The influence of sleep on emotional and cognitive processing is primarily trait- (but not state-) dependent ☆

Itamar Lerner [a,*], Shira M. Lupkin [a], James E. Corter [b], Suzanne E. Peters [a], Lee Anne Cannella [a,1], Mark A. Gluck [a]

[a] Center for Molecular and Behavioral Neuroscience, Rutgers University, 197 University Avenue, Newark, NJ 07102, USA
[b] Department of Human Development, Teacher's College, Columbia University, 525 W. 120th Street, New York, NY 10027, USA

ARTICLE INFO

ABSTRACT

Human studies of sleep and cognition have established that different sleep stages contribute to distinct aspects of cognitive and emotional processing. However, since the majority of these findings are based on single-night studies, it is difficult to determine whether such effects arise due to individual, between-subject differences in sleep patterns, or from within-subject variations in sleep over time. In the current study, we investigated the longitudinal relationship between sleep patterns and cognitive performance by monitoring both in parallel, daily, for a week. Using two cognitive tasks – one assessing emotional reactivity to facial expressions and the other evaluating learning abilities in a probabilistic categorization task – we found that between-subject differences in the average time spent in particular sleep stages predicted performance in these tasks far more than within-subject daily variations. Specifically, the typical time individuals spent in Rapid-Eye Movement (REM) sleep and Slow-Wave Sleep (SWS) was correlated to their characteristic measures of emotional reactivity, whereas the typical time spent in SWS and non-REM stages 1 and 2 was correlated to their success in category learning. These effects were maintained even when sleep properties were based on baseline measures taken prior to the experimental week. In contrast, within-subject daily variations in sleep patterns only contributed to overnight difference in one particular measure of emotional reactivity. Thus, we conclude that the effects of natural sleep on emotional cognition and category learning are more trait-dependent than state-dependent, and suggest ways to reconcile these results with previous findings in the literature.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Over the last two decades, sleep has been repeatedly shown to play a central role in memory consolidation and emotional cognition. Using polysomnography (PSG) measures, human studies demonstrated that specific sleep stages tend to affect specific cognitive abilities. Generally, Rapid-Eye-Movement sleep (REM) has been linked to procedural memory, high-level linguistic processes and the processing of emotional stimuli, whereas non-REM sleep (NREM) – and Slow-Wave Sleep (SWS) in particular – have been implicated in processes such as declarative memory, context sensitivity and relational learning (e.g., Groch, Wilhelm, Diekelmann, & Born, 2013; Gujar, McDonald, Nishida, & Walker, 2010; Plihal & Born, 1997; for reviews, see Rasch & Born, 2013; Walker, 2009).

In human studies, the standard methodology for examining the effects of sleep on cognitive function involves participants learning a cognitive task, spending a night (or, in case of nap studies, an afternoon) in a sleep laboratory during which their sleep is monitored with PSG, and in some studies manipulated, and then being retested. Results from these individuals are then compared to a control group in which no sleep period is interposed between the two sessions, or no sleep manipulation is exercised. Any performance benefits in the experimental group over the control group are attributed to sleep and compared to specific sleep parameters that were measured during the night (or the afternoon nap).

One key limitation of this single-night methodology is that it is difficult to determine whether correlations between sleep and cognitive performance are due to specific "**state-dependent**" properties of sleep that each participant happened to experience in the studied night, or, conversely, result from "**trait-dependent**" individual differences in participants' general sleep patterns. Often, single-night studies implicitly assume the former, failing to take into account individuals' baseline patterns. However, given that there is evidence suggesting that both sleep architecture (e.g., De Gennaro et al., 2008; Linkowski, 1999) and various cognitive and affective processes (e.g., Neta, Norris, & Whalen, 2009; Volk, McDermott, Roediger, & Todd, 2006) are in fact stable traits whose variability is lower within-subjects than between-subjects, it is possible that the interaction between sleep and these processes also follows a trait-like pattern.

Traditional human sleep studies suffer from several other limitations. First, due to their reliance on data from a single night, they are insufficient to address the effects of sleep on the learning of complex tasks that require multiple days to master (e.g., Shohamy, Myers, Onlaor, & Gluck, 2004). Second, the use of PSG is known to precipitate several sleep disturbances that contribute to poor sleep quality (e.g., increased awakenings and a decreased percentage of REM sleep; Agnew, Webb, & Williams, 1966). These effects have been shown to persist for up to three nights (the so called "First –night effects"), even when the PSG system is employed in participants' homes (Le Bon et al., 2001). Consequently, the ecological validity of such studies may be jeopardized, as exemplified by studies showing effects of sleep on cognition that appear only when sleep occurs without the use of PSG, but not when repeated in a laboratory (e.g., Djonlagic et al., 2009). While many of these limitations may be addressed by observing participants for extended periods of time, the nature of PSG studies typically renders long-term investigations both cost-prohibitive and logistically unfeasible. As a result, few controlled longitudinal studies that measure sleep and cognition in parallel have been performed to date (cf., Burke, Scheer, Ronda, Czeisler, & Wright, 2015).

In the present study we sought to address the limitations of traditional single-night studies by examining the longitudinal effects of sleep on behavioral performance. To that end, we utilized a combination of easy-to-use mobile devices that allowed participants to both monitor their sleep and administer cognitive tasks for multiple days, by themselves and in their own homes. We examined the effect of sleep on two behavioral tasks; one that tested emotional cognition, specifically reactivity to emotional facial expressions, and a second examining memory consolidation during category learning. Thus, our study tapped into both cognitive and affective processing, two central themes in the human sleep-cognition literature. The specific tasks were chosen for several reasons. First, both were compatible with (or could be adapted to) a long-term study that requires repeated administration over multiple days. Second, it was previously shown in single-night studies that performance in these and similar tasks is influenced by sleep (Barsky, Tucker, & Stickgold, 2015; Djonlagic et al., 2009; Gujar et al., 2010; Lara-Carrasco, Nielsen, Solomonova, Levrier, & Popova, 2009; Van Der Helm, Gujar, & Walker, 2010). Lastly, results regarding the precise role of sleep—and specific sleep stages—on performance in these previous studies have been inconsistent at best, raising the possibility that investigating these relations over a single night is insufficient.

Overall, we aimed to answer two fundamental questions: First, what are the relative contributions of daily and baseline sleep patterns on cognitive performance. If the state-like hypothesis is correct, we expected to see daily fluctuations in performance in accordance with properties of sleep during the preceding night. If, however, the nature of the relationship were more trait-like, we would expect an effect when comparing average performance

and sleep levels between-subjects. Second, we sought to determine whether new relationships between sleep stages and performance emerge when taking under consideration multiple nights of sleep, and whether these can shed light on inconsistencies in previous studies.

2. General methods

2.1. Participants

Twenty-three healthy students (n = 11 females) from Rutgers University and the New Jersey Institute of Technology participated in this study for monetary compensation (Table 1). Exclusion criteria included personal or family history of sleep, neurological or psychiatric disorders, drug or alcohol abuse, and/or intake of medications that have any effect on sleep. Three participants were discarded from the study due to a lack of reliable use of equipment, resulting in three or more experimental days of unusable sleep and/or behavioral data (see Section 1.2.1 in the *Supplemental Materials*). Throughout the experiment participants were asked to not increase their daily caffeine intake, to maintain their regular sleep schedule, and to refrain from alcohol consumption and daytime napping. All participants provided informed consent in line with the procedures approved by the Institutional Review Board of Rutgers University.

2.2. Sleep monitoring and cognitive testing devices

2.2.1. Mobile sleep monitoring system

The mobile sleep monitoring system included an automated wireless sleep-monitoring headband (Zeo Inc., Newton, MA), an actigraphy bracelet (Micro-MotionLogger Sleep watch, Ambulatory Monitoring, Inc., Ardsley, NY), and an Android tablet (Amazon.com, Inc., Seattle, WA).

The sleep-monitoring headband is equipped with a single bi-polar fabric sensor that transmits data wirelessly to the Android tablet, which acts as a base station. The sensor is fitted with three silver-coated electrodes used to detect brain waves (EEG), eye movements (EOG), and the movement of the frontalis muscle (EMG). The signals from these electrodes are analyzed in real time to produce sleep staging in 30-s epochs. This sleep staging, the accuracy of which was validated for nocturnal sleep compared to PSG in multiple studies (e.g., Griessenberger, Heib, Kunz, Hoedlmoser, & Schabus, 2013; Shambroom, Fabregas, & Johnstone, 2012), is a reduced version of the official staging criteria by the American Association of Sleep Medicine (Iber, Ancoli-Israel, Chesson, & Quan, 2007) and differentiates between four stages rather than five – wake, N1/N2 (combined N1 and N2 stages, termed 'Light sleep'), SWS ('Deep sleep'), and REM sleep.

