Sleep Renormalizes Negative Emotional Generalization

Ella Bar[1,2,3], Max Bringmann[1], Gennadiy Belonosov[2.3], Kristoffer Aberg[1], Reut Weissgross[1], Yuval Nir[2, 4,5,6, *], Rony Paz[1, *]

[1] Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel.
[2] Department of Physiology and Pharmacology, Gray Faculty of Medical and Health Sciences, Tel Aviv University, Israel, Tel Aviv, Israel.
[3] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel.[4] Sagol Brain Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel.
[5] Department of Biomedical Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv, Israel.
[6] The Sieratzki-Sagol Center for Sleep Medicine, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel.
* Equal senior contribution

Abstract

Overgeneralization of negative experiences, in which aversive responses spread to otherwise safe stimuli, often co-occurs with sleep disruption, and both are central features of anxiety and posttraumatic stress disorder (PTSD). Here, we show that sleep shifts emotional generalization away from the negative and toward the positive. Across three experiments combining behavior, fMRI, and sleep electrophysiology, participants learned to associate three faces with positive, negative, or neutral outcomes. Participants were tested using morphed faces that blended the original stimuli in varying proportions. Immediate generalization was assessed post-learning, and delayed generalization was assessed after overnight sleep or daytime wakefulness. Behaviorally, we found that sleep selectively promotes positive generalization, whereas prolonged daytime wakefulness favors generalization of the negative face. In the fMRI scanner, amygdala and limbic activity during outcome processing predicted stronger immediate generalization of the negative face, and subsequent shift from negative to positive face generalization that occurred only after sleep. Overnight sleep spindle activity, extracted from high-density sleep EEG, positively correlated both with positive face generalization and with the shift from negative to positive generalization following sleep. These findings reveal a potential neural mechanism by which sleep attenuates negative bias and enhances positive representations, suggesting a potential method to buffer against maladaptive generalization in anxiety and PTSD.

Introduction

Imagine walking down the street to your favorite café when a dog suddenly jumps out from around a corner, knocking you to the ground. Upon cautiously returning to the same street, you see a large dog; even if only vaguely resembling the original, you may believe it is the same dog and instinctively retreat. This response reflects emotional generalization*:* an adaptive process that allows individuals to extend past experiences to similar, novel situations[1,2]. Generalization can occur consciously or unconsciously, often influencing how we interpret current stimuli based on past outcomes[2,3]. It is particularly sensitive to emotionally salient events, with a stronger bias toward negative associations. For example, tones paired with negative reinforcement produce broader generalization gradients than those linked to positive reinforcement, which in turn produce broader generalization than neutral stimuli [3,4].

While generalization is considered adaptive in many contexts, overgeneralization, especially of negative experiences, can become maladaptive. For example, a single dog attack may lead to fear of all dogs or even avoidance of outdoor settings altogether. Such behavioral patterns are implicated in psychiatric disorders, including major depression, phobias, generalized anxiety disorder, and posttraumatic stress disorder (PTSD)[1,4,5].

Sleep disturbances are also a hallmark of psychiatric disorders**,** reflecting the bidirectional relationship between sleep and mental health[6–9]. Sleep disruption can impair emotional regulation and increase anxiety, while anxiety itself can interfere with sleep quality [10]. This raises a critical question: *What role does sleep play in shaping emotional generalization?*

Research increasingly shows that sleep facilitates the consolidation of both rewarding[11–15] and aversive[16,17] emotionally salient memories, and supports the extraction of patterns from experience[18–20]. Thus, sleep may help to regulate emotional generalization by limiting the spread of negative associations. Accordingly, we hypothesized that sleep serves a corrective function, reducing negative overgeneralization while enhancing the generalization of positive events.

Participants learned associations between neutral faces and positive, negative, or neutral outcomes. Generalization was assessed using a face discrimination task involving morphs between the original stimuli, to capture how participants extended their learned associations to novel yet similar inputs. Participants completed this task before learning, immediately after, and following a consolidation period of either overnight sleep or daytime wakefulness. We found that

sleep reduced the generalization of negative stimuli and enhanced the generalization of positive stimuli. Subsequent fMRI and sleep electrophysiology experiments revealed potential neural mechanisms subserving these core behavioral findings. Overall, our work aims to clarify how sleep shapes emotional generalization, with implications for understanding and treating anxiety and PTSD[21].

Material and Methods

Experimental tasks

Face stimuli used in the tasks: Twenty-six participants (14 women) rated faces with neutral expression sourced from an online database (the siblings DB)[22], arranging them along a continuum from most malicious to most kind. Three males with similar average ratings as well as similar skin tones, hair, and eye colors were selected as the original faces (Figure 1A). These faces were positioned as vertices of a triangle space comprising 66 faces that were morphed composites of the three original (vertex) faces (FantaMorph software). The proximity of each face to each vertex corresponded to the proportion of the original face in the morph. Thus, the face positioned in the center of the triangle was an average representation derived from all three original faces (Figure 1A). The assignment of original faces to negative, positive, or neutral valence vertices was counterbalanced across participants. Twenty-nine (18 women) participants completed an online discrimination experiment to ensure no bias in perceiving morphed faces as similar to the original faces (see also 'Discrimination Task' below).

