



ORIGINAL ARTICLE

# Do your troubles today seem further away than yesterday? On sleep's role in mitigating the blushing response to a reactivated embarrassing episode

Faya L. Reinhold<sup>1</sup>, Anna M. V. Gerlicher<sup>1,2</sup>, Eus J. W. van Someren<sup>3,4,5,6</sup> and Merel Kindt<sup>1,\*</sup>

<sup>1</sup>Department for Clinical Psychology, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WS Amsterdam, The Netherlands,

<sup>2</sup>Department for Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands,

<sup>3</sup>Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands, <sup>4</sup>Department of Integrative Neurophysiology, Center for Neurogenomics and Cognitive Research (CNCR), Amsterdam

Neuroscience, Vrije Universiteit, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands and <sup>5</sup>Department of Psychiatry, Amsterdam UMC, Amsterdam Neuroscience, Vrije Universiteit, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

<sup>6</sup>University of Amsterdam at which the research was conducted.

\*Corresponding author. Merel Kindt and Faya Louisa Reinhold, University of Amsterdam, Nieuwe Achtergracht 129B, 1018 WS Amsterdam, The Netherlands. Email: [m.kindt@uva.nl](mailto:m.kindt@uva.nl).

## Abstract

The “*sleep to forget and sleep to remember hypothesis*” proposes that sleep weakens the emotional tone of an experience while preserving or even enhancing its content. Prior experimental research however shows contradictory findings on how emotional reactivity changes after a period of sleep, likely explained by methodological variations. By addressing these inconsistencies, we investigated the mitigating effect of overnight sleep on emotional reactivity triggered by memory reactivation. Using a karaoke paradigm, we recorded participants' singing of two songs, followed by exposing them to one of the recordings (rec1) to induce an embarrassing episode. After a 12-hr period of either day-time wakefulness ( $N = 20$ ) or including nighttime sleep ( $N = 20$ ), we assessed emotional reactivity to the previously exposed recording (rec1) and the newly exposed recording (rec2). Emotional reactivity was assessed with a physiological measure of facial blushing as the main outcome and subjective ratings of embarrassment and valence. Sleep and wake were monitored with diaries and actigraphy. The embarrassing episode was successfully induced as indicated by objective and subjective measures. After controlling for an order effect in stimulus presentation, we found a reduction in blushing response to the reactivated recording (rec1) from pre- to post-sleep compared to wakefulness. However, emotional reactivity to the reactivated recording (rec1) and the new recording (rec2) did not differ after sleep and wakefulness. This study shows that facial blushing was reduced following overnight sleep, while subjective ratings were unaffected. Whether the beneficial effect of sleep is due to changes in memory representation or rather emotion regulation remains elusive.

## Statement of Significance

Sleep disturbances are highly prevalent among a broad range of affective disorders and contribute to the persistence of symptoms. While evidence for a generic effect of sleep on emotions is growing, there is still no consensus on whether and in which direction sleep alters the emotional tone of a memory. In contrast to numerous studies relying on subjective measures of emotional reactivity to negatively valenced pictures, we added a physiological measure and created an autobiographical memory to increase the reliability and ecological validity of our study. Here we show that the autonomic blushing response was weakened after overnight sleep, but further research is needed to dissect a memory-specific beneficial effect from a more general effect of sleep on emotion regulation.

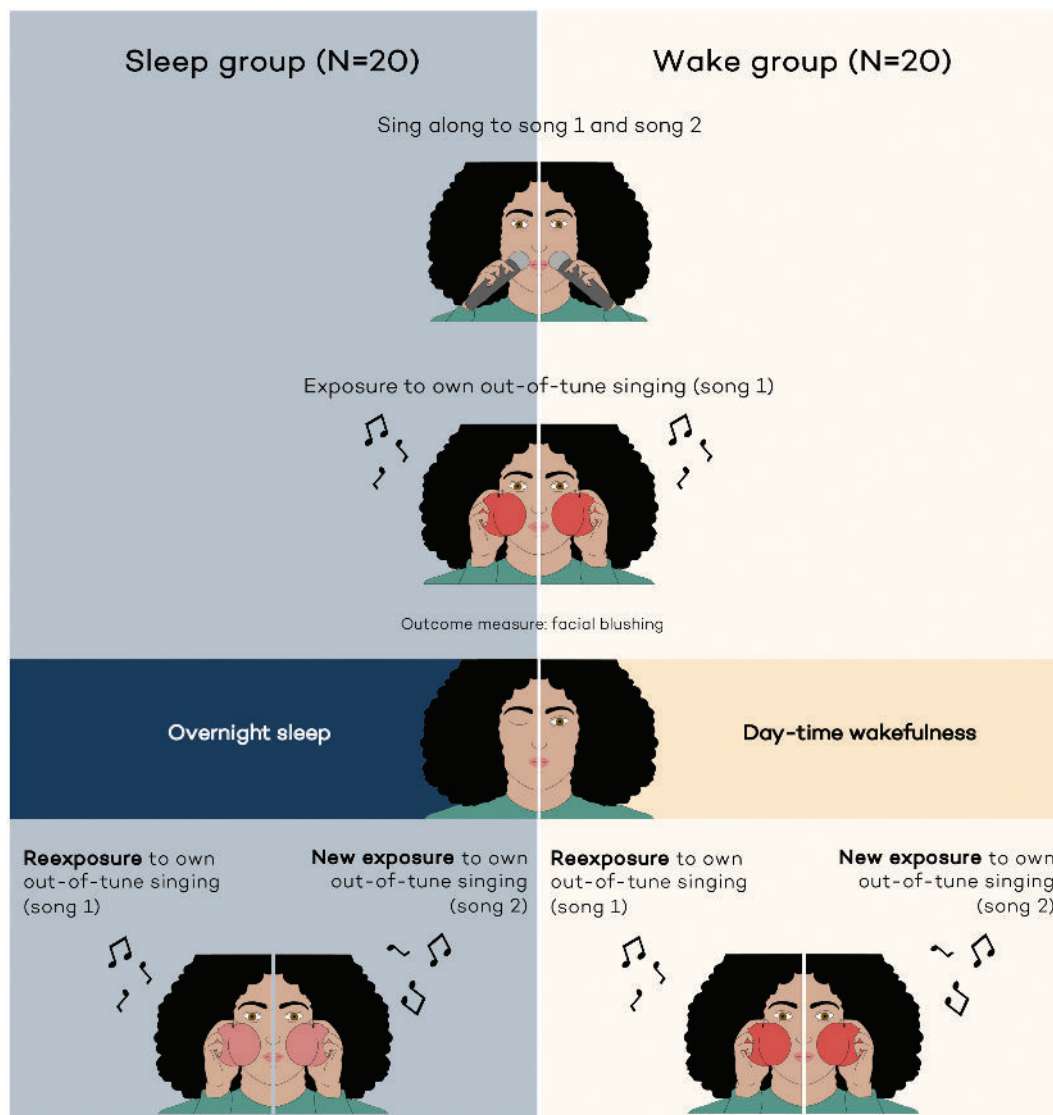
Submitted: 29 April, 2022; Revised: 17 August, 2022

© Sleep Research Society 2022. Published by Oxford University Press on behalf of the Sleep Research Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Graphical Abstract

## ON SLEEP'S ROLE IN MITIGATING THE **BLUSHING RESPONSE** TO A REACTIVATED EMBARRASSING EPISODE



**CONCLUSION** While facial blushing was reduced after overnight sleep compared to daytime wakefulness, the smaller response was not unique to the re-exposed song, but also occurred for the newly exposed song. Whether the beneficial effect of sleep on facial blushing is due to changes in memory representation or rather emotion regulation remains elusive.

