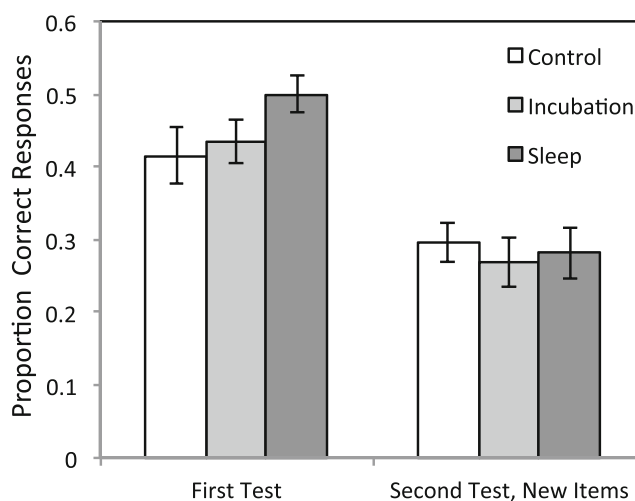


Fig. 1 Proportions of correct responses for the control, incubation, and sleep groups for solution accuracy for the first test and the new items in the second test. Error bars indicate $\pm 1 SE$

only previously unsolved RAT problems were presented to each participant in the second test session, there was a greater proportion of difficult (.663) than of easy (.347) RAT problems presented in the second test. A one-sample *t*-test comparison with an equal distribution of .5 was significant, $t(60) = 7.742$, $p < .001$, Cohen's $d = 1.999$. However, the proportion of difficult problems used in the second test session did not differ among the three groups, $F(2, 58) = 0.795$, $p = .457$, $\eta_p^2 = .027$.

In order to determine whether difficult and easy RAT problems differed in terms of associations between given words and the target word in the problems, we examined associative strength of word pairs from the Nelson, McEvoy, and Schreiber (1998) free-association norms. These associations were assessed in four ways in order to determine the factors underlying problem difficulty. First, strength of association between given words and the target word was assessed, but this did not differ significantly between easy ($M = .056$, $SD = .063$) and difficult ($M = .034$, $SD = .079$) RAT problems, $t(27) = 0.83$, $p = .412$, Cohen's $d = 0.320$. Second, the size of the initial search space was assessed in terms of the number of words given as associates to the three given words. No significant difference was found between the difficult ($M = 41.571$, $SD = 9.070$) and the easy ($M = 48.400$, $SD = 14.272$) RAT problems, $t(27) = 1.525$, $p = .139$, Cohen's $d = 0.586$.

Third, a recent study of Gupta, Jang, Mednick, and Huber (2012) revealed that some RAT problems are difficult because their target words are low-frequency words and individuals are usually biased to consider high-frequency incorrect words when solving these problems. Hence, word frequency of the target should be correlated with RAT solution rate. We assessed two types of word frequency: the written word frequency (Kučera & Francis, 1967) and the associate frequency (AF; Griffiths, Steyvers, & Firl, 2007; Gupta et al., 2012), which is the sum of the associative strengths of all words that are associated to the target word. No significant differences in word frequency were found between difficult and easy RAT problems [written word frequency, $t(27) = -0.053$, $p = .958$, Cohen's $d = 0.020$; AF, $t(27) = -0.069$, $p = .946$, Cohen's $d = 0.027$].

The fourth measure of associations in the problems that we conducted was the number of given word stimuli for which the target word was given as an associate in the database. This measure was significantly different between difficult ($M = 1.000$, $SD = 0.784$) and easy ($M = 1.933$, $SD = 0.799$) RAT problems, $t(27) = 3.171$, $p = .004$, Cohen's $d = 1.221$. Hence, if the solution search process is guided by spreading activation from associated task stimuli, it should be more effective if focused when the easy RAT problems are solved and more effective if activation is spread more broadly across a semantic network for the difficult RAT problems, as predicted by Mednick's (1962) descriptive model.

In order to determine the characteristics of the new RATs in the second test, we conducted similar analyses on the given word stimuli associated with the target word. The number of given word stimuli for which a target word was given as an associate differed between new, difficult, and easy RAT problems, $F(2, 38) = 10.521$, $p < .001$. The new RATs ($M = .667$, $SD = .651$) had fewer given word to target word associations than the easy RATs, $p < .001$, but did not differ from the difficult RATs, $p = .804$. This suggests that the new RATs in the second test were qualitatively similar to the difficult RATs and this accounts for the relatively low accuracy rates for new RATs on the second test, as compared with the overall accuracy of RAT solutions on the first test (including both easy and difficult items).

Effect of sleep

The facilitatory effects of incubation or sleep are demonstrated by significant differences as compared with the control group. Comparisons between easy/difficult items and new items are not able to reveal facilitatory effects, since the retested problems were those that participants generally found difficult to resolve. Previous studies revealing effects of sleep or incubation in comparison with new items have typically repeated presentation of all problems, regardless of whether they were solved or unsolved in the first testing.