Learning task. We used a probabilistic learning task (two-armed bandit[23]) to associate valence with the original face stimuli. After a short training to ensure understanding, participants were instructed to choose one face in each trial and maximize their points according to feedback. During each trial, participants were presented with two out of three original faces simultaneously on the screen, with the left and right positions randomized for each trial. All possible pairs of faces were presented throughout the task. Each original face was probabilistically assigned (p=0.9) to either positive, negative, or neutral feedback (gain +10, loss -10, or zero points), and participants learned these associations as the task progressed. After their selection, the chosen face was highlighted by a yellow frame on the screen. Feedback regarding the participant's response (gain +10, loss -10, or zero points) was then provided based on the selected face (see Figure 1B and Supplementary Figure 1A-C). Because face–feedback associations were highly predictable (p=0.9), participants quickly learned the value of each face and consistently chose the more rewarding option. This might lead to disproportionate exposure to positive feedback

and limited exposure to negative feedback. To overcome this potential bias, we included 'helpless' trials in which both presented stimuli were identical faces associated with negative or neutral feedback, ensuring balanced exposure across feedback types. Trial composition included a total of 132 trials comprising 30 positive-negative, 30 positive-neutral, 30 negative-neutral, 18 neutral-neutral, and 24 negative-negative face pairs trial.

In the fMRI experiment, fixation crosses appeared both between face selection and feedback, and during intervals between trials. The duration of the fixation cross was jittered uniformly between 1 and 7 seconds, with a mean of 4 seconds (Supplementary Figure 1B). The task lasted approximately one hour for the fMRI experiment, and 15 minutes for both the behavioral and sleep EEG experiments. In the sleep EEG experiment, participants first viewed two faces; after 1500 ms, arrows appeared beneath the faces, cueing participants to make a choice (Supplementary Figure 1C).

Discrimination task. The discrimination task operationalized generalization as the perceived similarity between images within the face space and each one of the original vertex faces. This task was performed in three sessions at different time points throughout the experiment: (i) before the first session, participants received instructions and practiced the task in a short training session. In the behavioral and fMRI experiments, each trial in the discrimination task begins with a presentation of one morphed face image from within the face space. Then, participants are asked to choose one of the three original faces that resemble the presented morphed face image the most (Figure 1C, Supplemental Figure 1D). Participants indicated their choice by pressing one of three arrows. As a result, the same screen was shown with a yellow rectangle surrounding the chosen face. The same trial was repeated if participants didn't indicate a choice within three seconds. The correct choice granted the participants points, while the wrong choice caused a loss of points. Notably, feedback on a given trial was not presented regarding the correctness of the choice. Instead, the final score of points collected throughout the task was presented at the end of each session. The accumulated points from all sessions of the discrimination task and the gained points from the learning task were translated to a bonus monetary reward at the end of the experiment.

In the EEG version of the experiment, participants initially had the opportunity to observe three original faces, and memorize their appearance before the trials began. Subsequently, in each trial, the three original faces are presented briefly, followed by the appearance of the morphed face (random order). After 1000 ms (while the morph is still visible), arrows appear on the screen, allowing participants to make their selections (refer to Supplemental Figure 1E). In all

experiments, each morph was presented six times (a total of 396 trials). The task lasted approximately 25 minutes in all experiments. (See Figure 1C and Supplementary Figure 1D, E).

Psychomotor Vigilance Task (PVT). Participants were instructed to press a mouse buttonwhen numbers, representing a running RT timer, appeared on a screen. The inter-stimulus interval (ISI) varied randomly between 2–10 seconds, and the task lasted five minutes[24]. Sustained attention performance was assessed using mean and SD of RT, 1/RT, as well as lapse counts (RT > 500 ms) and errors (button presses without a stimulus). These metrics are highly sensitive to sleepiness, making the PVT a standard tool for vigilance-related assessments[25–27].

Experiments*:*

Behavioral Experiment

Participants: Sixty healthy participants took part in the behavioral experiment, randomly assigned to a **day group** (n=30) or a **night group** (n=30). Written informed consent was obtained from each participant, and the study was approved by the IRB of the Weizmann Institute of Science. No participants were excluded. The final sample included 30 participants in the night group (19 females, mean age=25.57, SD=3.11) and 30 in the day group (14 females, mean age=28.13, SD=6.4).

Procedure: Participants completed two experimental sessions ~8–12 hours apart. Day group participants were tested in the morning and again in the evening; night group participants were tested in the evening and returned the next morning after sleeping at home. In the first session (~1 hr 15 min), participants completed sleepiness (SSS) and vigilance (PVT) assessments (~10 min), followed by two discrimination task sessions (~25 min each) separated by a learning task (~15 min). In the second session (~1 hr 15 min), they completed another SSS and PVT (~10 min) and a final discrimination task (~25 min), before receiving monetary compensation. Session started at 8:00–9:30 a.m. or 8:00–9:30 p.m., depending on group assignment. Sleep in the night group and wake activity in the day group were monitored via actigraphy between the sessions. (See Figure 1D).

fMRI Experiment

Participants: Ninety healthy, right-handed participants were recruited for the fMRI experiment. They were assigned to a **day group** (n=45) or a **night group** (n=45). Written informed consent was obtained, and all procedures were approved by the IRB of the Weizmann Institute of

Science and the Medical IRB of the Tel-Aviv Sourasky Medical Center. Ten participants were excluded due to technical issues with the scanner or task scripts (n=5), falling asleep during scanning (n=1), voluntary withdrawal (n=2), or excessive motion inside the scanner (n=1), and their data were neither preprocessed nor analyzed. The final analysis included 40 participants in the night group (21 females, mean age=26.97, SD=5.66) and 40 in the day group (25 females, mean age=26.1, SD=3.71).

Procedure: Each participant completed two sessions spaced ~8–10 hours apart. The day group was tested in the morning (7:00–9:00 a.m.) and returned that evening; the night group was tested in the evening (7:00–9:00 p.m.) and returned the following morning. The first session (~2.5 hrs) included sleepiness and vigilance assessments (~10 min), the first discrimination task (~25 min), and the learning task conducted inside the MRI scanner (~1 hour + preparation). After scanning, participants completed a second discrimination task (~25 min) and left the lab. The second session matched the behavioral format, including SSS, PVT, and the final discrimination task. Sleep behavior in the night group and wake behavior in the day group were tracked using an actigraph watch between the sessions. (See Figure 1D).