**Key words:** embarrassment; sleep; emotional memories; facial blushing; karaoke

### Introduction

Sleep disturbances are very common among a broad range of affective disorders: from depression to posttraumatic stress disorder [1–3]. There is growing evidence for a bidirectional link between sleep and affect where sleep disturbances do not only follow from psychopathology but also contribute to the persistence of symptoms [4, 5]. Beyond a generic effect of sleep on

emotions, it has been proposed that sleep has an influence on emotional memories as well (e.g. intrusive memories). By processing emotional events, sleep alters the emotional tone of the memory, resulting in a modified emotional reactivity at recall, or reexposure. Contrasting hypotheses have been put forward with respect to a strengthening or weakening effect of sleep on

the emotional tone of a memory. On the one hand, it has been proposed that sleep enhances the consolidation of both the content and the emotional tone of an emotional memory [6]. The suggested underlying neural mechanism is the particularly high brain activity in regions that have been linked to emotional processing during rapid eye movement (REM) sleep [7], thereby promoting the reactivation of emotional memory traces [8, 9]. An alternative hypothesis that has its roots in animal research is the “sleep to weaken” hypothesis, also known as the “*sleep to forget and sleep to remember hypothesis*” [10], which states that during sleep only the consolidation of the content of the emotional memory is enhanced, while the emotional tone is actually weakened. The strengthening and weakening of different memory components suggest that during sleep memory traces are not only reactivated but also reorganized. This latter hypothesis proposes that especially during REM sleep, the activity of brain regions related to emotional processing reinforces the reactivation of memory traces of emotional experiences, however, in the absence of their associated adrenergic tone. During REM sleep, the activity of the locus coeruleus (LC) is inhibited, resulting in a decrease in noradrenaline release, which is also coined as silencing of the LC [11–13]. Hence, the reactivation of memory traces under decreased noradrenaline release is supposed to strengthen the consolidation of the content, while the emotional tone of the memory is de-potentiated or weakened.

Over the last years, several studies have been conducted in which the weakening/strengthening effect of sleep has been tested in mostly healthy and sometimes symptomatic participants (e.g. depression; for reviews see [14, 15]). In these studies, changes in the emotional tone were typically assessed by asking participants to rate their emotional reactivity to an emotional stimulus (e.g. a negatively valenced picture) before and after a period of either sleep or wakefulness. The produced findings are contradictory with either a decrease, maintenance, or increase of emotional reactivity after a period of sleep. This inconsistency might be explained by the use of various experimental designs, different assessments of emotional reactivity (e.g. subjective and objective), as well as distinct methodological approaches for data analyses [14]. For instance, studies were often conducted [16–24] (or statistically analyzed [25–32]) without the inclusion of control stimuli—that is, emotional stimuli that have not been encoded during the initial exposure (i.e. new emotional stimuli at reexposure). Without proper control stimuli, a general effect of sleep on affect cannot be distinguished from a memory-specific effect. Hence, results may well reflect the effect of sleep on emotional reactivity rather than on emotional memory [33, 34]. Another varying factor in experimental paradigms concerns the manipulation of sleep, with many studies using partial sleep deprivation. In line with the previously mentioned bidirectional link between sleep and affect, a night of interrupted sleep relates to lower mood the following day, which can negatively influence general affect (e.g. [34]). A change in general affect after participants have been deprived of sleep makes it difficult to assess a memory-specific change. Support for the hypothesized strengthening effect of sleep seems to be mostly driven by studies that have partially deprived their participants of sleep (but see [9, 20, 21]). Even though sleep deprivation studies also bear valuable information that, so far, cannot otherwise be assessed in humans, they also have their drawbacks such as a general increase in stress and low mood (e.g. [35]).

Support for the “sleep to weaken” hypothesis has been found in a recent study in which Wassing et al. [23] investigated the emotional reactivity to a re-exposed emotional stimulus in healthy sleepers compared to people diagnosed with insomnia disorder. A hallmark of insomnia is hyperarousal which is suggested to interfere with the silencing of the LC during REM sleep [36, 37]. In this online study, participants sang along to a song in a karaoke-like setup while their voice was recorded. As of 2 days later, participants were repeatedly exposed (four times in total) to their own singing. Over a period of three consecutive days, exposures were separated by either day-time wakefulness, nighttime sleep, or both. After each exposure, their emotional reactivity was measured using a shame scale (Experiential Shame Scale (ESS) [38]), which assesses the experienced levels of shame on three subscales—physical, emotional, and social. Results showed that immediate sleep after the previous emotional experience led to reduced ratings on the physical and emotional subscales of the ESS in healthy participants when re-exposed to their recorded singing. Interestingly, the opposite was the case in the insomnia group, where participants showed an increase in physical ratings of shame after immediate sleep. In line with the “sleep to weaken” hypothesis, these findings indicate that healthy immediate sleep promotes an overnight adaptation to a distressing event by reducing emotional reactivity, whereas disrupted sleep in people with insomnia interferes with synaptic reorganization and with that hinders this overnight adaptation. In a later study, Wassing et al. [39] substantiated their original findings in the lab, where they indeed demonstrated that the exposure of participants’ own out-of-tune singing activated objective indicators of emotional reactivity (Galvanic skin response (GSR) and amygdala activation). Subsequently, they found a correlation between restful sleep and amygdala activation [22], but it bears mentioning that a wake control group was missing. Hence, we here examined previously unaddressed questions relevant to the use of the karaoke paradigm and the interpretation of findings. First, previous work used electrodermal activity (GSR) as an autonomic nervous system indicator of shame or embarrassment induced by the karaoke paradigm. However, electrodermal activity merely indicates that stimuli are physiologically arousing, not that they are specifically inducing the emotion of embarrassment. In this study, we, therefore, recorded facial blushing, an autonomic indicator of embarrassment [40]. Second, previous studies had limitations in the control stimuli used. More precisely, reexposure to embarrassing stimuli after a period of wake or sleep has not been optimally matched by similarly embarrassing stimuli that were newly exposed.