The actigraphy bracelet is a research-grade device that contains a built-in accelerometer used to infer sleep/wake decisions in one-minute epochs based on participants' arm movements (Ancoli-Israel et al., 2003; de Souza et al., 2003). Participants wore the actigraph on the non-dominant wrist throughout the entire study. Data was extracted from the devices at the end of the experiment, and was used to assess the sleep/wake validity of the sleep-monitoring headband (see detailed description in Section 1.2 of the *Supplemental Materials*).

2.2.2. Mobile cognitive assessment

The cognitive tasks were delivered using a separate application on the same Android tablet used to collect and transmit data from the sleep-monitoring headband. In each experimental session, participants completed an emotional reactivity task followed by a probabilistic category-learning task, described in detail below.

2.3. General procedure

Prior to commencing the cognitive experiments, participants monitored their sleep at home for a period of 4–7 days ('habituation phase;' see Table S1). In addition to receiving training on how to use the sleep monitoring devices, they were instructed to keep a sleep log noting their sleep/wake times and any nocturnal awakenings, as well as to send both the data from the monitoring devices and a picture of their log to a secure email that was monitored daily by experimenters. This period allowed participants to adapt to the use and operation of the devices and allowed experimenters to collect participants' baseline sleep measurements and assess participants' reliability in following the sleep-monitoring protocols (i.e., consistently emailing data, ensuring that no data was missing because of the sleep-monitoring headband falling off at night, etc.).

Following the habituation phase, participants began cognitive testing while continuing to monitor their sleep, for seven additional days ('experimental phase'). Testing occurred twice daily at home, once in the morning (between 8 am and 10 am, at least half an hour after waking to prevent the effects of sleep inertia) and once in the evening (between 8 pm and 10 pm, at least half an hour before going to sleep to reduce the effects of tiredness), for a total of 14 sessions. Each session included both cognitive tasks. The first session was conducted in the lab and included a short training period prior to testing. Training stimuli were different than the ones used in the actual experiment. Fig. 1 illustrates the general experimental design.

2.4. Data analysis

Data from the sleep-monitoring headband and the actigraphy bracelet were integrated for each subject to yield measures of total time spent in each sleep stage (see details in *Supplementary Material* Section 1.3). These measures were then used in a mixed-model Analysis of Variance (ANOVA) to predict the daily behavioral measures. Comparisons of within- and between-subjects effects were carried out in SAS 9.3 (SAS Institute) using the *Mixed* procedure with a covariance structure for error defined by Kronecker products, specifying unstructured covariances for the Time-of-day (mornings vs. evening) factor and a first-order autoregressive AR (1) structure for Day. Other analyses, including multiple regression, Principle Components Analysis (PCA), and curve-fitting were carried out in Matlab R2015a (MathWorks). Specific details for each analysis are described in the corresponding sections in the main text.

3. Experiment 1: Emotional Face Valence Rating Task

3.1. Task description

The Emotional Face Valence Rating Task assesses participants' emotional reactivity to images portraying different emotional expressions (cf. Gujar et al., 2010). Stimuli were comprised of 392 photographed faces obtained from the Radboud Faces Database (RaFD; Langner et al., 2010). The images used were of

57 individuals, each portraying seven different emotions (sad, angry, disgusted, fearful, neutral, surprised and happy). The images were converted to grayscale, equalized on luminance, and counterbalanced for average and standard deviation contrast using the image processing toolbox in Matlab (Willenbockel et al., 2010). In each of the 14 testing sessions, participants were presented with four different individuals, each portraying one of the seven emotions, followed by another four portraying a different emotion, and so on, until each participant had seen four instances for each of the seven emotions (a total of 28 images). Images appeared on screen for two seconds, after which participants were asked to rate the valence of the image on a scale of 1–5, where 1 is negative and 5 is positive (Fig. 2A). No time constraint was imposed on decision-making. After a response was made, the scale remained on the screen for 1 s, after which a new trial began. Different images were used in each session such that no image was repeated over the course of the experiment. The order of the images was randomized across all sessions for each participant.

3.2. Results

Raw scores for each picture were standardized using the mean ratings and the standard deviation obtained from the RaFD database such that participants' answers reflected standardized deviations from the mean "population" responses to that particular picture. In order to decrease the number of multiple comparisons and focus on the main contributors to variance in performance, the standardized scores were subject to dimensionality reduction using PCA. For each participant, the scores given to each emotion were averaged across the 14 sessions, resulting in a 7-dimensional vector for each participant. We then obtained the 7 eigenvectors and eigenvalues comprising the PCA components. The results showed two main components ('factors') responsible for the variation in the normalized ratings (Fig. 2B). Inspecting the component loadings of the original facial expressions composing these two dimensions (that is, the degree to which the original dimensions contribute to the new ones; Fig. 2C) we found that the loadings on the first component were dominated by the *Surprised* expression and, to a lesser extent, by *Disgust* and *Fearful* expressions (all with positive loadings). Because *Surprised*, *Disgust* and *Fearful* are often perceived as relatively complex and ambiguous facial expressions (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003; Neta et al., 2009; Pantic & Rothkrantz, 2000; Widen, 2013), we denote the first component as "Ambiguity", reflecting the tendency to evaluate ambiguity/complexity in faces as having a more positive or negative valence (see Section 3.3, Discussion, regarding other interpretations). The loadings of the second component had positive values for the *Angry*, *Disgust*, *Fearful* and *Sad* expressions, and negative values the *Happy* and *Neutral* expressions, all with similar absolute values. This means that high valence ratings given to expressions associated with negative emotions (angry, etc.) increased the value on this dimension, whereas low valence values decreased it, and the opposite was true for valence values given to expressions associated with emotions that are not negative (happy, neutral). Therefore, this component can approximately be interpreted as a tendency for 'Emotional

Fig. 1. Experimental design of the study. A habituation phase, in which participants monitored their sleep each night, was followed by a 7-day experimental phase in which participants continued to monitor their sleep and also had bi-daily assessments of cognitive and affective performance, in the mornings and evenings (for a total of 14 sessions, marked by s1, s2… in the figure). See text for details.
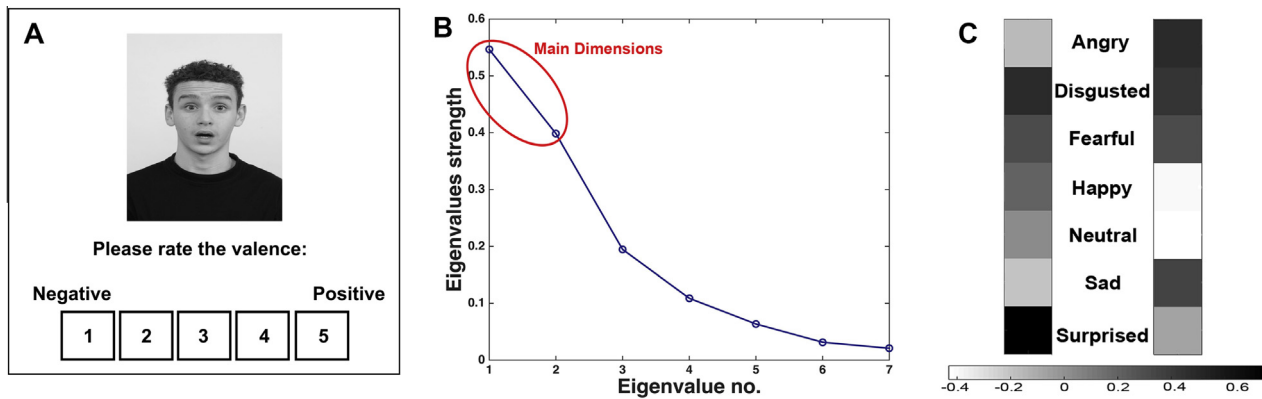
Fig. 2. A: Example of a trial in the Emotional Face Valence Rating Task; Results of PCA analyses: B: Eigenvalues, sorted from highest to lowest; C: Component loadings of the original dimensions on the first two Eigenvectors (corresponding to the two largest eigenvalues).

Blunting' (please see Supplementary Material for further discussion). To summarize, the PCA indicated that participants' evaluations of the stimuli differed primarily according to how ambiguous/complex they viewed a facial expression, and the degree to which they tend to blunt the emotion in that expression. The remaining statistical analyses examined the relations between sleep and each of these two components.