Table 2 presents a breakdown of the number of RATs presented and solved, and the proportion of correct responses, in the first and the second tests, by group and problem type. The proportion of correct responses on both difficult and easy RAT problems in the second test was positively skewed—for easy RATs, skew = .729 ($SES = .306$); for difficult RATs, skew = .603 ($SES = .306$)—due to participants finding the previously unsolved problems somewhat difficult. A logarithmic transformation corrected the skew: .447 ($SE = .306$) for easy and .397 ($SE = .306$) for difficult RATs, both $ps > .1$.

We conducted an ANOVA on the transformed data with group (control, incubation, or sleep) as a between-subjects factor and problem type (difficult or easy) as a within-subjects factor. There was no significant main effect of group, $F(2, 58) < 1$, indicating no overall effects of incubation or sleep, as compared with the control condition. However, the effect of problem type was significant, $F(1, 58) = 5.229$, $p = .026$, $\eta_p^2 = .083$. This was due to lower solution rates for difficult problems ($M = .189$, $SD = .177$), as compared with easy RAT problems ($M = .323$, $SD = .341$), in the second test (for clarity, we report the untransformed proportion correct values throughout the results). Critically, there was a significant interaction between group and problem type, $F(2, 58) = 3.693$, $p = .031$, $\eta_p^2 = .11$ (see Fig. 2). This interaction was due to the significantly better performance on difficult items by the sleep group, as compared

Table 2 Number of RATs presented and solved in each session by problem type and group

| Condition | Easy RAT | | | Difficult RAT | | |
|-------------|------------------------------|---------------------------|--|------------------------------|---------------------------|--|
| | Presented Mean (<i>SD</i>) | Solved Mean (<i>SD</i>) | Proportion of correct responses Mean (<i>SD</i>) | Presented Mean (<i>SD</i>) | Solved Mean (<i>SD</i>) | Proportion of correct responses Mean (<i>SD</i>) |
| First test | | | | | | |
| Control | 7.217 (3.232) | 4.391 (3.187) | .557 (.279) | 7.391 (2.692) | 2.438 (2.352) | .292 (.220) |
| Incubation | 7.091 (2.20) | 4.454 (2.283) | .600 (.211) | 7.455 (2.017) | 2.273 (1.518) | .289 (.175) |
| Sleep | 8.067 (2.658) | 5.800 (2.484) | .692 (.167) | 7.867 (1.727) | 2.600 (1.454) | .312 (.119) |
| Second test | | | | | | |
| Control | 2.826 (1.557) | 0.826 (0.887) | .289 (.355) | 4.957 (1.581) | 0.870 (1.058) | .156 (.181) |
| Incubation | 2.636 (1.177) | 1.181 (1.140) | .405 (.347) | 5.182 (1.296) | 0.773 (0.752) | .135 (.108) |
| Sleep | 2.250 (0.856) | 0.563 (0.629) | .260 (.310) | 5.313 (.793) | 1.563 (0.892) | .293 (.169) |

with the control and incubation groups, $F(2, 58) = 4.260$, $p = .019$, $\eta_p^2 = .128$ [sleep ($M = .293$, $SD = .168$) vs. control ($M = .156$, $SD = .181$), $p = .040$; sleep vs. incubation ($M = .147$, $SD = .155$), $p = .032$]. Performance on easy items was not significantly different across the three groups, all $ps > .5$. The interaction demonstrates that the facilitatory effect of sleep was observed only for the difficult RAT problems.

For the combined sleep group, participants slept, on average, 7.47 h ($SD = 1.54$). The two sleep subgroups had similar sleep onset time, $p = .92$. The sleep time for the evening sleep group ($M = 8.56$, $SD = 6.37$) was longer than that for the morning sleep group ($M = 6.375$, $SD = 1.188$), $t(14) = 4.019$, $p < .001$, Cohen's $d = 2.148$. However, there was no correlation between the total sleep time and the degree of sleep-dependent performance improvement [difficult RAT problems, $r(14) = .048$, $p = .861$; easy RAT problems, $r(14) = .173$, $p = .522$].