To investigate the effect of sleep on emotional memory, we first created an emotional autobiographical memory in line with the study design of Wassing et al. [23]. The embarrassing episode was induced by a karaoke-like setup during which participants first had to sing along to two songs while we recorded their voices. Subsequently, they were exposed to a fragment of one of their recordings (rec1) in presence of an audience to create an embarrassing memory. Twelve hours later, participants were *re-exposed* to their recording (rec1; i.e. heard before) as well as *newly exposed* to a fragment of the other recording (rec2; i.e. not heard before). The 12-hr period in between the two sessions was either filled with day-time wakefulness (wake group) or nighttime sleep (sleep group). With the use of an infrared-reflective photoplethysmograph, we recorded facial blushing and collected subjective ratings of embarrassment and

valence. Facial blushing represented our main outcome measure of emotional reactivity as subjective measures demand the ability to self-reflect on one's emotions [41]. Hence, we tested the emotional reactivity (i.e. physiological and subjective) triggered by reactivation (rec1) and the emotional reactivity evoked by a newly exposed embarrassing stimulus (rec2) after either a period of wakefulness or sleep. If sleep indeed reduces the emotional tone of an embarrassing memory, we expected to observe decreased facial blushing in response to the re-exposed recording (rec1) compared to the newly exposed recording (rec2) after a night of sleep, but not following 12 hr of wakefulness.

## Methods

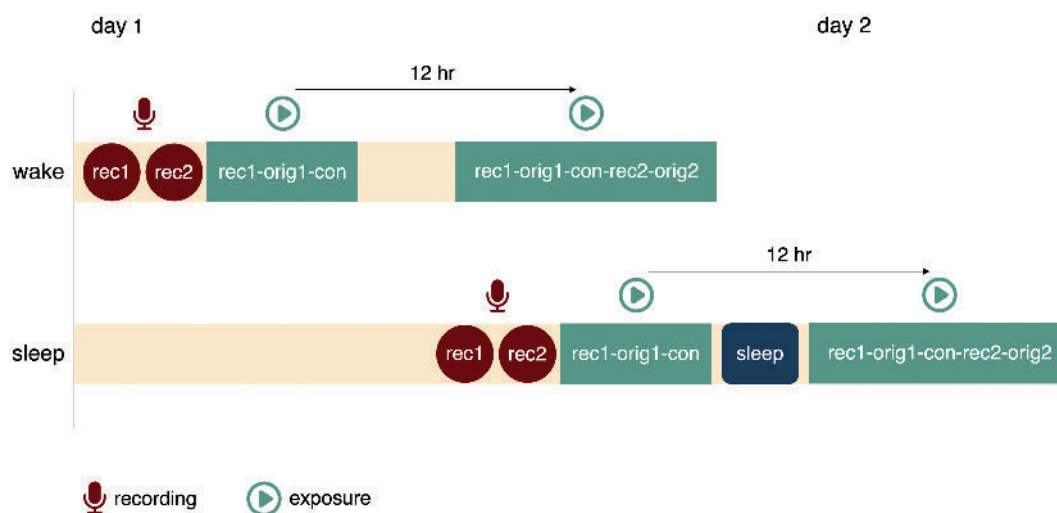
### Participants

As preregistered (<https://osf.io/3am65/>), the final sample size was determined by employing sequential Bayes Factor hypothesis testing [42]. Evaluated by an uninvolved researcher, the Bayes Factor (BF) met the previously defined grade of evidence ( $BF_{10} \leq \frac{1}{3}$  or  $BF_{10} \geq 3$ ) after the minimum sample size of  $N = 20$  participants per group ( $n = 40$ ) was reached. We collected data from  $n = 45$  participants of which five were excluded:  $N = 1$  quit after the first session due to illness,  $N = 1$  was excluded due to technical issues, and  $N = 3$  were excluded due to counter-balancing mismatches between groups caused by experimenter errors. The mismatch concerned the order in which participants' own recordings were presented during the second session (i.e. whether they were first re-exposed to rec1 or first exposed to rec2). We based the exclusion of these participants only on their (1) presentation order and (2) age, sex, and scores on the Blushing Propensity Scale (BPS) [43] and Liebowitz Social Anxiety Scale (LSAS) [44]. As the two groups were matched on these demographical data, the latter preserved the similarities between the two groups. Our outcome measures (objective and subjective) were not considered. Hence, our sample for statistical analyses consisted of  $n = 40$  participants (age range: 18–59 years). Participants were recruited via the Lab website of the University of Amsterdam (UvA) and had to fit the following

inclusion criteria: age between 18 and 70 years, being familiar with the two songs they had to sing along to, and availability during the morning and evening sessions. Exclusion criteria were any current or past diagnosed sleep disorder, social anxiety disorder, or epilepsy, use of any medication that might influence facial blood flow (e.g. antihypertensives), indifference or a positive attitude towards performing karaoke, and having extensive experience with karaoke (>5 times). The criteria for karaoke attitude and experience were chosen to enhance the chance of inducing a feeling of embarrassment during the experiment. After passing the screening procedure through online questionnaires (exclusion rate: 85%), participants were assigned to one of the two groups using stratified randomization carried out by an uninvolved researcher. The groups were matched on scores on the BPS and LSAS [44], age, and sex (see [Supplementary Table S1](#) for demographics and statistical tests). To avoid a biased enrollment of people who actually enjoy singing, participants were unaware of what to expect during the experimental task when they filled in the online screener. At invitation, all participants were informed about the exact study procedure. The study was approved by the local ethics review board of the UvA including an informed consent procedure. Participation was compensated with either research credits or € 20 (during the first academic year dealing with the Coronavirus Disease 2019 pandemic, participants received additional € 10).

### Procedure

The experiment was conducted over two sessions taking place 12 hr apart. During these 12 hr, one group stayed awake during the day (wake group), while the other group slept during the night (sleep group). Hence, the wake group had their first session in the morning (between 08:00 and 10:00) and the second session in the evening (between 20:00 and 22:00) on the same day. In contrast, the sleep group attended the first session in the evening (between 20:00 and 22:00) and the second session in the morning on the following day (between 08:00 and 10:00). For a detailed overview of the design, we refer to [Figure 1](#). One day before and throughout the experiment, we asked participants to



**Figure 1.** Design overview. We tested two groups (wake and sleep), both attending two sessions. In the first session, after singing along to two songs (recording), participants were exposed to a fragment of one of their own recordings (rec1), of the original song of that recording (orig1), and of the control song (con). During the second session, participants were re-exposed to these fragments as well as newly exposed to a fragment of their other recording (rec2) and of the original song of that recording (orig2). The sessions were 12 hr apart during which one group stayed awake during the day (wake group), while the other group slept during the night (sleep group).

refrain from alcohol and other drugs. During the experimental day(s), they were additionally not allowed to take any naps, and caffeinated drinks should not be consumed after noon. The data were collected by two experimenters, one guiding the participants through the procedure and the other one acting as a listener during the exposures. The role of the experimenters was consistent throughout both sessions.