Sleep EEG Experiment

Participants: Forty-five healthy participants were recruited for the sleep EEG experiment. Written informed consent was obtained, and procedures were approved by the Medical IRB of the Tel Aviv Sourasky Medical Center. Three participants were excluded due to noisy or corrupted EEG data, resulting in a final sample of 42 participants (26 females, mean age=25.71, SD=3.17).

Procedure: All participants were assigned to a **night group** and slept in the lab under full polysomnographic and high-density EEG monitoring. They arrived ~3 hours before their regular bedtime. Setup included EEG net application (~1 hr), followed by SSS and PVT assessments (~10 min), and two discrimination task sessions (~25 min each) separated by a learning task (~15 min), totaling ~2.5 hours. After overnight sleep and a short recovery period, participants completed a second sleepiness and vigilance assessment (~10 min), and a final discrimination task (~25 min). The EEG net was removed, and participants were compensated. (See Figure 1D).

Actigraphy

In the fMRI and behavioral experiments, we provided the participants with an actigraph watch (Ambulatory Monitoring, Inc. motionlogger Micro Watch). This watch measured acceleration, light, and temperature. These measures were used to validate sufficient sleep in the night groups and the absence of daytime naps in the day groups.

fMRI

fMRI acquisition: Functional imaging data were collected using a 3T Siemens MAGNETOM Tim-Trio scanner with a 32-channel head coil. T2*-weighted images were acquired with a gradient-echo EPI sequence (TR=1450 ms, TE=30 ms, flip angle=90°), with 60 descending slices (3 mm thick, 10% gap) at a 30° coronal angle from the ACPC plane. Voxel size was 2.3 mm^3 (FOV=220), and phase encoding was anterior-to-posterior. Functional scanning occurred in two sessions with up to a 2-minute interval. High-resolution anatomical T1-weighted images were also acquired (TR=2300 ms, TE=2.29 ms, flip angle=8°, voxel size=0.9 mm^3, FOV=240). Functional scans covered nearly the entire brain, excluding a small dorsal parietal region. The anatomical scan covered the whole brain. **fMRI preprocessing:** Preprocessing was performed using *fMRIPrep* 21.0.0 [35,36], based on Nipype 1.6.1 framework [37,38]. Fieldmaps for B0 distortion correction were estimated using topup on EPI reference images [39]. T1-weighted anatomical images were corrected for intensity nonuniformity (N4BiasFieldCorrection [40] implemented in ANTs version 2.3.3.[41]), skull-stripped (Nipype implementation of ANTs-based brain extraction using the OASIS30ANTs template), and segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using FSL's FAST [42]. Nonlinear spatial normalization to a standard space (MNI152NLin2009cAsym [43]) was completed using ANTs (antsRegistration). For each BOLD run, a reference volume was generated for motion correction (FSL mcflirt [44]). Slice timing correction was applied (AFNI's 3dTshift [45]). Co-registration to the T1-weighted image was done using FreeSurfer's *bbregister* with boundary-based registration [46]. Functional images were resampled to MNI space using antsApplyTransforms (with Lanczos interpolation to minimize smoothing). Functional images were resampled to standard space using *antsApplyTransforms* with Lanczos interpolation [50], and surface-based resampling was done with FreeSurfer's *mri_vol2surf* [47].Confound regressors included framewise displacement (Power [48] and Jenkinson [44]), DVARS, and global signals (CSF, WM, whole brain). Temporal and anatomical CompCor were computed [49] temporal (tCompCor) components were derived from the top 2% most variable voxels, and anatomical (aCompCor) components were extracted from masks of CSF, WM, and combined compartments, with GM faction minimized by excluding voxels with >5%

GM partial volume [49]. For each CompCor method, components explaining 50% of nuisance variance were retained. All regressors, including motion estimates and their derivatives/quadratics, were included in the confound file. Volumes exceeding FD > 0.9 mm or DVARS > 1.5 SD were flagged as motion outliers.

Whole-brain GLM analysis: Voxel-wise statistical analyses were conducted across the whole brain. At the first level, individual events were modeled using a canonical hemodynamic response function (HRF), and framewise displacement estimates from preprocessing were included as nuisance regressors to control for motion. Each participant's two imaging sessions were modeled separately using a first-level general linear model (GLM), then combined in a second-level group analysis. The event-related design included six regressors, corresponding to the appearance of the face's stimulus and the feedback, each with positive, negative, or neutral valence. For the face stimuli, valence was determined by the expected feedback associated with the participant's choice. Regressors were time-locked to event onset and modeled over the full event duration: three seconds for the stimulus and two seconds for the feedback. All participants were included in the group-level model regardless of sleep condition (day/night), enabling unbiased voxel selection for downstream analysis. Group-level statistical maps were corrected for multiple comparisons using a family-wise error rate (FWER) threshold of $p < 0.001$, with a minimum cluster size of 10 voxels.

ROI selection*:* A predefined list of 28 regions of interest (ROIs) was compiled based on the Harvard-Oxford Atlas, guided by prior research implicating these regions in reward/punishment processing, decision-making, emotion regulation, and face perception. ROIs were included in the final analysis only if they (1) passed the corrected GLM threshold for both contrasts mentioned below and (2) were part of the predefined list. Fourteen regions met these criteria. See Table 1.