Previous studies have shown the evaluation of emotional faces to correlate with the total amount of REM sleep (or the percentage of REM sleep out of total sleep time) during the preceding night (Wagner, Fischer, & Born, 2002; Walker, 2009). We aimed at exploring whether such relations are driven by daily fluctuations in sleep and behavior within-subjects or by individual differences between-subjects, and whether other sleep stages may contribute to the evaluation when multiple nights of sleep are taken under consideration. To investigate these questions, we ran mixed-model ANOVAs examining the correlations between the daily amounts of time spent in REM, SWS and N1/N2 sleep with the Ambiguity and the Emotional Blunting factors underlying the face ratings (separately for each component). The dependent variables for the two models were the average scores for each session, for each participant, for each of the two PCA-derived factors, Ambiguity and Emotional Blunting, across all 14 sessions (7 days × 2 sessions/day). In order to assess if each rating factor was affected by stable, individual differences in sleep parameters, the ANOVA model included the average time spent in REM, SWS and N1/N2 (across all experimental nights) as "between-subjects" factors. To assess whether "within-subject" fluctuations in sleep patterns from one night to another affected the dependent variables, the daily deviations of REM, SWS and N1/N2 of each participant from his/her corresponding averages were entered as continuous (centered) covariates crossed with blocks (i.e., participants). The sleep variables of each night (termed REMcovdev, SWScovdev and N1/N2covdev in Table S2) were compared with the performance variables of the following day. Additional within-subject factors included Day (1–7) and Time of day (morning vs. evening). Finally, the Day × Time interaction was also included.

The statistical analysis showed (see Table S2 in the Supplementary Materials for the full results) that values of the Ambiguity component were significantly correlated with the REM factor ($F_{(1,16)} = 18.84$, $p = 0.0005$) – the average time across all seven days that each participant spent in REM sleep. In contrast, values of the second component, Emotional Blunting, were significantly correlated with the average time spent in SWS ($F_{(1,16)} = 18.20$, $p = 0.0006$). None of the day-to-day fluctuations in sleep measures contributed significantly to the prediction of either component, nor did Time of day (morning versus evening) appear to affect these

aspects of the ratings. Emotional Blunting scores, however, did vary by Day ($F_{(6,105)} = 2.60$; see Discussion for details). Fig. 3 illustrates the relations between REM and Ambiguity, and SWS and Emotional Blunting. Whereas no consistent effect exists within-subjects (comparing dots of the same color), across participants these relations are clearly evident, with each score increasing with its respective sleep stage modulator.[2]

Since raw amount of time in a sleep stage is necessarily confounded by total sleep time, it is common to corroborate such effects using the proportion of time spent in a sleep stage. Examining the effects of total sleep time on each component (separating, as before, within and between subject effects), we found a significant between-subject effect of total sleep time on Ambiguity ($F_{(1,18)} = 5.62$, $p = 0.0291$) though not for emotional Blunting. To ensure that our stage-specific results did not stem from variations in total sleep time, we ran separate mixed-model ANOVAs examining the effect of REM on Ambiguity scores and SWS on Emotional Blunting scores using relative measures of sleep, namely, the proportion of time spent in REM and SWS out of total sleep time, with similar between- and within-subject factors. These analyses yielded similar results to the ones using raw time in each sleep stage: A significant effect of mean proportional REM sleep on Ambiguity scores ($F_{(1,18)} = 16.31$, $p = 0.0008$) and of mean proportional SWS on Emotional Blunting scores ($F_{(1,18)} = 19.82$, $p = 0.0003$; Table S3, Supplementary).

If, indeed, the relationship between the amount of time spent in specific sleep stages and emotional reactivity to facial expressions is a stable trait, our results should remain unchanged even when these measures are decoupled in time. To verify this, we next examined whether the same correlations were replicable using data from the baseline measures of sleep (i.e., from the habituation phase) in place of data from the experimental period. For each participant, the average amount of time spent in each sleep stage, over the whole habituation phase, was used as a between-subject factor in the model (including also Time, Day and Time * Day interaction as within-subject factors). This analysis yielded, again, a significant effect between REM and Ambiguity, $F_{(1,16)} = 5.26$, $p = 0.0357$, and between SWS and Emotional Blunting, $F_{(1,16)} = 15.13$, $p = 0.0013$.

---

[2] The mixed-model ANOVAs shown in Table S2 contain only main effects of sleep stages. We also ran more complex mixed models analyses for each component that included interactions of Day or Time-of-day with SWS, REM, and N1/N2 (both the between- and within-subjects components). None of these interactions were significant and, consistent with the simpler model, the estimates in these models reconfirmed that there is a main effect of REM sleep on Ambiguity scores and that Emotional Blunting scores are affected by both mean SWS and Day (results not shown). All of the alternative models that included the additional interactions were found to be inferior to the additive model based on goodness of fit measures.
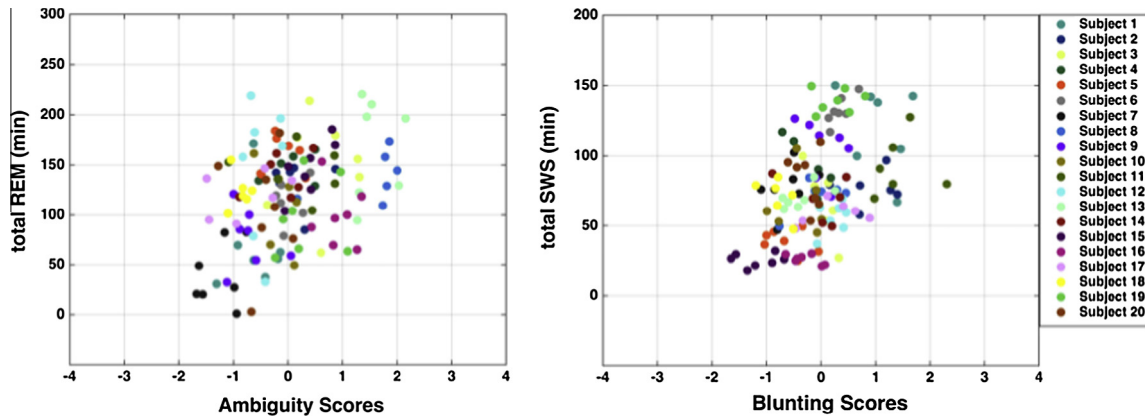
Fig. 3. Amount of nightly REM/SWS plotted against the corresponding scores of the Ambiguity (left) and Emotional Blunting (right) components the following day (averaged over the morning and evening sessions), respectively. Each dot represents one measure for one participant. Measures for the same participant are plotted in the same color. Some dots may not appear due to overlap, or due to missing data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this last analysis the effect of Day also remained significant. Like before, these analyses were re-run using the proportional measures of sleep to ensure that the noted effects did not stem from variations in total sleep time. This analysis, too, yielded significant effects ($F(1,18) = 7.93$, $p = 0.0115$); $F(1,18) = 10.14$, $p = 0.0051$, for average REM/Ambiguity and SWS/Emotional Blunting, respectively).

Next, to make a more direct comparison between the results of our study and those of previous studies examining the effect of a single night of sleep on the evaluation of facial expressions, the data was re-analyzed, this time computing correlations between Ambiguity and REM, and Emotional Blunting and SWS, separately for each of the seven days of the experiment. Only the morning scores were used for this analysis, comparable to the typical procedures in single-night studies (e.g., Gujar et al., 2010). Results are presented in Fig. 4, alongside a correlation computed over the average sleep and performance values across morning sessions. In three out of seven days there were significant correlations between the daily amount of REM sleep and the valence scores for Ambiguity at the following morning, with two additional days showing trend level effects. With regards to SWS and the scores for Emotional Blunting, only two days had significant correlations and one additional day showed results at a marginal level. With scores averaged over sessions, the correlations were evident and stronger for both factors, corresponding to our results using the mixed-model analysis.[3]

Finally, to conclude this analysis, rather than examining whether sleep modulated the raw scores of Ambiguity and Emotional blunting, we examined whether sleep affected the difference between their morning scores and the preceding evening's scores. This before-after comparison is often used in single nights studies to investigate overnight changes in emotional reactivity. For each participant, we calculated the difference scores of each component (since there was no evening score taken before the first night, only 6 nights were used in this analysis; see Fig. 1). These scores were then subject to a mixed-model ANOVAs with all within and between subject sleep factors, as well as Day, as predictors. Results are presented in Table S4. Whereas Ambiguity difference scores were not affected by any sleep measure, difference scores of

Emotional Blunting were significantly correlated with the within-subject SWS factor, SWScovdev ($F(1,84) = 6.67$, $p = 0.0116$). Parameter estimates suggested that this effect was opposite to the one found for between-subject SWS effects: the more SWS a participant had during a night, the lower the tendency for emotional blunting was at the following morning compared to the previous evening. This within-subject effect remained significant also when re-analyzing using the proportion of SWS out of total sleep time ($F(1,86) = 5.25$, $p = 0.0244$).

3.3. Discussion

Our findings suggest that individual differences between global sleep patterns are the main contributors to the effects of sleep on reactivity to emotional faces. Specifically, the average time spent in REM and SWS significantly predicted valence ratings of ambiguous and negative facial expressions, respectively. These strong correlations were evident irrespective of whether the sleep data was taken during the experimental phase or during the earlier (habituation) phase; nor did they change when using proportional amount of time instead of raw time spent in each slept stage. Overall, these results support the notion that trait-like properties of sleep, characterizing each participant as a whole, are strongly linked to emotional reactivity. In contrast, with the exception of difference scores of emotional blunting, no significant effects of day-to-day fluctuations in sleep patterns were evident in our results, suggesting that state-like properties of sleep are less indicative of emotional reactivity to facial expressions, at least as long as natural sleep variability (as opposed to extreme variations) is involved.