#### Session 1.

**RECORDINGS.** In the first session, participants were seated at a desk with a screen, speakers, headphones, and a microphone in front of them (room temperature approximately 21°C). After signing the consent form and filling in the same questionnaires as in the online screener (demographics, inclusion/exclusion criteria checks, BPS, and LSAS), we asked the participants to sing along to two widely known and rather challenging songs (“Hey Jude” by The Beatles and “Dancing Queen” by ABBA) while their voice was recorded. The lyrics were handed out on a sheet, and the experimenter instructed participants to sing loud and at their best performance before leaving the room. To promote out-of-tune singing, the original songs were played via headphones at a high volume (average peak: 80 dB). Thereby, participants could hardly hear their own voices which made pitch correction difficult. The order of songs was counterbalanced such that half of the participants in both groups first sang the song “Hey Jude.” Before the start of each recording, participants were informed about which song would be played. After finishing the recordings, participants stayed in a different room where they completed a distraction task (sudoku) for the next 30 min. The purpose of the distraction task in another context was to separate the singing experience from the subsequent exposure experience. In the meantime, the experimenter selected a 40-s-long fragment of each recording that contained the most out-of-tune singing. To note, the selected fragments, therefore, contained different parts of the songs across participants.

**FIRST EXPOSURE.** After 30 min, participants were brought back to the first room to experimentally induce the embarrassing episode. The exposure task was preceded by the attachment of a photoplethysmography sensor to physiologically record facial blushing. Afterward, the exposure task commenced with the first of three test blocks. Each block started with a 1-min baseline measure of facial blood flow. During the following 40 s, we exposed participants to one song fragment—this time played through speakers—which either contained the embarrassment-inducing recording (rec1) [22, 23, 39, 45], the original version of that recording (orig1), or a new song (con). Fragments of the original and new songs were added as potential neutral stimuli (i.e. songs that do not induce embarrassment). As it was unclear whether the original songs would also induce an emotional reactivity by reminding the participants of their own recordings, we added a song that participants did not sing along to as an additional control stimulus (“Respect” by Aretha Franklin). Hence, the exposed songs were the selected fragment of one of their own recordings (rec1; 50% “Hey Jude” in both groups), the same fragment of the original song of that recording (orig1), and a fragment of a new song (con). For every participant, the new song was played from 1’44”. In the 2 min after each exposure, participants rated their embarrassment and valence evoked by the exposed song fragment on two separate visual analog scales (VASs), one presented at a time in random order. The

VASs ranged from not embarrassed at all/extremely negative (0) to extremely embarrassed/extremely positive (100). To avoid carry-over effects on subjective and objective measures from one block to the next, participants performed a simple distraction task (1-back task) in between exposures. All participants were exposed to all three song fragments once in a randomized order. The time between exposure and the next baseline measure lasted 2 min in total, including subjective ratings followed by the distraction task. There was no time limit for participants to respond to the subjective ratings. Hence, there was a 3-min period before participants were exposed to the next fragment, which is deemed sufficient for blushing responses to return to baseline (see e.g. [46]). This was further supported by exploratory analyses that can be found in [Supplementary Figure S3](#). Throughout the task, a second experimenter sat at a table located diagonally left behind the participant. We instructed the participants to not communicate with the listener, but to focus on the screen and complete the task. With the additional observer, we intended to increase the feeling of embarrassment during exposure to their own singing.

#### Session 2.

**SECOND EXPOSURE.** Twelve hours later, participants returned to the lab for the second session. During the first 10 min, we asked participants again to complete a distraction task (sudoku). The purpose of this 10-min filler task was to acclimatize the participant to room temperature to reduce external influences on facial blood flow. After the attachment of a photoplethysmography sensor, we exposed participants again to song fragments as in the first session. However, this time the exposure task contained five test blocks instead of three: participants were re-exposed to the same fragments of rec1, orig1, and con, as well as newly exposed to fragments of their other recording (rec2) and its original version (orig2). The original version of the second recording was added to the design as an additional newly exposed control stimulus. The order in which the different song fragments were played was again randomized, where one recording would never directly follow the other. Whether they were first exposed to rec1 or rec2 was again counterbalanced across both groups.

#### Facial blushing

Our main outcome measure was a physiological measure of embarrassment, facial blushing. We measured facial blushing using an infrared-reflective photoplethysmograph that was both AC and DC coupled. Hence, the transducer recorded the increase of blood volume pulse (AC) as well as blood volume (DC) over time at a sampling speed of 200 Hz. The sensor was placed on the participants’ forehead, 2–3 cm from the center to their right, and subsequently covered with a black headband to reduce the signal interference of external light (e.g. [46]). For our analysis, we focused on the amplified and filtered fast-responding AC component (low-pass 4.0 Hz, 4th order, 24 dB/oct; high-pass 0.5 Hz, 6th order; 36 dB/oct). To further minimize facial movement artifacts in the signal, sharp spikes that were asynchronous with the pulse waveform have been manually removed. We used the program Versatile Stimulus Response Registration Program—1998 (VSRRP98 v 11.04) [47] to record the blood volume pulse amplitudes. The pulse amplitudes were then standardized due to individual differences in skin characteristics [48] as follows. From the average blood volume pulse amplitude evoked by

each song fragment, the average blood volume pulse amplitude during the 1-min pre-fragment baseline was subtracted. The resulting baseline-corrected signal was divided by the average baseline amplitude and multiplied by 100 to calculate the percentage change from baseline to song fragment. The resulting percentage change was then used for further statistical analyses and is referred to as *blushing response*.

### Sleep measures

As an index of sleep duration and quality, participants filled in the online Consensus Sleep Diary (CSD) [49] every morning during the experimental day(s). For each night, we calculated the individual subjective total sleep time (sTST) by subtracting the sleep onset latency (SOL) and total duration of wake after sleep onset (WASO) from the period between lights out and final get-up time. Subjective sleep quality (sSQ) was measured on a 5-point Likert scale ranging from “poor” (0) to “very good” (4). In addition to subjective sleep measures, we also collected physiological activity with an actigraph (ActTrust2 Condor Instruments Model AT20101) that participants wore during the 12 hr between the two sessions. We employed this microelectromechanical system accelerometer to record their physical (in)activity at a sampling rate of 25 Hz. The data were uploaded to and converted by the computer software ActStudio (Condor Instruments Ltda., SP, Brazil). Based on the actigraphic data, we computed objective total sleep time (oTST). Participants in the sleep group were instructed to press the actigraph’s button once at the moment they intended to sleep (lights out), thereby adding a marker to the recording. We used get-up times reported in the CSD entries for the actigraphy analysis. Actigraphic data were scored according to standards using the pyActigraphy package [50] to estimate the oTST and objective sleep quality (oSQ; i.e. sleep efficiency: % time asleep after sleep onset).