ROI name	**Included ?**	**ROI name**	**Included ?**
Caudate	V	Cuneus	X
Putamen	V	Thalamus	X
Pallidum	V	Posterior Superior temporal gyrus	X
Hippocampus	V	Occipital Fusiform Gyrus	X
Amygdala	V	Fusiform Face Area	X
Accumbens	V	Occipital Face Area	X
Precuneus	V	Posterior Temporal Fusiform Cortex	X

Orbitofrontal Cortex	V	Anterior Temporal Fusiform Cortex	X
Medial Frontal Cortex	V	Temporo-occipital fusiform cortex	X
Subcallosal Cortex	V	Superior Temporal Sulcus	X
Anterior Cingulate Cortex	V	Anterior Middle Temporal Gyrus	X
Posterior Cingulate Gyrus	V	Posterior Middle temporal Gyrus	X
Anterior Superior Temporal Gyrus	V	Postcentral Gyrus and the Brainstem	X
Insula	V	Brainstem	X

Table 1: List of 28 a priori regions of interest (ROIs) based on the Harvard-Oxford Atlas. Fourteen ROIs that passed statistical thresholding (FWE-corrected $p < 0.001$, cluster size ≥ 10 voxels) were included in the final analysis (left column).

For each of the 14 ROIs, a 3 mm-radius sphere was defined around the peak activation voxel. Activity was extracted from both the peak voxel and its surrounding sphere using each participant's unthresholded statistical maps. If a sphere extended beyond anatomical ROI boundaries, only the portion within the ROI was retained. To assess differences in neural responses, we computed two contrasts: Gain (positive > neutral feedback) and Loss (negative > neutral feedback). Contrast values were averaged across both scanning sessions for each participant and used in subsequent correlational analyses with behavioral generalization measures.

Sleep EEG

Data Acquisition: Continuous high-density EEG and full polysomnography were recorded throughout the entire experimental session, which included both cognitive tasks and sleep periods. A 256-channel hydrogel Geodesic Sensor Net (Electrical Geodesics, Inc., EGI) was used for data acquisition. Each electrode consisted of a silver chloride carbon fiber pellet, a lead wire, and a gold-plated pin, and was filled with conductive gel (Electro-Cap International). Signals were amplified using an AC-coupled, high-input-impedance amplifier (NetAmps 300, EGI) with an integrated antialiasing analog filter. Recordings were referenced to Cz and digitized at a sampling rate of 1000 Hz. Prior to recording, electrode impedance was verified to be below 50 kΩ for all sensors.

Sleep Scoring: Sleep stages were manually scored offline according to the American Academy of Sleep Medicine (AASM) guidelines. Each recording was scored by two trained raters to

ensure reliability. Standard sleep parameters, including sleep latency, total sleep time, and time spent in each sleep stage, were calculated for subsequent analysis.

Preprocessing: Analyses were conducted in Python using the sleepEEGpy package[50], an open-source "wrapper" pipeline that integrates MNE-Python[51], PyPREP[52], YASA[53], SpecParam[54], and other tools. Each EEG recording was down-sampled from 1000 Hz to 250 Hz and band-pass filtered between 0.75–40 Hz using a zero-phase FIR filter with a Hamming window. Manual inspection of raw signals was performed using the "butterfly" view, complemented by topographical maps to identify and exclude abnormal channels. Subsequently, Independent Component Analysis (ICA) was applied using the extended infomax algorithm[55] with 40 Principal Component Analysis (PCA) components. Components associated with ECG artifacts were identified through visual inspection and removed.

Power Spectral Analysis: Power Spectral Density (PSD) was computed separately for each sleep stage using the Welch method as implemented in SleepEEGpy [50]. Parameters included a Fast Fourier Transform (FFT) length of 2048 samples, a Hamming window, and a 50% segment overlap (1024 samples). For each participant, the maximum sigma band power (12.5–15 Hz) across all channels was identified. The sigma power from the peak channel was used in subsequent correlation analyses.

Sleep Spindle Detection: Analyses were conducted in Python using sleepEEGpy[50] employing the YASA algorithm [53]. Spindles were detected during Non-rapid eye movement (NREM) sleep from each participant's peak sigma channel. Fast spindles were defined within the 13–15 Hz frequency range, with durations constrained between 0.5–2.0 seconds and a minimum inter-spindle interval of 500ms. Detection criteria included: Correlation ≥ 0.65 (moving correlation between the raw and sigma-filtered signal), Relative power ≥ 0.2 (ratio of power in the spindle band to broadband power), root-mean-square (RMS) amplitude ≥ 1.5 standard deviations above the mean of the moving RMS of the sigma-filtered signal.

Generalization analysis

Generalization assessment: We analyzed each participant's response patterns separately for each session. We calculated the proportion of responses classifying each morphed face as most similar to one of the original faces, organizing them into three triangular spaces based on valence: positive, negative, or neutral. As face–valence pairings were counterbalanced, triangles were labeled by valence rather than face identity (Figure 1E, top left). We interpolated

each triangle to generate continuous response gradients (Figure 1E, top right). A 50% threshold defined perceptual boundaries, areas where faces were most likely judged similar to the original. The size of these areas served as a measure of generalization (Figure 1E, bottom left; for visualization, all three are shown on one triangle). To assess changes over time, we computed two metrics: (1) immediate generalization: the area change from Session 1 to Session 2 (after learning), and (2) delayed generalization: the change from Session 2 to Session 3 (after sleep or wake; Figure 1E, bottom right). Note: While a 66% threshold was used in the figures for clearer visualization, all analyses were conducted using the 50% threshold.

Generalization score: In the fMRI experiment, we computed a single score to capture the balance between generalization of positive and negative faces. This generalization score was calculated by subtracting negative face generalization from positive face generalization. Scores above zero indicate a bias toward generalizing the positive face, while scores below zero indicate greater generalization of the negative face. This measure reflects the idea that increased generalization of one emotional valence occurs at the expense of the other.

Renormalization index: To quantify changes in generalization over time, we computed a renormalization index for each participant by subtracting the immediate generalization score from the delayed generalization score. In the fMRI experiment, we used the generalization bias score (positive minus negative) for this calculation. In the EEG experiment, we focused on generalization of the positive face only.