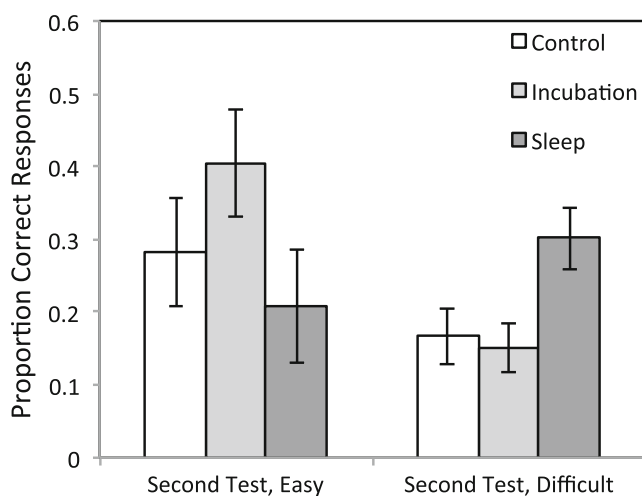


Fig. 2 Proportions of correct responses for the control, incubation, and sleep groups for the previously unsolved easy and difficult items in the second test. Error bars indicate $\pm 1 SE$

Discussion

This study demonstrated a sleep-dependent improvement in problem solving, in comparison with a group that had a similar period of incubation between first and second exposures to a problem, and also in comparison with a group that continued with the problems without a long interruption. The sleep-dependent improvement was not general across all the problems but was evident only for difficult problems. When attempting difficult RAT problems in the second test, the sleep group demonstrated a significant improvement over the control group and the incubation group. When easy RAT problems were solved, the degree of improvement in the second test across the three groups was less distinct. This supported our prediction that sleep would provide a particular benefit for difficult problems.

Previous studies of the effect of sleep on problem solving have proposed spreading activation among a network of associated concepts as the process boosted by sleep (Cai et al., 2009; Stickgold & Walker, 2004). To solve a RAT requires activation passing from the stimulus words to the target word with sufficient strength to activate the target word to a sufficient level that it can be retrieved as a potential solution (Mednick, 1962). Cai et al. found that implicit presentation of the target word in an unrelated task could be assimilated with the problem as a consequence of REM sleep, thereby resulting in improvements in solution rate for these primed answers, but we instead found such facilitatory effects of sleep without additional cues to the answer, but only for the more difficult problems. Payne (2011) proposed that the beneficial effect of sleep is due, in some instances, to reactivation of information during sleep. Reactivation of the stimulus words of a problem during sleep would result in an increase in activation throughout the network of associated words connected to the stimulus words and would provide an advantage for RAT problem solving particularly for problems requiring a

larger spread of activation to discover the answer—that is, those problems that participants initially found difficult to solve.

In our study, we have provided a greater specification of the properties of the associative network that result in such sleep-dependent improvements in solution rate. The distinction between easy and difficult problems corresponds to a distinction in terms of the number of stimulus words that have direct associations to the target word. Thus, the difficult problems, with fewer direct associative links from stimuli to target, required a broader spread of activation in order to increase activation of the remotely associated target word to enable the RAT's solution. This could be achieved by spreading activation directly to weak associates, a process supported by sleep (Stickgold et al., 1999), or via mediating associative concepts. In contrast, the easy RAT problems had a greater number of stimulus words that were directly associated with the target, and so a focused search of closely associated words would be sufficient for discovering the answers to the easy RAT problems. For these easy problems, immediate continuation with the problem appeared to be sufficient to support the local spreading activation for solution, and incubation and sleep did not sufficiently improve performance, as compared with the control condition.

As with all incubation studies with a long interval between testing, it is not possible to ensure that participants do not consciously return to the problem between test sessions. However, the effect of sleep appears to be additional to any effect of possible conscious attempts at the problem. As was previously mentioned, the two sleep subgroups did not differ in terms of accuracy of RAT solutions on the second test session, despite their differing opportunities to return to the problem during wakefulness. In addition, any effects of conscious return to the problems would have been demonstrated by distinct performance between the incubation and control groups. Since these differences were small as compared with the effects of sleep, we can be confident that the observed effects were due to processing associated with sleep.

An alternative explanation to the spreading activation account for the differential impact of sleep for the difficult problems is that sleep allows the forgetting of competing ideas, which block access to the correct solution and may lead to impasse (Ash & Wiley, 2006; MacGregor, Ormerod, & Chronicle, 2001). When difficult RATs are solved, it is possible that a greater number of incorrect words may be retrieved during the initial attempt and, consequently, a forgetting mechanism is required in order to resolve the retrieval competition. Forgetting of misleading concepts is considered to be one of the mechanisms that facilitate solution discovery (Storm & Angello, 2010).

However, as was reported above, the overall number of associates to the task stimuli were not significantly different for difficult than for easy RAT problems; thus, the total

number of incorrect, potentially misleading answers that were directly related to the stimulus words was not sufficient to account for problem difficulty. Equally, the nonsignificant difference in terms of the AF of the target words between difficult RATs and easy RATs suggests that these two sets of RATs induced similar levels of retrieval bias. Therefore, the same level of retrieval competition would be expected when easy and difficult RATs are solved. If forgetting is the core mechanism underlying the sleep effect in this study, sleep-dependent improvement should be comparable between the difficult and easy RATs.