The effect of REM on Ambiguity ratings is interesting for two reasons. First, the trait-like characteristic of our results corroborates findings by others suggesting that the ratings of ambiguous faces is a stable property that does not change significantly over time (Neta et al., 2009), much like sleep patterns (De Gennaro et al., 2008; Linkowski, 1999). Second, prior investigations showing an effect of REM on ratings of emotional stimuli have not been consistent. Some studies have found REM to blunt valence ratings, whereas others have found it to intensify such ratings (Gujar et al., 2010; Lara-Carrasco et al., 2009; Rosales-Lagarde et al., 2012; Wagner et al., 2002). Our findings suggest that the discrepancy between these studies may be the result of the rigid classification of emotional stimuli into positive, negative, or neutral, failing to take into account the degree of ambiguity associated with their valence. For example, Gujar et al. (2010) found REM sleep to decrease the degree of negativity with which fearful

---

[3] In a separate analysis (not shown), we also computed similar correlations for each of the original 7 emotional expressions, correcting for multiple comparisons. Results showed significant correlations between REM and valence ratings for surprised faces for four out of the seven days, as well as a highly significant correlation between the average amount of REM and the average rating of surprised faces across days. No other correlation was significant.
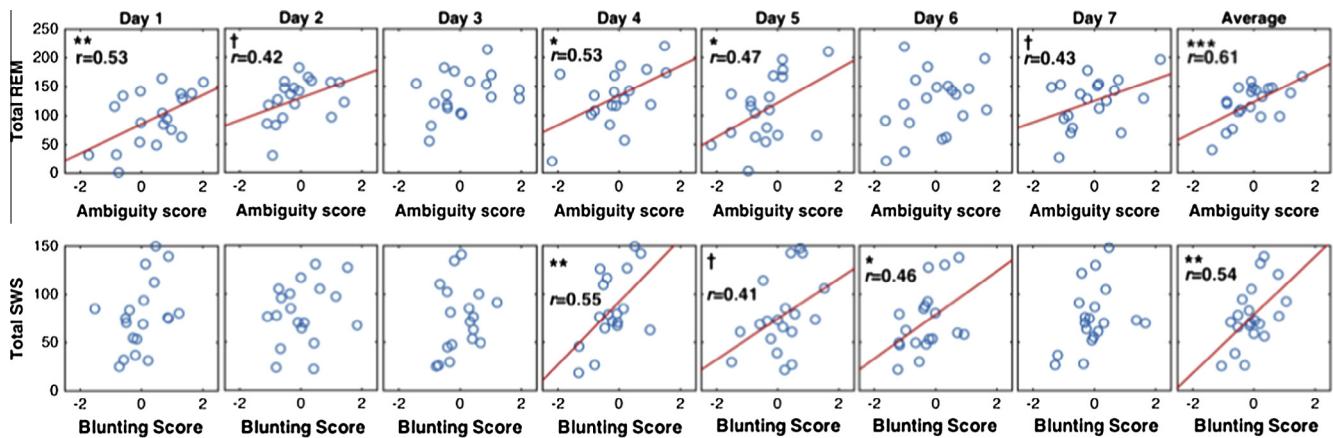
Fig. 4. Correlations between daily REM/SWS (in minutes) and scores of the Ambiguity/Emotional Blunting components, respectively. $^{*}p < 0.05$; $^{**}p < 0.02$; $^{***}p < 0.005$; $^{\dagger}p < 0.08$.

expressions are perceived. In contrast, Wagner et al. (2002) found REM sleep to intensify the perceived negativity of negative stimuli. Whereas these results seem contradictory, we suggest that the "negative" value of the stimuli in the two studies cannot be treated equally. The findings by Gujar and colleagues were based on averaging stimuli varying from neutral to fearful rather than using strictly fearful expressions, whereas Wagner and colleagues' negative stimuli were purely negative. Our study suggests that fearful expressions contribute to an ambiguity dimension and may be perceived as less negative following REM sleep only as part of such dimension (indeed, we did not find a correlation between REM and valence ratings of strictly fearful expressions; see footnote 2). It may therefore be that by averaging over fearful and neutral dimensions, as was done in the study by Gujar and colleagues, the intensity of the expressions were inadvertently dampened, making them more ambiguous and consequently rendering them more susceptible to correlations with REM sleep. This interpretation is consistent with another study (Van Der Helm et al., 2010) showing the effect of sleep on valence ratings to be particularly pronounced when the stimuli in question are more ambiguous.

One could argue that our first PCA component does not strictly represent evaluation of ambiguous emotions. Rather, while the main expression that contributes to this component is *Surprise* – which is widely considered ambiguous (Neta et al., 2009) – it was also influenced by expressions such as *Disgust* and, to a lesser extent, *Fear*, which are not ambiguous per se. An alternative interpretation of the components contributing to this dimension is that they represent expressions that are more complex and that often involve orientation to external stimuli, making them more difficult to learn. According to Widen (2013; see also Jack, Garrod, & Schyns, 2014), different facial expressions are learned at different rates throughout development. Whereas *Happy*, *Sad* and *Angry* are already learned at an early age as part of a basic emotional dichotomy between what "feels good" and what "does not feel good", *Surprised*, *Disgust* and *Fearful*, being emotions that depend not only on internal states but also on external stimuli, require more experience and are therefore learned later in life. According to this interpretation, our results indicate that REM sleep influences how positively/negatively we conceive complex emotions rather than ambiguous ones. Nevertheless, given that ambiguity and complexity are related to each other, both being characteristics of facial expressions that are multi-faceted, our analysis of how REM influences this component holds regardless of whether it is interpreted as one or the other.

Unlike the relationship between REM sleep and emotional reactivity to faces, previous findings in the literature have rarely implicated SWS in this processing. There are several possible

reasons for this discrepancy. First, our use of PCA to extract the most significant dimensions influencing valence scores of emotional faces may have increased the sensitivity of our measures in detecting pure emotional components (by reducing, for example, noise stemming from less relevant processes occurring in parallel, such as face recognition). Indeed, no SWS effect was detected when we examined each of the original facial expressions at each night, independently (see footnote 2). Second, the use of multiple nights may have been, by itself, crucial in detecting SWS effects; as is evident in Fig. 5, when examining the correlations between SWS and Emotional Blunting scores separately for each night, significant or trend-level effects were reached in only three out of the seven nights (compared to five out of seven nights when examining REM and Ambiguity scores). This suggests that such an effect may be missed when conducting single night studies. Finally, and most important, our finding of opposite effects of between- and within-subject SWS on emotional blunting, depending on whether raw or difference scores are used, suggests that previous studies could have missed such effects by either choosing an improper design, or possibly even due to the opposing effects canceling each other. For example, Van Der Helm et al. (2010) have examined sensitivity to facial expressions following sleep versus sleep deprivation, without comparing it to a baseline measure in the previous evening. They found sleep enhances blunting of positive and negative emotions, very much like the between-subject effect we found for SWS on raw Emotional Blunting scores (in their study no sleep-staging measures were taken, so there was no way of attributing the effect to a specific sleep stage). On the other hand, two other studies (Gujar et al., 2010; Wagner et al., 2002) have found some evidence of SWS contributing to blunting of negative emotions when comparing scores before and after sleep to a control group that stayed awake, but the effect was marginal. These results might reflect a mixture of conflicting between- and within- subject influences, summing up to a weak effect. If, indeed, effects of SWS on emotional processing are repeatedly missed in single night studies due to lack of reliable measures of baseline sleep patterns, this calls for significant modifications in the standard procedures when investigating the effects of sleep on various aspects of cognition.