Overgeneralization Calculation: Overgeneralization was defined as instances where participants incorrectly classified morphed faces as resembling one of the original faces, beyond the boundaries of the correct generalization zone. In the generalization task, such incorrect responses resulted in point loss and were considered maladaptive. Regions of the triangular face space that extended beyond the adaptive generalization boundaries were designated as overgeneralization areas. See Supplementary Figure 7A, B for illustration.

Statistics: Group-level effects in generalization were assessed using mixed-model ANOVAs. Significant interactions were followed by paired or independent-samples two-tailed t-tests, depending on the comparison. To examine whether brain ROI activity predicted generalization, we applied both linear regression and LASSO (least absolute shrinkage and selection operator) regression. All predictors were z-scored prior to LASSO to ensure comparability. The

regularization parameter was optimized using 10-fold cross-validation and averaged across folds. Linear regression models were then constructed using the subset of predictors selected by LASSO.

Pearson correlations were used to assess associations between behavioral generalization scores and neural measures. Specifically, we examined correlations between amygdala activation (from ROI-based fMRI contrasts) and generalization indices, as well as between EEG-derived metrics (e.g., spindle count and sigma power) and changes in generalization across sessions. Statistical significance was set at $p < .05$.

Learning task analysis

To confirm that participants learned the associations between faces and outcomes, we calculated the percentage of correct choices for each participant across all trials. In all experiments, all participants performed above chance level, with a minimum accuracy of 60%. The average accuracy across all three experiments was 93.1% (SD=5.9%). The learning curve shown in Figure 1B reflects data from all participants across the three experiments (N=182), smoothed using a moving average over a window of five trials.

Results

To quantitatively study the role of sleep in the generalization of sensory stimuli with different emotional valence in healthy participants, we first employed a series of repeated behavioral discrimination tasks (N=60). Participants were presented with face images generated through morphing, depicting intermediate levels along a continuum between three original faces (Figure 1A). For each image, participants judged which original face it most closely resembled (Figure 1C). In a second task, using a probabilistic reinforcement-learning procedure (two-arm bandit), participants learned to associate the original faces with positive, negative, or neutral feedback (Figure 1B). Participants completed the discrimination task three times (Figure 1D): before (session 1) and immediately after (session 2) learning the associations between faces and feedback, as well as following an 8-10 hour interval of nighttime sleep or daytime wakefulness (session 3).

Behavioral responses from each session were organized into a triangular face space representing perceived similarity between each morph and the original faces. Areas around each vertex where similarity exceeded a 50% threshold were compared between two sessions

to quantify generalization (Methods, Figure 1E). Put simply, generalization was marked by a wider area of morphed faces in the triangle that was perceived as identical to the original face in the vertex. Throughout the paper, we refer to differences in generalization while learning associations as 'immediate generalization' (session 2 - session 1) and refer to differences across the long daytime wake or nighttime sleep intervals as 'delayed generalization' (session 3 - session 2).