In our study, we tested only accuracy rates for RATs, but future studies could collect more qualitative measures of insight problem solving to determine with greater precision yet the nature of facilitatory sleep effects for problem solving. For instance, previous studies of RATs have gathered data on participants' feelings of insight (Bowden & Jung-Beeman, 2003a), feelings of restructuring, and feelings of suddenness (Sandkühler & Bhattacharya, 2008). Such additional measures could indicate how sleep affects problem solution, extending our present work that provides the first step in demonstrating the effect sleep has on solution rate with regard to problem characteristics.

The differential impact of sleep on difficult and easy RAT problems implies that the effect of sleep is not general, even among tasks sharing the same type of representation. This suggests that tasks sharing similar superficial features do not necessarily require the same processes for their effective completion (Kuriyama et al., 2004) and, consequently, may not benefit from sleep-dependent processing to the same degree. Future studies whose aim is to identify which aspects of problem solving benefit from sleep should focus not only on the domain of the task, but also on the characteristics of the task. On the basis of the findings of this study, we suggest that easy and difficult problems require different degrees of spreading activation. In conclusion, we suggest that the idiom “sleep on it” should be elaborated to “sleep on it, but only if it is difficult.”

Acknowledgements This research was supported by ESRC Grant RES-000-22-4049.

References

- Ash, I. K., & Wiley, J. (2006). The nature of re-structuring in insight: An individual differences approach. *Psychonomic Bulletin and Review*, *13*, 66–73.
- Ball, L. J., & Stevens, A. (2009). Evidence for a verbally-based analytic component to insight problem solving. Chapter in N. Taatgen and H. van Rijn, J. (Eds.): *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society.
- Bowden, E. M., & Jung-Beeman, M. (2003a). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin and Review*, *10*, 730–737.

- Bowden, E. M., & Jung-Beeman, M. (2003b). One hundred forty-four Compound Remote Associate Problems: Short insight-like problems with one-word solutions. *Behavior Research Methods, Instruments, & Computers*, *35*, 634–639.
- Bowers, K. S., Regehr, G., Balthazard, C. G., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*, 72–110.
- Cai, D., Mednick, S. A., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences*, *106*, 10130–10134.
- Cartwright, R. (1974). Problem solving: Waking and dreaming. *Journal of Abnormal Psychology*, *83*, 451–455.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.
- Gall, M., & Mendelsohn, G. A. (1967). Effects of facilitating techniques and subject-experimenter interaction on creative problem solving. *Journal of Personality and Social Psychology*, *5*, 211–216.
- Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science*, *18* (12), 1069–1076.
- Gupta, N. G., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high-frequency responses. *Psychological Science*, *23*(3), 288–294.
- Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 431–440.
- Kenett, Y. N., Kenett, D. Y., Ben-Jacob, E., & Faust, M. (2011). Global and local features of semantic networks: Evidence from the Hebrew mental lexicon. *PLoS One*, *6*(8), e23912.
- Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor skill complexity. *Learning and Memory*, *11*, 705–713.
- MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information-processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 176–201.
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 220–232.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms*. Available: <http://www.usf.edu/FreeAssociation/>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Olton, R. M., & Johnson, D. M. (1976). Mechanism of incubation in creative problem solving. *The American Journal of Psychology*, *89*, 617–630.
- Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London: Harvester Wheatsheaf.
- Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York: Cambridge University Press.
- Payne, J. D. (2011). Sleep on it: Stabilizing and transforming memories during sleep. *Nature Neuroscience*, *14*(3), 272–274.
- Payne, J. D., Schacter, D. L., Propper, R., Huang, L.-W., Wamsley, E., Tucker, M. A., ... Walker, M. P. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory*, *92*, 327–334.
- Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep*, *17*, 201–207.
- Sandkühler, S., & Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLoS One*, *3*(1), e1459.
- Sio, U. N. (2010). *The underlying mechanism of incubation period in problem solving* (Unpublished doctoral dissertation). Lancaster University, Lancaster, UK.
- Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, *135* (1), 94–120.
- Smith, S. M. (1995). Fixation, incubation, and insight in memory and creative thinking. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 135–146). Cambridge, Mass: MIT Press.
- Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *The American Journal of Psychology*, *104*, 61–87.
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*, 1272–1278.
- Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-induced changes in associative memory. *Journal of Cognitive Neuroscience*, *11*, 182–193.
- Stickgold, R., & Walker, M. P. (2004). To sleep, perchance to gain creative insight? *Trends in Cognitive Sciences*, *8*, 191–192.
- Storm, B. C., & Angello, G. (2010). Overcoming fixation: Creative problem solving and retrieval-induced forgetting. *Psychological Science*, *21*, 1263–1265.
- Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*, 352–355.
- Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory and Cognition*, *26*, 716–730.
- Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 187–205.