Finally, with regard to the effect of Day on Emotional Blunting scores, visual inspection of the average valence scores across participants suggested that this might have stemmed from a mild decrease in the scores along the experimental sessions. One possibility to explain this pattern is that participants have gotten better over time in distinguishing between negative and more ambiguous faces (such as surprised), increasing their confidence in giving lower scores to the negative expressions and resulting in reduced ratings. However, since this effect did not interact with sleep, we
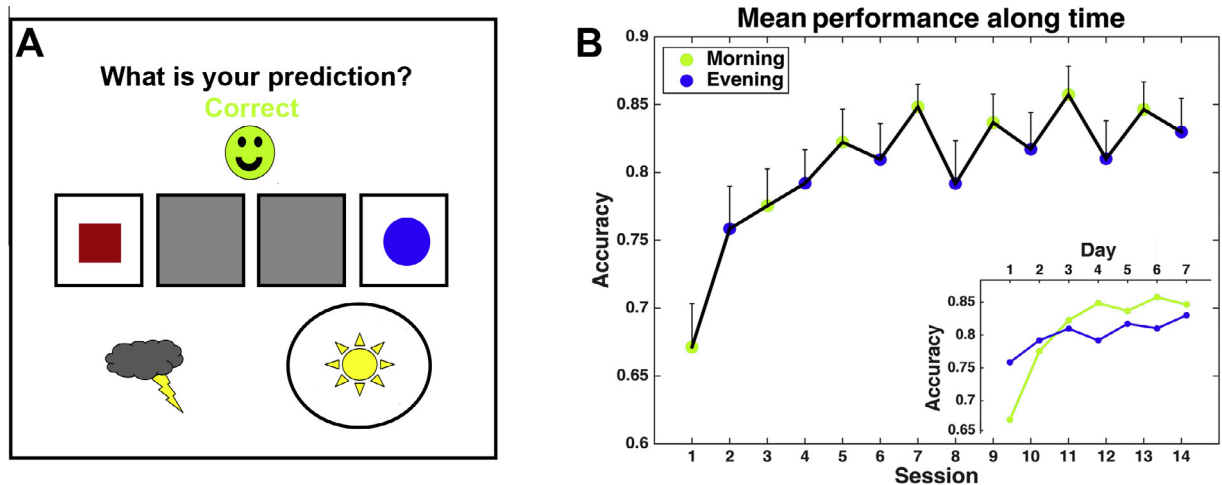
Fig. 5. A: Example of a trial in the Weather Prediction Task; B: Mean accuracy in the Weather Prediction Task along time. Sessions interchanged between mornings and evenings for 7 days, with the first session always performed in the morning. Inset: Mean performance by day, broken down to morning and evening sessions.

did not consider it to be central to our investigations and thus did not pursue it further.

4. Experiment 2: Weather Prediction Task

4.1. Task description

The Weather Prediction Task (Gluck, Shohamy, & Myers, 2002; Knowlton, Squire, & Gluck, 1994) assesses participants' ability to learn stimulus-outcome pairings in a feedback-driven, probabilistic, category-learning paradigm. The task utilizes four cards, each containing a single, colored, geometric shape (red square, green triangle, blue circle, purple diamond). In each trial, a combination of one, two, or three of the cards is presented (14 possible combinations). Participants are required to determine whether the cards predicted one of two outcomes, 'rain' or 'sun'. Each of the 14 combinations is associated with a probability of predicting the two outcomes (e.g., the probability of rain being the "correct" response given that the red[4] square and the blue circle are present, as pictured in Fig. 2, is 16.7%; for a full list of outcomes and associated probabilities see Gluck et al., 2002). The correct mapping between the card combinations and the required responses was fixed for all sessions throughout the experiment, although the order of the trials was randomized. Participants were informed about the probabilistic nature of the task and, as such, were instructed to select the answer that was correct "most of the time". Once participants selected one of the two possible outcomes (see Fig. 5A) they were given immediate feedback as to whether or not they were correct. Subsequent trials began one second after participants made their selection. Participants experienced 100 trials during each of the 14 testing sessions.

4.2. Results

4.2.1. Sleep and daily scores

Participants' accuracy rates in each session were calculated as the total number of optimal responses (divided by 100), with an optimal response for a given combination of cards being the one that offers the higher probability of correct response for that combination. In other words, if a certain combination predicts 'rain' 85% of the time and 'sun' 15% of the time, the optimal answer for

such a combination was considered 'rain' even for trials where the feedback given indicated that 'sun' was correct.

In order to assess how sleep affects daily learning, we ran mixed-model ANOVAs, with mean percentage of correct responses (PC) for each session as the predicted variable. Within- and between-subjects factors were defined as in the first experiment. Tests of covariates and fixed effects for this model are presented in Table S4. As expected, there was a significant effect of Day $(F_{(6, 105)} = 6.06, p < 0.0001)$, signifying an improvement in performance as participants learned the task; however, this effect was qualified by a Time × Day interaction, indicating that morning and evening performance improved at different rates $(F_{(6, 105)} = 2.86, p = 0.0128)$. The main trends in performance are illustrated in Fig. 5B. Whereas during the first two experimental days the evening testing sessions were better than morning sessions due to rapid increase in performance, around day 3, after learning reached a plateau, this ordering was reversed, with performance during morning sessions becoming better than the evenings. Importantly, the analysis also showed a significant between-subjects effect of N1/N2 across sessions $(F_{(1, 16)} = 11.76, p = 0.0034;$ Table S5). No other sleep effect was significant.[5]

Because of the Time × Day interaction on session accuracy, we proceeded to analyze morning and evening sessions separately to achieve a clearer picture of participants' learning. Separate mixed-model ANOVAs were run on the morning and evening PC scores across days, with Day and N1/N2 (including both between- and within-subjects effects) as predictors. A first-order autoregressive AR(1) error structure was used across the levels of Day. The results are presented in Table S6. For morning sessions, mean accuracy increased across days, as indicated by a significant Day effect $(F_{(6, 104)} = 8.98, p < 0.0001)$. N1/N2 was also a predictor of accuracy, although this effect was only marginally significant $(F_{(1, 18)} = 3.65, p = 0.0720)$. For evening session accuracy, the effect of Day was again significant $(F_{(6, 104)} = 2.22, p = 0.0471)$. The effect of mean total N1/N2 was significant, with parameter

---

[4] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

[5] As in the first experiment, we also examined models that included interactions of the sleep factors with Day and Time-of-day. These models were inferior to the one reported based on goodness of fit measures. These models did, however, reveal a significant SWScovdev ∗ Time-of-day interaction. Investigating this effect further by separating analysis to mornings and evenings revealed a marginally significant effect of SWScovdev on evening PC scores. This effect, however, was not replicated when using proportional measures of SWS. Given that the effect appeared only with raw scores, was only marginally significant, and was based on an inferior model in terms of goodness-of-fit measures, we did not pursue it further.

estimations showing accuracy increasing as the average amount of N1/N2 increases ($F(1,16) = 7.58$, $p = 0.0131$).

As in the first experiment, we next examined whether total sleep time (both within and between-subject) was predictive of the behavioral measures. We found a marginally significant effect of between-subject total sleep time on evening scores ($F(1,18) = 2.89$, $p = 0.0641$), but no effect on morning scores. To ensure that our stage-specific results did not stem from variations in total sleep time, we then examined how the accuracy scores were influenced by the proportion of time spent in N1/N2 out of total sleep time. Results are presented in Table S7. For morning session accuracy, the effect of Day was significant ($F(6,104) = 9.39$, $p < 0.0001$), while the effect of mean proportional N1/N2 was marginally significant ($F(1,18) = 3.01$, $p = 0.0999$). Interestingly, the effect of nightly fluctuations in proportional N1/N2 (pN1/N2covdev in Table S6), previously null, also became marginally significant ($F(1,104) = 2.97$, $p = 0.0877$). For evening session accuracy, the effect of N1/N2 was significant ($F(1,18) = 6.55$, $p = 0.0197$), as was the effect of Day ($F(6,104) = 2.29$, $p = 0.0403$). Accuracy improved as the proportional amount of N1/N2 increased.

To conclude this part of the analysis, we examined, as before, whether the between-subject effect of N1/N2 is maintained when using the mean values from the habituation phase rather than the experimental phase. To that end, we ran a model using the average amount of time spent in each sleep stage over the whole habituation phase as between-subject factors, and Day as a within-subject factor, separately for mornings and evenings. This analysis yielded a significant effect of N1/N2 on PC scores on both times ($F(1,16) = 6.85$, $p = 0.0186$; $F(1,16) = 9.29$, $p = 0.0077$; respectively). Once again, the effect was maintained when using the proportional measure of N1/N2 as the between-subject factor ($F(1,16) = 5.42$, $p = 0.0317$; $F(1,18) = 8.61$, $p = 0.0312$; for mornings and evenings, respectively).

Finally, as in the first experiment, rather than examining whether sleep modulated the raw PC scores, we examined if it modulated day-to-day improvements in those scores. For each participant, we calculated the difference in accuracy between each morning session and the preceding evening session (for the morning of Day 1, we assumed a 0.5 success rate as the baseline, representing random chance, allowing us to use data from all 7 nights). These scores were then subject to a mixed-model ANOVAs with all within and between-subject sleep factors, as well as Day, as predictors. Results (see Table S8) showed no effects of sleep. Only the effect of Day was significant ($F(6,102) = 5.10$, $p = 0.0001$), an expected result signifying improvements in learning were slowed down as the experiment progressed (see Fig. 7).