### Statistical analysis

Bayesian statistical analyses were carried out in JASP [51] and RStudio (v 1.4.1717) [52] as preregistered (<https://osf.io/3am65/>). Assumptions were checked using frequentist analyses (significant at  $p < .05$ ). We corrected for violations of normality using nonparametric alternatives for t-tests. Before we performed our main analysis, we tested whether the induction of the emotion embarrassment was successful during the first session. For this, we used a  $2 \times 3$  Bayesian repeated measures ANOVA with the between-subject factor Group (sleep vs. wake) and the within-subject factor Song (rec1, orig1, and con), followed by post-hoc comparisons. The successful emotion induction was indicated by a higher blushing when exposed to rec1 compared to both orig1 and con. To test the equivalence of control stimuli we compared the blushing response to orig1 and con in a post-hoc comparison. Default priors were applied (i.e.  $r = 0.5$  for fixed effects), and we compared the alternative hypotheses to the null hypothesis (no effect). We considered Bayes Factors (BFs) between 0 and  $\frac{1}{3}$  as evidence for the null hypothesis, whereas BFs between 3 and 10 were considered as moderate evidence and BFs greater than 10 as strong evidence for the alternative hypotheses [53]. In the main analysis, we evaluated whether

sleep had a mitigating effect on the blushing response to the recording participants had been exposed to before (rec1), but not to the recording they were exposed to for the first time (rec2). To avoid discarding data of participants with one or more missing or removed data points and controlling for order effects (see Session 2. and [Supplementary Figure S1](#)), we deviated from the preregistered analysis and employed Bayesian linear mixed-effect models which allows for the inclusion of all available observations. We performed a Bayesian linear mixed-effect model using the rstanarm package with default prior settings (v. 2.21.3 [54]) to predict the blushing response during the second session. We included Group (wake vs. sleep; reference level “wake”) and Exposure of the recordings (re-exposed vs. newly exposed; reference level “newly exposed”), and their interaction as fixed effects. The intercept varied by participant. In an additional analysis, we tested the change in blushing response to rec1 from the first session to the second session in a Bayesian linear mixed-effect model. We included Group (wake vs. sleep; reference level “wake”) and Session (s1 vs. s2; reference level “s1”), and their interaction as fixed effects and allowed the intercept to vary between participants. Expected values (median) and their 95% credible intervals (CI) indicated by the highest density intervals (HDI [ $HDI_{low}$ ,  $HDI_{high}$ ]) are reported. Evidence for a difference ( $\delta$ ) was interpreted as convincing if the 95% CI did not include zero and the posterior probability [ $P(\delta > 0)$ ] for a difference was close to one. The same analyses were applied to the subjective ratings of emotional reactivity. For additional control and exploratory analyses on the embarrassment-evoking stimuli and the relationship between different outcome measures, we refer to the [Supplementary Analyses](#).

## Results

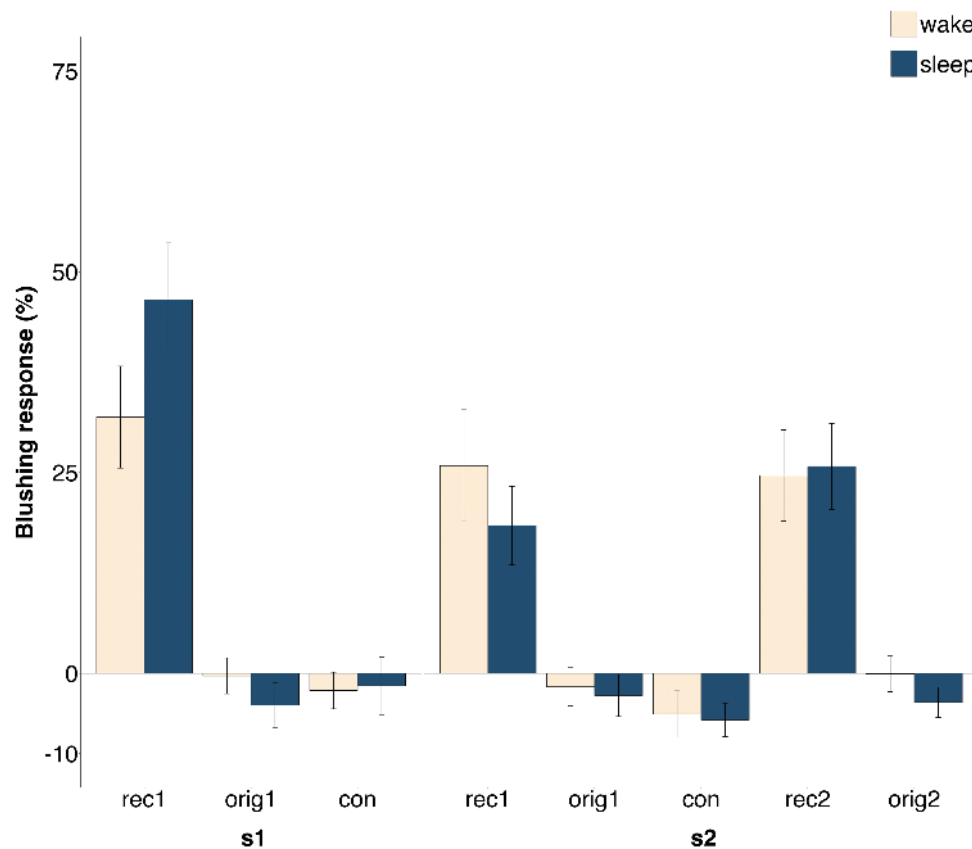
Groups were matched on age, sex, and BPS and LSAS questionnaire scores (see [Supplementary Table S1](#)).

### Facial blushing

**Session 1.** One data point was missing due to experimenter error, and two data points were removed as outliers ( $\pm 3$  SDs from the overall mean of the song). This resulted in 18 data points for rec1 in the sleep group and 19 data points for orig1 in the wake group.

**MANIPULATION CHECK.** The induction of the emotion of embarrassment during the first session was equally successful in both groups. That is, we found strong evidence for a main effect of Song and no or inconclusive evidence for a main effect of Group and the interaction term (Song:  $BF_{incl} = 2.4e+14$ ; Group:  $BF_{incl} = 0.41$ ; Song\* Group:  $BF_{incl} = 0.90$ ). More specifically, we observed a strong increase in blushing response across groups when exposed to rec1 compared to orig1 and con (rec1 vs. orig1:  $BF_{10,U} = 2.0e+7$ ; rec1 vs. con:  $BF_{10,U} = 575252.61$ , see [Figure 2](#)).

**CONTROL STIMULUS CHECK.** There was no evidence that the original song evoked more embarrassment compared to the control song. Both songs elicited a similar blushing response (orig1 vs. con:  $BF_{10,U} = 0.18$ , see [Figure 2](#)). Therefore, the original songs (orig1 and orig2) were included as control stimuli in the main analysis.



**Figure 2.** Mean blushing response to the different songs per group (wake and sleep) in both sessions (s1 and s2). The songs were fragments of the recordings (rec1 and rec2), the original songs (orig1 and orig2), and the control song (con). After controlling for an order effect in stimulus presentation, main analyses showed a reduction in blushing response to the reactivated recording (rec1) from session 1 to session 2 in the sleep group only (Group\* Session:  $\beta = -25.51\%$ , HDI [-48.56, -0.40],  $P(\delta > 0) = 0.98$ ). However, this effect was not specific to the previously exposed recording (rec1): emotional reactivity to the reactivated recording (rec1) and the new recording (rec2) during session 2 did not differ after sleep and wakefulness (Group\* Exposure:  $\beta = -11.79\%$ , HDI [-31.94, 7.69]). \* represents convincing evidence for a difference between groups while n.e. represents no evidence for a difference between groups. Error bars represent the standard error of the mean.

Session 2. One data point was missing due to experimenter error, resulting in 19 data points for rec1 in the wake group. No outliers could be identified.