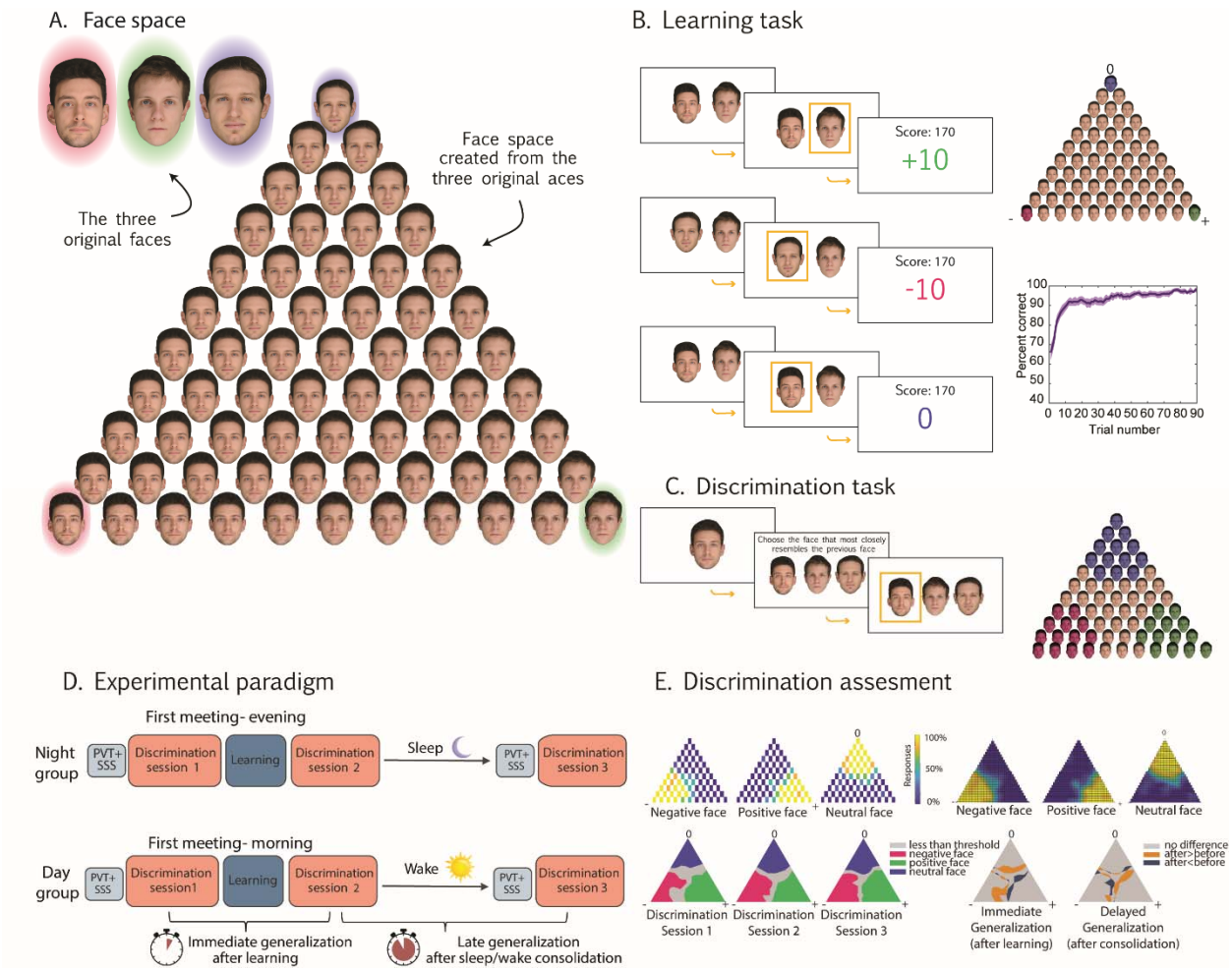


Figure 1: Experimental design: A. Face Space: Top left: the three original faces used in the experiment. Right: The triangular "face space" generated from the three original faces. Faces at the vertices are the original faces, while intermediate positions represent morphed faces created by blending the three originals. Each morph's proximity to a vertex indicates the proportion of that original face in the morph. The face at the center is the average of all three original faces. **B.** Learning task trial timeline. Left: During each trial of the learning task, two faces (selected from the three original faces) were displayed. Participants chose one face and received feedback (+10, -10, or 0 points) based on their selection. Each original face was probabilistically associated with positive, negative, or neutral feedback (90% probability per face for each category). Feedback-face associations were randomized across participants. Detailed timing for different experiments is provided in Supplemental Figure 1. Top-right: Feedback association illustration for the original faces at the vertices of the face space (randomized for each participant). Bottom-right: learning curve from participants across all three experiments (behavioral, fMRI, EEG). **C.** Discrimination task trial. Left: Each trial began with a figure of a morphed face from within the face space. Then, participants were asked to choose from the three original faces the face that most resembled the presented morph. Selections were indicated by

pressing one of three arrow keys, and the chosen face was highlighted with a yellow rectangle. Detailed timing for different experiments is provided in Supplemental Figure 1. Right: illustration of generalization areas around each original face**. D.** Schematic timeline of the experimental procedure: Participants completed two meetings, approximately 8–12 hours apart, either across the day (Day group) or overnight (Night group). Each meeting included assessments of sleepiness and vigilance. The first meeting included two sessions of the discrimination tasks, separated by the learning task, and the second meeting included the last session of the discrimination task. In the behavioral and fMRI experiments, participants were assigned to either the Day or Night condition, and sleep or wakefulness between sessions was monitored via actigraphy. In the sleep EEG experiment, only a Night group was tested**. E.** Generalization Assessment Example. Top left: Triangular response space for a single participant, with colors indicating the proportion of responses identifying a morph as similar to the negative (left), positive (middle), or neutral (right) associated face. Top right: Interpolated representation of the triangular response space. Bottom left: Simplified triangular spaces generated by applying thresholds for each original face and merging areas from the same session into a single triangle. Bottom right: Segment changes across learning – immediate generalization (left). And post-sleep/wake intervals – delayed generalization(right).

Sleep selectively promotes generalization of positive associations

We observed distinct generalization patterns between day and night groups, finding that wakefulness and sleep had different effects on generalization. Specifically, participants in the night group were more likely to generalize the positive associated face, whereas participants in the day group more often generalized the negative associated face (see Figure 2A for representative examples from each group). To quantify and assess the effect of the sleep/wake interval on generalization, we conducted a two-way, mixed-model ANOVA on delayed generalization, with Group (Day/Night, between subjects) and Face (Positive/Negative/Neutral, within-subjects) as factors. The analysis revealed a significant Face × Group interaction ($F(2,116)=3.61$, $p=0.03$, Figure 2B, 2C). Post-hoc comparisons indicated that the negative face was generalized more during daytime wakefulness than during overnight sleep ($t(58)=-2.3$, $p=0.025$). By contrast, the positive face was more generalized during sleep than wake ($t(58)=2.09$, $p=0.049$).

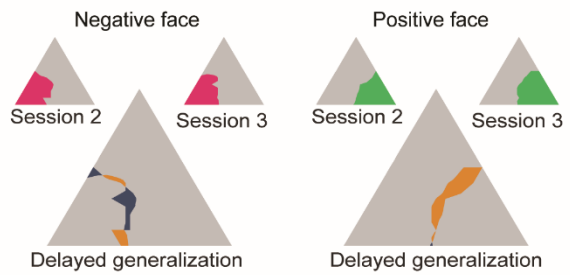
To further understand this effect, we conducted separate two-way, mixed-model ANOVAs for the negative and positive faces, considering the total classified area in each session (rather than only the session-to-session difference; Methods, Figure 2D). For negative face generalization, a mixed-model ANOVA with Group (Day/Night) and Session (Session 2/Session 3) as factors revealed a significant Group × Session interaction ($F(1, 59)=2.34$, $p=0.02$; Figure 2D, left). Post-hoc comparisons indicated a significant increase in the area classified as negative in Session 3 compared to Session 2, but only in the daytime wake group ($t(29)=–2.44$, $p=0.02$). A similar analysis for positive face generalization also revealed a significant Group × Session interaction ($F(1, 59)=2.04$, $p=0.04$; Figure 2D, right). Here, post-hoc comparisons showed a significant

increase in the area classified as positive in Session 3 relative to Session 2, but only in the sleep group ($t(29)=-2.14$, $p=0.04$). No significant group differences were observed for neutral face generalization (Figure 2E, 2F), nor were there differences in immediate generalization prior to the sleep/wake interval (Figure 2G, 2H). Together, these results suggest a dissociation in generalization patterns: overnight sleep preferentially enhances generalization of positive stimuli, whereas daytime wakefulness enhances generalization of negative stimuli.