4.2.2. Sleep and overall learning trends

In addition to success rates in each session, we investigated whether average sleep predicts overall learning trends. To accomplish this, the success rates across the 14 sessions of each participant were fitted with a sigmoid learning curve unique to that participant. Each participant's performance was then characterized by two measures based on these curves: Maximum performance level (defined as the accuracy when learning reached a plateau), and learning time (defined as the time it took to reach accuracy level of 95% of the maximum performance level).[6] Examples of

---

[6] Each learning curve, representing the participant's accuracy level along time, was hypothesized to initiate when $t = 0$ at an accuracy level of 0.5 (random chance) and described by the sigmoid equation $F(t) = c/(1 + \exp(-a(t - b))) + 0.5 - c/(1 + \exp(ab))$. For each participant, we then fitted a,b and c by search through the parameter space to find the minimum mean squared error between the curve at time points $t = 1,2,\ldots0.14$ and the data from the 14 experimental sessions. One subject did not show a good fit to a sigmoid curve (and also had atypical learning based on strategy analysis as presented in Gluck et al., 2002, not detailed here for lack of space), therefore this subjects' learning plateau was determined manually.

the fitted curves are presented in Fig. 6, showing that all but three participants learned the task well.

A multiple regression model with average total time in N1/N2, SWS and REM sleep as predictor variables was run separately for the two performance measures, maximum learning level at plateau and speed of learning. There was only a weak trend on the maximum performance level measure, $F(3,16) = 2.38$, $p = 0.1077$, with N1/N2 being the sole predictor to significantly contribute to that trend ($p = 0.0355$; see Fig. 7A). In contrast, there was a significant effect of sleep on learning time ($F(3,16) = 4.04$, $p = 0.0257$), driven by a positive correlation with SWS ($p = 0.0056$). The higher their average SWS, the quicker participants reached their learning plateau (Fig. 7B). This effect also held when examining the correlation between the average proportional time spent in SWS and speed of learning ($r = -0.50$, $p = 0.0244$. Like before, we also examined correlations between total sleep time and each of the two performance measures but no significant effects were found). Finally, when using sleep-stage measures taken from the habituation phase rather than the experimental phase, we found, again, a significant effect of sleep on learning time, $F(3,16) = 5.38$, $p = 0.0113$, which was driven by both N1/N2 and SWS ($p = 0.0209$, $p = 0.0032$, respectively). This last effect did not hold, however, when using the proportional measures of sleep during habituation.

4.3. Discussion

The main results from our second experiment echo those from the valence-rating task: Sleep affects performance in the Weather Prediction Task in a between-subject, trait-like manner. Individual differences in average N1/N2 predicted individual differences in accuracy, with a positive correlation between the two. On the other hand, individual differences in average SWS predicted individual differences in learning time, with more SWS leading to faster learning. These effects were evident with both raw and proportional sleep measures, as well as when using measures from the habituation phase rather than the experimental phase.

Performance in the weather prediction task is known to employ both the hippocampus and the striatum in a competitive way, with the former being most conspicuous during the very early stages of learning when initial coding of the stimuli takes place, and then again later on following training over hundreds of trials (Poldrack et al., 2001). Since performance in hippocampal-related tasks is sometimes modulated by SWS (Diekelmann & Born, 2010), one could expect a similar effect in the Weather Prediction Task. Previous studies from our lab and others examined the effects of a single night of sleep on performance in several variations of the Weather Prediction Task (Barsky et al., 2015; Djonlagic et al., 2009). It was found that sleep (compared to wake) enhances performance in some conditions (see below); however, no correlation between SWS and performance the following day was found. Instead, both studies found a correlation with REM sleep (though of a different kind; whereas one study reported correlations between REM and performance improvement following sleep, the other found a correlation between REM and performance in the *preceding* night, with a lack of any performance benefits the next day).

The discrepancy between the aforementioned studies and the current one may be explained by key methodological differences. First, participants in the earlier studies did not experience regular nocturnal sleep; in one study participants experienced a daytime nap, and in the other, despite the sleep occurring at night, participants reportedly slept significantly less than their typical sleep schedule (likely due to frequent awakenings in the unnatural environment of a sleep lab; see Supplementary Material in Djonlagic et al., 2009). Second, the effects in the previous studies were only obtained in a specific "observational" variation of the task wherein
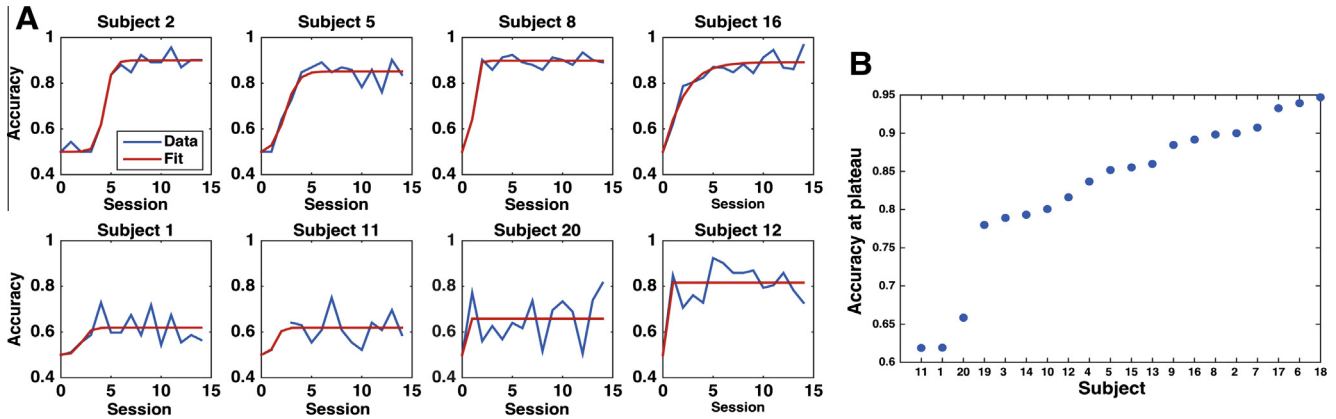
Fig. 6. Learning trends of participants in the Weather Prediction Task. A: Examples of learning data and their fitted curves. Upper row: Typical subjects. Lower row: Subjects with poor learning, and an additional subject that showed a non-sigmoid learning curve. B: Accuracy at learning plateau, sorted by subject from worst to best.
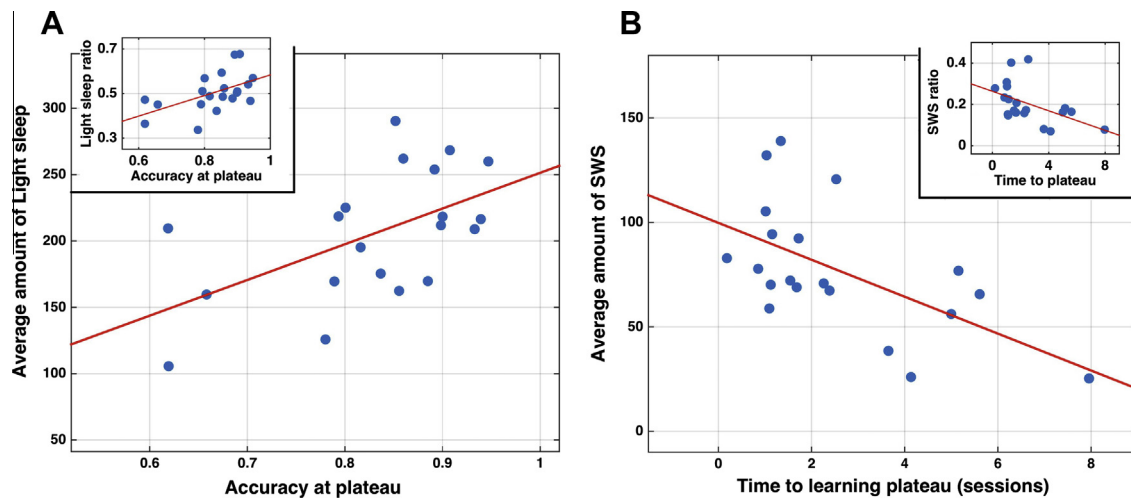

Fig. 7. Between-subject correlations of sleep and learning curve properties. A: Average amount of N1/N2 as a function of accuracy at learning plateau. Inset: Same, for proportional amount of N1/N2 out of total sleep time. B: Average amount of SWS as a function of time to reach learning plateau. Inset: Same, for proportional amount of SWS out of total sleep time.

subjects were trained by immediately receiving the correct answer to each card combination instead of learning through trial and error. This version of the task may be processed and consolidated differently that the feedback-based version used in the current work. Results from neuroimaging studies suggest that in the observational version of the task, hippocampal involvement in learning does not compete with striatal regions, unlike the feedback-based version (Poldrack et al., 2001). Moreover, when using the observational version, there is evidence that the effect of REM sleep on performance involves a different processing mechanism than the ones used for initial encoding of the stimuli, presumably within the hippocampus (Barsky et al., 2015). Taken together, it is tempting to hypothesize that the effects of REM on performance in these previous studies were unrelated to hippocampal-dependent consolidation. On the other hand, given that in the feedback version of the task the hippocampus is involved in both early and late stages of training, the focus of our study on global measures of performance over many trials and days may have been better suited to tap into hippocampal-based consolidation and thus allowed us to detect the hypothesized SWS effect.