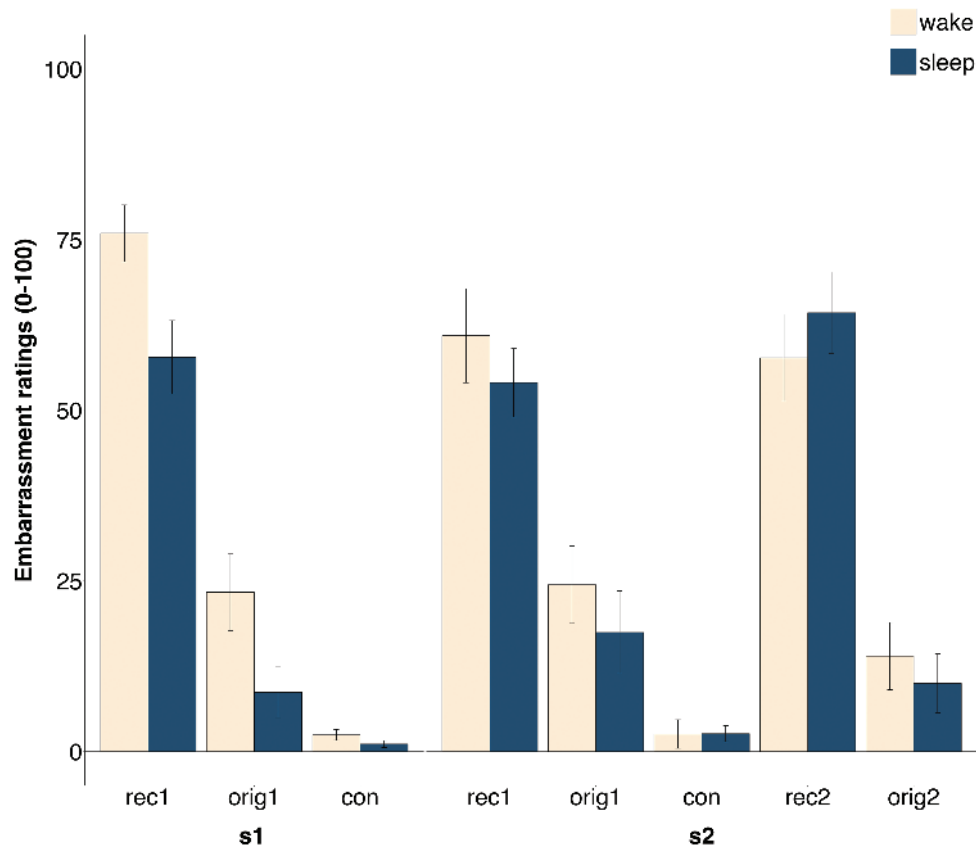
**CONTROL STIMULUS CHECK.** Before we performed our main analysis, we first ensured that groups did not differ with respect to evoked embarrassment by the control stimuli (orig1 and orig2) during the second session. For this, we ran a  $2 \times 2$  Bayesian repeated measures ANOVA with the between-subject factor Group (sleep vs. wake) and within-subject factor Exposure (re-exposed vs. newly exposed) for the original songs (orig1 and orig2). We found no evidence for a main effect of Group ( $BF_{incl} = 0.30$ ) or Exposure ( $BF_{incl} = 0.17$ ), nor for an interaction (Group\* Exposure:  $BF_{incl} = 0.08$ ), indicating that overall the blushing response was neither different between groups nor songs (see Figure 2).

**MEMORY-SPECIFIC EFFECT OF SLEEP ON FACIAL BLUSHING.** To examine whether sleep mitigated the emotional reactivity to an embarrassing memory, we compared groups with respect to the blushing response to the re-exposed rec1 with the newly exposed rec2 in the second session (see Figure 2). As the order in which rec1 and rec2 were presented affected the outcome variable (see Supplementary Figure S1), we added Order as a covariate to our model. Against our expectations, we found no evidence for a difference between groups in blushing response when comparing

rec1 with rec2 (Group\* Exposure:  $\beta = -11.79\%$ , HDI [-31.94, 7.69]), nor was there evidence for a main effect of Group or Exposure on the blushing response (Group:  $\beta = 1.95\%$ , HDI [-14.62, 17.62]; Exposure:  $\beta = 3.05\%$ , HDI [-10.23, 17.20]). These findings suggest that, compared to day-time wakefulness, a night of sleep did not differentially affect the level of embarrassment evoked by the reactivated recording (rec1) and the recording participants had been exposed to for the first time (rec2). As groups seemed to differ in blushing response at initial exposure to rec1 in the first session (albeit not statistically significant), we included the blushing response to rec1 in the first session in the main analysis. Adding the blushing response to rec1 in the first session as a covariate to the analysis did however not change the result (HDI includes zero).

#### Additional exploratory analyses.

**GENERAL EFFECT OF SLEEP ON FACIAL BLUSHING.** To test whether a night of sleep compared to wakefulness reduced emotional reactivity more generally and not memory specific, we compared the blushing response to rec1 between sessions 1 and 2. As rec1 was the only recording presented during the initial exposure (i.e. session 1), we have no control stimulus here and cannot disentangle any effect of sleep related to changes in memory representation from effects of sleep on emotion regulation. Hence, we refer here to a general instead of a memory-specific effect.



**Figure 3.** Mean embarrassment ratings per group (wake and sleep) after being exposed to one of five songs in both sessions (s1 and s2). The songs were fragments of the recordings (rec1 and rec2), the original songs (orig1 and orig2), and the control song (con). Ratings ranged from 0 (not embarrassed at all) to 100 (extremely embarrassed). In contrast to the blushing response, main analyses on embarrassment ratings showed that the decrease for rec1 from the first to the second session did not differ between groups (Session\* Group:  $\beta = 12.36$ , HDI [0.11, 26.11],  $P(\delta > 0) = 0.03$ ). Furthermore, there was no indication of a memory-specific effect of sleep on embarrassment ratings when comparing the reactivated recording (rec1) to the new recording (rec2) during session 2 (Group\* Exposure:  $\beta = -12.88$ , HDI [-31.05, 6.55]). n.e. represents no evidence for a difference between groups. Error bars represent the standard error of the mean.

Again, the order of rec1 presentation (i.e. before/after rec2) was included as a covariate in the analysis. Results showed convincing evidence for a pronounced reduction in blushing response from session 1 to session 2 after a night of sleep (Group\* Session:  $\beta = -25.51\%$ , HDI [-48.56, -0.40],  $P(\delta > 0) = 0.98$ ). Group and Session alone did not affect the blushing response (Group:  $\beta = 15.86\%$ , HDI [-0.86, 34.30]; Session:  $\beta = 0.45\%$ , HDI [-18.81, 19.32]). These results indicate that sleep indeed reduced the blushing response from the first session to the second session more than a similar interval of wakefulness did (see Figure 2).