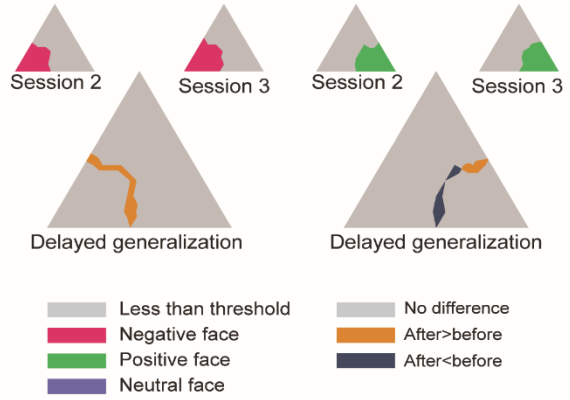
To control for potential confounds, we compared subjective sleepiness (SSS) and vigilance (PVT) between groups. No significant group differences were found in either measure during the first meeting. In the second meeting, participants in the day group reported higher sleepiness than the night group (Day: 3.87 ± 1.63 vs. Night: 3.03 ± 1.34; $t(58)=2.15$, $p=0.036$); however, no group differences were observed in vigilance as measured in the PVT (a highly sensitive measure of sleepiness[26,56]). Additionally, neither SSS nor PVT scores were significantly correlated with generalization measures in either group (see Discussion).

A. Examples for delayed generalization

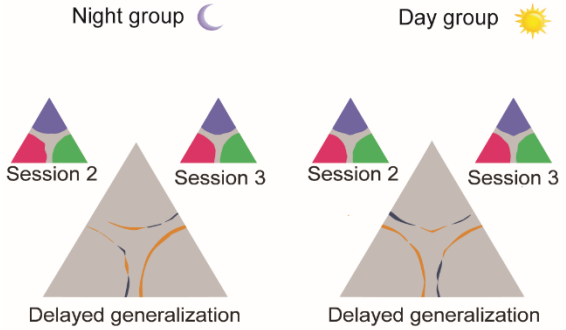
(i) Example from night group



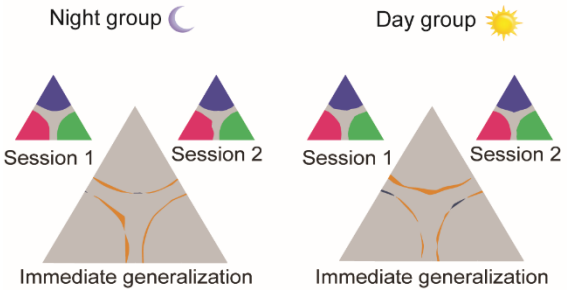
(ii) Example from day group



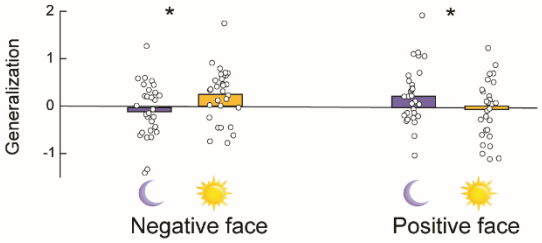
B. Averages for day and night groups (all faces) delayed generalization



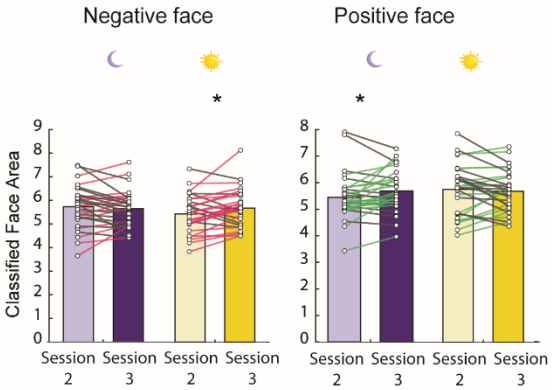
G. Averages for day and night groups (all faces) Immediate generalization



C. Delayed generalization: negative and positive faces

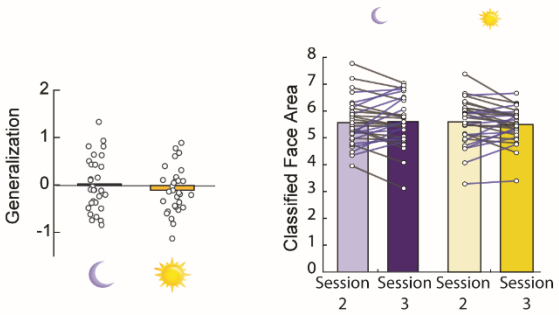


D. Sessions 2 and 3: negative and positive faces



E. Delayed generalization: Neutral face

F. Sessions 2 and 3: Neutral face



H. Immediate generalization: all faces

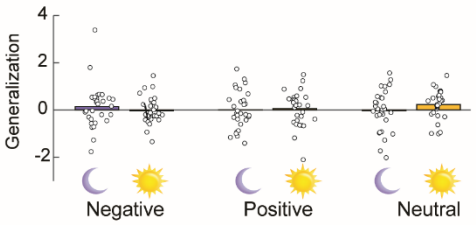


Figure 2: Sleep selectively promotes generalization of positive associations.