Alternatively, our SWS results could be explained by the synaptic homeostasis hypothesis (Tononi & Cirelli, 2006). According to this theory, learning during wake causes a net increase in synaptic strength that is both toxic and energetically costly. In order to compensate for this increase, during SWS a "renormalization"

process takes place, which restores synaptic homeostasis. As part of this process, weak synaptic connections are pruned, leading to an increased signal-to-noise ratio when accessing stored memories and, consequently, to improved performance following sleep (see Tononi & Cirelli, 2014 for a more detailed description of this process). In light of this hypothesis, one could argue that SWS, rather than enhancing performance in the weather prediction task through hippocampal-dependent processes, increases the stability and accessibility of memories as a whole, a process that is inherently accumulative and ongoing. As a result, it is mostly pronounced when examining the relations between SWS and performance on a trait level rather than direct effects from one day to another.

The finding that individual differences in N1/N2 predict individual differences in average accuracy is more surprising. N1/N2 includes sleep stages N1 and N2, with N2 comprising the large majority (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004). Therefore, results attributed to N1/N2 are most likely tied to N2. Previous studies in humans have found correlations between performance in hippocampal-dependent tasks and time in N2 (Meier-Koll, Bussmann, Schmidt, & Neuschwander, 1999; Peigneux et al., 2004; Van Der Helm, Gujar, Nishida, & Walker, 2011), as well as between performance in such tasks and sleep spindles, which occur mostly in N2 (e.g., Gais, Mölle, Helms, & Born, 2002; Schabus et al., 2004; Van Der Helm et al., 2011).

However, overall, findings relating performance in hippocampal tasks to SWS are far more common. It is therefore possible that the repeated monitoring of sleep over multiple days was crucial in allowing us to identify the involvement of N1/N2.

5. General discussion

Using two very different experimental paradigms, one pertaining to hippocampal-dependent learning and the other to emotional reactivity, we showed that effects of sleep on performance are, to a large extent, trait-dependent. Individual differences in sleep characteristics were correlated with individual differences in performance in both paradigms, whereas within-subject variations in sleep were, with one exception, uncorrelated with variations in subsequent performance (see summary in Table 1). These results serve as a warning against a simplified interpretation of the sleep-cognition relationship shown in single-night studies as stemming from a short-term effect of a particular night of natural sleep on behavior the following morning. Further, our results suggest that because the effect of sleep on cognition is trait-like, single night studies may not be sufficient to get a reliable measure of this trait (even when first-night effects are controlled for) since any particular night, given the daily sleep fluctuations, is simply too noisy a measurement (see Fig. 5). Moreover, even when utilizing daily monitoring of sleep, our results suggest that it is insufficient to estimate the trait effect of sleep on cognition based only on total sleep time (for example, using actigraphy) as many of our findings pertain to specific sleep stages.

Several possibilities exist that may explain this basic pattern in our results. One technical possibility is that the variability of sleep measures within-subjects was larger than the one between-subjects, essentially yielding noisy measurements that blurred correlations of the former with other factors. However, this is unlikely. First, we did find some within-subject effects in our first experiment. Second, calculating intra and inter variability over the data, we found lower values for within- than between-subjects ratios for every sleep measure except Total sleep (see Supplementary Table S10; also, see Fig. 5 for qualitative impression). Therefore, it does not seem to have been "easier" to detect between-subjects effects in our data. Another possibility is that the correlations between sleep and cognition are not indicative of a causal relation between the former and the latter; instead, it might be that performance on one day affects sleep the following night rather than the other way around; or, alternatively, that a common factor affects both sleep and behavior independently, yielding a correlation between them that does not reflect a direct influence. However, such relations per se do not seem to be sufficient to account for the full spectrum of results in our and others' studies, for several reasons. First, our finding that sleep patterns in the habituation phase are predictive of performance a week later cannot be accounted for by a relationship in which performance affects subsequent sleep. Second, there is clear and strong evidence in

the literature, coming from studies employing manipulations of sleep in both human and animals, that sleep can affect subsequent behavior directly (see Rasch & Born, 2013, for review). A support for this effect was also seen in the within-subject influence of SWS on Emotional Blunting difference scores in our first experiment.

Given this literature, we hypothesize two alternative mechanisms that may explain our results in the context of previous findings. One possible mechanism is that there is a direct causal effect of sleep on subsequent performance, but this effect has an accumulative nature such that natural variations in sleep from one night to another are barely enough to significantly influence cognitive capabilities in the following day; rather, it takes incremental changes over many days to create a substantial effect. Thus, persistent differences in sleep patterns are generally required to produce differences in performance. Nevertheless, if large changes in sleep patterns are introduced, such as when pulling an all-nighter before an exam or when extreme experimental manipulations of sleep are implemented; or, alternatively, when increasing the sensitivity of the behavioral measures by comparing scores before and after sleep, an effect on behavior can become substantial after a single night (see illustration in Fig. 8A). Another possible mechanism is that sleep and performance are correlated indirectly by a third factor that affects them both (e.g., a genetic tendency for high stress levels; a developmental factor such as the efficiency of functional networks in the brain, etc.). However, there is also a minimum amount of sleep (or a specific sleep stage) that is required to maintain daily functioning. Natural variations in sleep are not extreme enough to go below that minimum level and therefore do not cause aversive effects on performance; but, if sleep is drastically disturbed during a specific night, the threshold is crossed and subsequent cognitive performance may deteriorate as a result (Fig. 8B). Ultimately, more research is required to determine which of the two models, if any, is correct, and if these results generalize to additional paradigms beyond the ones presented in this study.

Our findings have practical implications as well. They suggest that better cognitive and emotional functioning may not be achieved by, for example, improving one's sleep on a certain important night (e.g., before taking an exam) beyond one's typical average; rather, if sleep-related improvement can be achieved at all, it will likely require general change in sleep hygiene that takes place over an extended period of time, thus altering average sleep patterns. This too, however, is uncertain given the possibility that an independent developmental factor is responsible for the correlations we observed.

Finally, it should be noted that the interest in and demand for consumer-level sleep-monitoring devices has been constantly increasing in the last decade, with more and more products available on the market (for review see Chen et al., 2013). Some of these devices, being easy to use for multiple nights while maintaining relative validity, have the potential to not only provide easy access to quantifiable information regarding one's own sleep but also allow the expansion of sleep and cognition research into large, longitudinal studies in natural environments. Our study provides one

Table 1
Summary of sleep effects in the two experiments.

Type of Effect	Task	Sleep Stage		
		REM	SWS	N1/N2
Between-Subject Effects	Valence Rating Task	↑ Evaluation of ambiguity as positive in emotional stimuli	↑ Tendency towards blunting of emotional stimuli	___
	Weather Prediction Task	___	↓ Time to reach maximum performance	↑ Average performance
Within-Subject Effects	Valence Rating Task	___	↓ Overnight increase in emotional blunting	___
	Weather Prediction Task	___	___	___

Note. Direction of arrow represents the direction of change in performance as the corresponding amount of time in a sleep stage goes up.
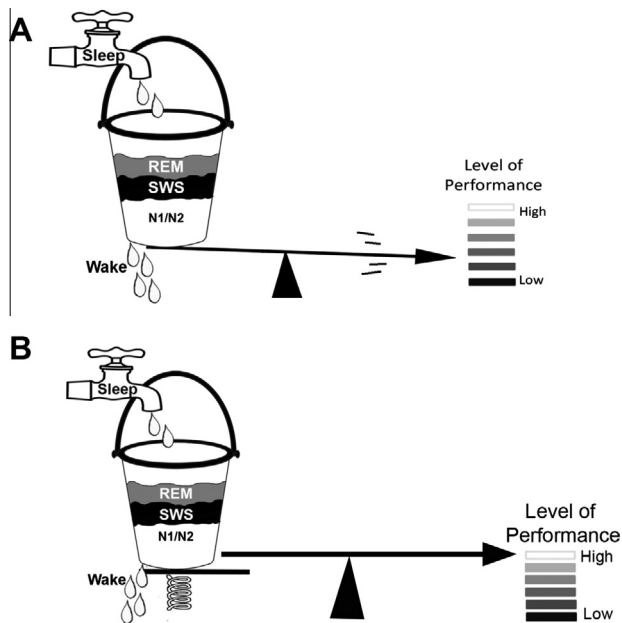
Fig. 8. Two possible causal mechanisms for the effect of sleep on performance. A: Sleep affects performance directly; however, small daily variations in sleep vs. wake do not have a significant effect, and only large, accumulated variations are translated to detectable performance modulation. B: Sleep and performance are affected by a third factor, but sleep can also affect performance directly if it is reduced below a certain threshold. See text for details.

of the first pieces of evidence that utilization of such devices in parallel to cognitive testing can replicate known results in the field that were previously obtained in PSG-based, single-night studies, as well as extend these studies to allow both the discovery of new sleep-cognition relations that depend on reliable, individual baselines, and a better appreciation of how these relations progress over time.