### Embarrassment ratings

The subjective ratings of the embarrassment of one participant were excluded from the analysis due to response invariance throughout songs (sleep group:  $N = 19$ ). We interpreted the response invariance as an indication of disengagement with the task. Hence, the participant's subjective ratings of embarrassment would not represent their true emotional reactivity as indicated by the blushing response. For analyses of valence ratings and potential order effects we refer to Supplementary Figures S1 and S2.

**Session 1.** Two data points were removed as outliers ( $\pm 3$  SDs from the overall mean of the song). This resulted in 17 data points for the control song in the sleep group.

**MANIPULATION CHECK.** The induction of the emotion of embarrassment as assessed by subjective rating during the first session was successful in both groups. Accordingly, we found strong evidence for the main effect of Song ( $BF_{\text{incl}} = \infty$ ). In detail, we observed a strong increase in embarrassment ratings across groups after being exposed to the recording compared to the original song and the control song (rec1 vs. orig1:  $BF_{10,U} = 1.22e+11$ ; rec1 vs. con:  $BF_{10,U} = 1.83e+17$ , see Figure 3). The data further suggested that groups differed in the strength of embarrassment. We found moderate evidence for a main effect of Group ( $BF_{\text{incl}} = 4.99$ ), but inconclusive evidence for an interaction (Group\* Song:  $BF_{\text{incl}} = 2.56$ ). To compare the two groups on the three songs using nonparametric Bayesian Mann-Whitney U tests (two-tailed), we adjusted the prior to control for multiple testing according to the guidelines by Westfall et al. [55] (Cauchy prior = 0.260). Results indicate moderate evidence for differences between the two groups in embarrassment ratings for the recording only (rec1:  $BF_{10} = 4.58$ ; orig1:  $BF_{10} = 1.17$ ; con:  $BF_{10} = 0.75$ ).

**CONTROL STIMULUS CHECK.** We found strong evidence for an overall difference in embarrassment evoked by the original song compared to the control song. Participants rated the original song as more embarrassing than the control song (orig1 vs. con:  $BF_{10,U} = 119.52$ , see Figure 3). Hence, the original songs cannot be



considered neutral, and the control song was included as a control stimulus in subsequent analyses.

**Session 2.** One data point was removed as an outlier ( $\pm 3$  SDs from the overall mean of the song). This resulted in 19 data points for the control song in the wake group.

**CONTROL STIMULUS CHECK.** Before performing further analyses without the control stimulus (con), we first ensured that groups did not differ with respect to evoked embarrassment by con during the second session. Hence, we ran a nonparametric Bayesian Mann–Whitney  $U$  test (two-tailed; Cauchy prior = 0.707) and found no evidence for a difference between the groups on embarrassment ratings for con,  $BF_{10} = 0.43$ . This means that no correction for group differences in embarrassment ratings for rec1 was necessary.

**MEMORY-SPECIFIC EFFECT OF SLEEP ON EMBARRASSMENT RATINGS.** To examine whether sleep mitigated the emotional reactivity to an embarrassing memory, we compared embarrassment ratings to the re-exposed rec1 with the newly exposed rec2 in the second session between the sleep and the wake group (see Figure 3), while controlling for the order of stimulus presentation. In line with the physiological data, embarrassment ratings showed no evidence for a main effect of Exposure ( $\beta = 2.88$ , HDI [-10.52, 15.77]), a main effect of Group ( $\beta = 6.92$ , HDI [-11.38, 23.65]), nor for a difference between groups when comparing rec1 with rec2 (Group\*Exposure:  $\beta = -12.88$ , HDI [-31.05, 6.55]). Again, these findings suggest that, compared to day-time wakefulness, a night of sleep did not differentially affect the level of embarrassment evoked by the reactivated recording (rec1) and the newly exposed recording (rec2). In an additional analysis, we controlled for the previously observed group difference in embarrassment ratings for rec1 in the first session (see manipulation check). Again, adding embarrassment ratings for rec1 in the first session as a covariate did not change the result (HDI includes zero).

*Additional exploratory analyses.*

**GENERAL EFFECT OF SLEEP ON EMBARRASSMENT RATINGS.** To test the general effect of sleep on embarrassment ratings, we compared the change in ratings for rec1 from the first to the second session. We found overall lower embarrassment ratings in the sleep group (Group:  $\beta = -18.68$ , HDI [-34.14, -3.45],  $P(\delta > 0) = 0.99$ ) and strong evidence for a decrease in embarrassment ratings over sessions (Session:  $\beta = -18.16$ , HDI [-29.51, -8.04],  $P(\delta > 0) = 1.00$ ). However, there was no evidence for an interaction between Session and Group ( $\beta = 12.36$ , HDI [0.11, 26.11],  $P(\delta > 0) = 0.03$ ). Thus, in contrast to the blushing response, the effects of sleep and wake on embarrassment ratings did not differ (see Figure 3).

### Sleep analyses

Subjective and objective sleep measures were taken as a check of our experimental sleep manipulation (wake vs. sleep) during the 12-hr period between the two sessions. Sleep data of one participant was missing, resulting in a sample size of  $N = 19$  participants in the sleep group. Means and SDs can be found in Supplementary Table S1. Comparing the actigraphy data with subjective sleep variables in the sleep group showed convincing evidence that subjective total sleep time (sTST) correlated

positively with objective total sleep time (oTST) and objective sleep quality (oSQ) in the night after the first session (sTST\*oTST:  $r = 0.85$ ,  $BF_{10} = 4583.28$ ; sTST\*oSQ:  $r = 0.61$ ,  $BF_{10} = 10.56$ ). This result supports the reliability of subjective measures in the sleep diary. Evidence for a correlation between subjective ratings of sleep quality (sSQ) and objective measures was inconclusive (sSQ\*oTST: Kendall's  $t = 0.35$ ,  $BF^{10} = 2.32$ ; sSQ\*oSQ: Kendall's  $t = 0.35$ ,  $BF^{10} = 2.26$ ). When assessing participant adherence to instructions, we found that one participant reported having taken sleeping pills (over-the-counter available supplement GABA), and four participants had caffeinated drinks after noon on the day of the first session. Excluding these participants from the main analyses did not affect the results.

To evaluate whether group differences in emotional reactivity during the first exposure might be caused by a potential effect of sleep on the next day's mood, we compared the two groups on quantity and quality of sleep the night before the first session. With a Bayesian independent samples  $t$ -test (two-tailed, Cauchy prior = 0.707), we tested the group difference in sTST. Results showed no evidence for a difference in sTST between groups ( $BF_{10} = 0.50$ ; for means and SDs see Supplementary Table S1). A Bayesian Wilcoxon rank sum test (two-tailed) showed no evidence for a difference between groups in subjective sleep quality (sSQ) ratings either ( $BF_{10} = 0.44$ ; for means and SDs see Supplementary Table S1). These findings indicate that the group differences in emotional reactivity during the first exposure cannot be explained by variations in sleep quantity and quality. To test whether the embarrassment induction during the first session led to poorer sleep and potentially influenced next-day emotional reactivity in the sleep group, we compared sTST and sSQ ratings between the night before and after session 1. We found convincing evidence that participants slept considerably less after the first session compared to the night before ( $BF_{10} = 117.52$ ; Bayesian paired-samples  $t$ -test, two-tailed, Cauchy prior = 0.707; for means and SDs see Supplementary Table S1). However, evidence for a direct correlation between sTST and emotional reactivity measures was not found (all  $BF_{10} < 1.83$ ). In contrast to sleep quantity, there was only inconclusive evidence that subjective sleep quality (sSQ ratings) differed between the two nights ( $BF_{10} = 1.13$ ; for means and SDs see Supplementary Table S1). Hence, while total sleep time (sTST) was reduced, ratings of sleep quality (sSQ) did not change from before to after initial embarrassment induction.