6. Conclusions

We have found that natural day-to-day variations in sleep architecture do not predict variations in performance in emotional and cognitive tasks the following morning; rather, individuals' trait-like characteristics of sleep architecture predict trait-like performance in these tasks. Our results highlight the necessity of obtaining reliable baseline sleep measures in future studies of sleep-cognition relations, as well as the possible advantage of monitoring sleep and behavior in natural home environments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.nlm.2016.07.032.

References

Agnew, H. W., Webb, W. B., & Williams, R. L. (1966). The first night effect: An EEG study of sleep. *Psychophysiology, 2*(3), 263–266. http://dx.doi.org/10.1111/j.1469-8986.1966.tb02650.x.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. (2003). The role of actigraphy in the study of sleep and circadian rhythms. American academy of sleep medicine review paper. *Sleep, 26*(3), 342–392.

Barsky, M. M., Tucker, M. A., & Stickgold, R. (2015). REM sleep enhancement of probabilistic classification learning is sensitive to subsequent interference. *Neurobiology of Learning and Memory, 122*, 63–68. http://dx.doi.org/10.1016/j.nlm.2015.02.015.

Burke, T. M., Scheer, F. A., Ronda, J. M., Czeisler, C. A., & Wright, K. P. (2015). Sleep inertia, sleep homeostatic and circadian influences on higher-order cognitive functions. *Journal of Sleep Research*. http://dx.doi.org/10.1111/jsr.12291.

Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., ... Campbell, A. T. (2013). *Unobtrusive sleep monitoring using smartphones*. Paper presented at the Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th international conference on. http://dx.doi.org/10.1145/2517351.2517359.

De Gennaro, L., Marzano, C., Fratello, F., Moroni, F., Pellicciari, M. C., Ferlazzo, F., ... Sforza, E. (2008). The electroencephalographic fingerprint of sleep is genetically determined: A twin study. *Annals of Neurology, 64*(4), 455–460. http://dx.doi.org/10.1002/ana.21434.

de Souza, L., Benedito-Silva, A. A., Pires, M. N., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep, 26*(1), 81–85.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114–126. http://dx.doi.org/10.1038/nrn2762.

Djonlagic, I., Rosenfeld, A., Shohamy, D., Myers, C., Gluck, M., & Stickgold, R. (2009). Sleep enhances category learning. *Learning & Memory, 16*(12), 751–755. http://dx.doi.org/10.1101/lm.1634509.

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-dependent increases in sleep spindle density. *The Journal of Neuroscience, 22*(15), 6830–6834.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory, 9*(6), 408–418. http://dx.doi.org/10.1101/lm.45202.

Griessenberger, H., Heib, D., Kunz, A., Hoedlmoser, K., & Schabus, M. (2013). Assessment of a wireless headband for automatic sleep scoring. *Sleep and Breathing, 17*(2), 747–752. http://dx.doi.org/10.1007/s11325-012-0757-4.

Groch, S., Wilhelm, I., Diekelmann, S., & Born, J. (2013). The role of REM sleep in the processing of emotional memories: Evidence from behavior and event-related potentials. *Neurobiology of Learning and Memory, 99*, 1–9. http://dx.doi.org/10.1016/j.nlm.2012.10.006.

Gujar, N., McDonald, S. A., Nishida, M., & Walker, M. P. (2010). A role for REM sleep in recalibrating the sensitivity of the human brain to specific emotions. *Cerebral Cortex*, bhq064. http://dx.doi.org/10.1093/cercor/bhq064.

Iber, C., Ancoli-Israel, S., Chesson, A., & Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology, and technical specifications* (1st ed.). Westchester, Illinois: American Academy of Sleep Medicine.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology, 24*(2), 187–192. http://dx.doi.org/10.1016/j.cub.2013.11.064.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport, 14*(18), 2317–2322. http://dx.doi.org/10.1097/01.wnr.0000101520.44335.20.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory, 1*(2), 106–120. http://dx.doi.org/10.1101/lm.1.2.106.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377–1388. http://dx.doi.org/10.1080/02699930903485076.

Lara-Carrasco, J., Nielsen, T. A., Solomonova, E., Levrier, K., & Popova, A. (2009). Overnight emotional adaptation to negative stimuli is altered by REM sleep deprivation and is correlated with intervening dream emotions. *Journal of Sleep Research, 18*(2), 178–187. http://dx.doi.org/10.1016/j.nlm.2015.02.008.

Le Bon, O., Staner, L., Hoffmann, G., Dramaix, M., San Sebastian, I., Murphy, J. R., ... Linkowski, P. (2001). The first-night effect may last more than one night. *Journal of Psychiatric Research, 35*(3), 165–172. http://dx.doi.org/10.1016/S0022-3956(01)00019-X.

Linkowski, P. (1999). EEG sleep patterns in twins. *Journal of Sleep Research, 8*(1), 11–13. http://dx.doi.org/10.1046/j.1365-2869.1999.00002.x.

Meier-Koll, A., Bussmann, B., Schmidt, C., & Neuschwander, D. (1999). Walking through a maze alters the architecture of sleep. *Perceptual and Motor Skills, 88*(3c), 1141–1159. http://dx.doi.org/10.2466/pms.1999.88.3c.1141.

Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion, 9*(5), 640. http://dx.doi.org/10.1037/a0029431.

Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. *Sleep, 27*, 1255–1274.

Pantic, M., & Rothkrantz, L. J. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1424–1445. http://dx.doi.org/10.1109/34.895976.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., ... Aerts, J. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron, 44*(3), 535–545. http://dx.doi.org/10.1016/j.neuron.2004.10.007.

Plihal, W., & Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience, 9*(4), 534–547. http://dx.doi.org/10.1162/jocn.1997.9.4.534.

Poldrack, R. A., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. (2001). Interactive memory systems in the human brain. *Nature, 414*(6863), 546–550. http://dx.doi.org/10.1038/35107080.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681–766. http://dx.doi.org/10.1152/physrev.00032.2012.

Rosales-Lagarde, A., Armony, J. L., del Río-Portilla, Y., Trejo-Martínez, D., Conde, R., & Corsi-Cabrera, M. (2012). Enhanced emotional reactivity after selective REM

sleep deprivation in humans: An fMRI study. *Frontiers in Behavioral Neuroscience, 6.* http://dx.doi.org/10.3389/fnbeh.2012.00025.

Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klosch, G., Anderer, P., ... Zeitlhofer, J. (2004). Sleep spindles and their significance for declarative memory consolidation. *Sleep, 27*(8), 1479–1485.

Shambroom, J. R., Fabregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research, 21* (2), 221–230. http://dx.doi.org/10.1111/j.1365-2869.2011.00944.x.

Shohamy, D., Myers, C., Onlaor, S., & Gluck, M. (2004). Role of the basal ganglia in category learning: How do patients with Parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676. http://dx.doi.org/10.1037/0735-7044.118.4.676.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10*(1), 49–62.

Tononi, G., & Cirelli, C. (2014). Sleep and the price of plasticity: From synaptic and cellular homeostasis to memory consolidation and integration. *Neuron, 81*(1), 12–34.

Van Der Helm, E., Gujar, N., Nishida, M., & Walker, M. P. (2011). Sleep-dependent facilitation of episodic memory details. *PLoS ONE, 6*(11), e27421. http://dx.doi.org/10.1371/journal.pone.0027421.

Van Der Helm, E., Gujar, N., & Walker, M. P. (2010). Sleep deprivation impairs the accurate recognition of human emotions. *Sleep, 33*(3), 335–342.

Volk, H. E., McDermott, K. B., Roediger, H. L., & Todd, R. D. (2006). Genetic influences on free and cued recall in long-term memory tasks. *Twin Research and Human Genetics, 9*(05), 623–631. http://dx.doi.org/10.1375/183242706778553462.

Wagner, U., Fischer, S., & Born, J. (2002). Changes in emotional responses to aversive pictures across periods rich in slow-wave sleep versus rapid eye movement sleep. *Psychosomatic Medicine, 64*(4), 627–634. http://dx.doi.org/10.1097/01.PSY.0000021940.35402.51.

Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences, 1156*(1), 168–197. http://dx.doi.org/10.1111/j.1749-6632.2009.04416.x.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review, 5*(1), 72–77. http://dx.doi.org/10.1177/1754073912451492.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*(3), 671–684. http://dx.doi.org/10.3758/BRM.42.3.671.