## Discussion

In this study, we aimed to address the dispute on the role of sleep in processing the emotional tone of an autobiographical memory. The emotional episode was induced by exposing participants to their flawed singing of one of two songs to create an embarrassing memory. Twelve hours later, participants were re-exposed to the same recording as well as newly exposed to their imperfect singing of the other song. The 12-hr period in between the two sessions was either filled with day-time wakefulness or nighttime sleep. Findings convincingly showed that the induction of an embarrassing autobiographical episode was successful in this paradigm, as is evidenced by physiological and subjective readouts. While the decrease in blushing response was more pronounced across a period of sleep than across a period of wakefulness, the smaller response was not unique to the re-exposed embarrassing stimulus, but also occurred for

the newly exposed embarrassing stimulus. Thus, in contrast to our prediction, we found no evidence for a stronger decrease in emotional reactivity to the previously encoded recording than to the new recording after a night of sleep compared to wakefulness for both physiological and subjective measures. Neither did we find evidence for the hypothesized strengthening effect of sleep. Exploratory analyses on the general change in emotional reactivity, however, showed a reduction in facial blushing after a period of sleep compared to wakefulness.

Our findings do not corroborate the stronger reduction in subjective ratings of physical and emotional distress after a night of immediate sleep that has been previously shown in healthy subjects by Wassing et al. [23]. Even though our study design was inspired by Wassing et al., the inconsistency in results could still be due to some methodological differences. Since they investigated group differences in emotional reactivity between healthy sleepers and people with insomnia disorder, we can only compare our results to the findings of the healthy sleep group. For both groups, Wassing and colleagues employed a within-subject design, where participants have been repeatedly exposed to the same stimulus over several days, separated by either day-time wakefulness, nighttime sleep, or both. Despite the counterbalanced schedule across participants to better control for time-of-the-day effects, multiple nights of sleep and several exposures to their singing could have influenced emotional reactivity differently. A second methodological discrepancy concerns the duration of exposure. In our study, participants have been exposed to 40-s fragments of their singing, whereas in Wassing et al.'s study the recordings lasted around 2.5 min. The extended exposure could have promoted habituation to the stimulus. Therefore, emotional reactivity ratings following this long duration might not be representative of the initial emotional response. Finally, we exposed participants to their singing in the presence of an audience to increase the strength of the induced emotion, while in Wassing et al.'s study participants were home alone. Referring to the bidirectional link between sleep and affect, distressing events can worsen sleep quality, while poor sleep may negatively influence next-day mood. In the current study, poorer sleep following initial exposure could have lowered next-day mood, overshadowing the potentiated overnight decrease in subjective emotional reactivity in the sleep group. When evaluating sleep parameters, our data showed that participants indeed slept significantly less after the first session compared to the night before. Although the hours of sleep the night before the second session did not correlate with emotional reactivity the following day, poor sleep may not be determined solely by the total number of hours slept. Qualitative aspects of sleep such as the duration and continuity of certain sleep stages may be as critical as the quantity of sleep (e.g. [56]). Consequently, future studies should consider reexposing participants after several nights of sleep instead of only one night. Of note, it remains unclear whether participants slept less because exposure to their own singing was distressing. The sleep reduction could also be explained by other factors, such as waking up earlier to attend the second session (see [Supplementary Table S1](#)).

The idea that more than one night of sleep is necessary to observe a reduction in emotional reactivity to a previously encoded stimulus has also been substantiated by Bolinger et al. [57]. In their study, participants were exposed to negative and neutral pictures either in the morning, followed by day-time

wakefulness (wake group), or evening, immediately followed by nighttime sleep (sleep group). Ten hours and again 1 week later, participants were re-exposed to half of those pictures and newly exposed to other pictures. Emotional reactivity was measured physiologically and subjectively. Results showed that after 10 hr, the physiologically measured emotional reactivity decreased in both groups when re-exposed to negative pictures. The decrease, however, was stronger after wakefulness, indicating a preservation of the emotional reactivity after immediate sleep. Interestingly, these findings stand in contrast to what has been observed one week later. At 1-week reexposure, Bolinger et al. found a further decrease in emotional reactivity compared to the 10-hr reexposure, when participants slept immediately after the first exposure. In the wake group, emotional reactivity increased again when compared to the 10-hr reexposure. Further, valence was rated as less negative only in the sleep group at 1-week reexposure, while valence ratings in the wake group did not change across all three (re)exposures. Analyses of the objective and subjective emotional reactivity to newly exposed pictures did not reveal any group differences. Thus, Bolinger and colleagues concluded that the mitigating effect of immediate sleep on emotional reactivity to encoded stimuli only emerges in the long term. To summarize, beyond the discussed methodological deviations in previous studies, differences in induced distress level and number of nights slept after the induction of an emotional episode might influence outcome measures of the memory-specific effect of sleep as well. Therefore, our results do not necessarily reject the hypothesis that sleep weakens the emotional tone of an emotional memory, but rather indicate that further research on varying distress levels and nights of sleep may be required to test the "sleep to weaken" hypothesis.

Apart from considering factors that possibly influence the change in emotional reactivity to emotional memory, there are more ecologically valid and clinically relevant ways to test the "sleep to weaken" hypothesis. To investigate the effect of sleep on intrusive thoughts, Wilhelm et al. [24] exposed participants to traumatic film clips, immediately followed by either a day-time nap or wakefulness. The amount of related intrusive thoughts and associated distress during the subsequent week was then tracked with an intrusion diary. Afterward, participants were re-exposed to traumatic film cues to measure changes in emotional reactivity. During both exposures, affective responses were assessed physiologically and subjectively. Results indicated that associated distress of intrusive thoughts was reduced in participants who entered REM sleep during the nap, while the number of intrusive thoughts equally declined in both groups. With regard to affective responses, the wake group only reported a stronger increase in one of the subjective measures (i.e. negative mood) after reexposure, compared to the nap group. However, the effect became marginal when splitting the nap group into REM vs. noREM sleep. These findings suggest that immediate REM sleep after a distressing experience has a reducing effect on associated distress of intrusive thoughts, while number of intrusive thoughts and emotional reactivity to cued memories of the distressing experience do not seem to be affected. Although Wilhelm et al.'s findings on intrusive thoughts are in accordance with the "sleep to weaken" hypothesis, it is unclear why emotional reactivity at reexposure was unaffected. As suggested in the current study, Wilhelm et al. found that strong emotional reactivity during initial exposure disturbed subsequent sleep (i.e. extended sleep latency and reduced REM