A. Examples of Delayed Generalization from Individual Participants: (i) Night group participant: In the small triangles, pink areas indicate morphed faces classified as the negative face in sessions 2 and 3 of the discrimination task; green indicates the positive face. The large triangle illustrates delayed generalization, changes in the generalization areas between sessions 2 and 3 across the night. This participant shows increased generalization of the positive face overnight. (ii) Day group participant: This participant shows delayed generalization of the negative face after a day of wakefulness. **B**. Group-Level Delayed Generalization: Small triangles show averaged morph areas classified as each of the original faces during sessions 2 and 3, combined onto one triangle. Large triangles show the change in generalization between sessions. Left: night group; right: day group. Night group shows increased delayed generalization of the positive face, day group of the negative face. **C.** Summary of Delayed Generalization: Bar graph shows average change for positive and negative faces from (B). Night group shows significantly greater generalization of the positive face; day group to the negative face. **D.** Session-wise Change in Classification Area: Bar graph shows size of areas classified as positive and negative in sessions 2 and 3. Night group shows increased positive face classification in session 3; day group shows increased negative face classification. **E.** Neutral Face Delayed Generalization: No significant difference between groups. **F.** Session-wise Neutral Face Classification: No significant difference from session 2 to 3 in either group. **G**. Group-Level Immediate Generalization: Same as (B), comparing sessions 1 and 2. No significant change in any face for either group. **H.** Summary of Immediate Generalization: Bar graph shows average immediate generalization for all faces. No significant differences in either group.

Sleep reverses amygdala-associated generalization of negative stimuli

Next, we set out to investigate the neural correlates associated with generalization of positive and negative faces and their dynamics using fMRI, employing a modified version of the learning task performed in the scanner (Methods). We could not replicate the group-level behavioral differences between generalization after extended wakefulness versus sleep (Supplementary Figure 2, see also Discussion), observing considerable inter-individual variability in generalization patterns. During immediate testing, some participants exhibited stronger generalization of the positive face (Figure 3A), whereas others showed a bias toward generalization of the negative face (Figure 3B). Similar individual variability was observed in delayed generalization after daytime wakefulness or after overnight sleep. We focused subsequent analysis on testing whether these individual variabilities are associated with specific brain activities.

First, we tested whether brain activity during learning, specifically when receiving feedback linked to point gains or losses, could predict immediate and delayed generalization behavior. To quantify generalization, we computed a composite score calculated as the difference between positive- and negative-face generalization scores (Methods), with values above zero indicating greater generalization of the positive face and below zero indicating greater generalization of the negative face. This score was used in multiple linear regression models to predict either (i) immediate generalization (for both groups combined, as it is prior to the sleep/wake

consolidation period) or (ii) delayed generalization following daytime wake or overnight sleep (separated between the groups). fMRI blood oxygenation level dependent (BOLD) activity underwent whole-brain general linear model (GLM) analysis to identify task-related activity, and 14 Regions of Interest (ROIs) were selected as predictors (Methods). To identify the most predictive ROIs, we applied Lasso regression [57] (elastic net regularization) with 10-fold cross-validation for model optimization.

We found that greater activation for losses (versus neutral feedback) in limbic regions, specifically the Amygdala, Caudate, Putamen, and Subcallosal Cortex, was correlated with increased immediate generalization of the negative face, but decreased immediate generalization of the positive face, This corresponds to a negative correlation with the generalization composite score (regression model: N=80, R^2=0.18, F(4,75)=4.17, p=0.004; Figure 3C, D). A similar pattern emerged during gains (versus neutral feedback): activity in the Amygdala, Orbitofrontal cortex, and Insula was positively correlated with generalization of the negative face and negatively with the positive face (positive correlation to the generalization score), whereas activity in the Precuneus (default mode network, DMN), showed the opposite trend: (regression model: N=80, R^2=0.18, F(4,75)=4.14, p=0.004, Figure 3E, F). Thus, regardless of feedback valence (gaining or losing points associated with monetary reward), limbic activity biases immediate generalization toward negatively associated stimuli, whereas engagement of the DMN is associated with positive stimuli immediate generalization. These effects were also evident when modeling positive and negative generalization separately, rather than using the generalization composite score. Specifically, limbic activity during learning mostly predicted increased immediate generalization of the negative face (see Supplementary Figure 3).

Next, we analyzed brain activity associated with delayed generalization after overnight sleep. Greater limbic activation during both loss- and gain-related feedback in the learning session was associated with a renormalization from negative to positive face generalization following overnight sleep. Specifically, activity in the Amygdala, Nucleus Accumbens, and Anterior Superior Temporal Gyrus during loss-related feedback predicted increased generalization of the positive face after sleep (regression model: N=40, R^2=0.27, F(3,36)=4.45, p=0.009; Figure 3G, H). A similar trend was observed for gain-related feedback, where higher activation in the Amygdala and Anterior Cingulate Cortex during gains was associated with a greater shift in generalization from the negative to the positive face following sleep (regression model: N= 40, R^2=0.18, F(2,37)=4.14, p=0.02; Figure 3, I-J). Similarly, separate models for delayed positive

and negative generalization, instead of the generalization composite score, showed that limbic activation during learning predicted a sleep-dependent shift toward greater generalization of the positive face (see Supplementary Figure 4).

Together, these findings suggest that heightened limbic engagement during emotionally salient learning—regardless of outcome valence—may promote immediate negative generalization, followed by a sleep-dependent renormalization that shifts toward positive generalization. In contrast, applying the same Lasso regression approach to predict delayed generalization after daytime wakefulness did not select any ROI predictors for either gain or loss, suggesting that this renormalization process may be specific to sleep-related consolidation (see Discussion).

Finally, no significant group differences in subjective sleepiness (SSS) or vigilance (PVT) were observed at either time point. Additionally, neither SSS nor PVT scores were significantly correlated with generalization measures in either group.

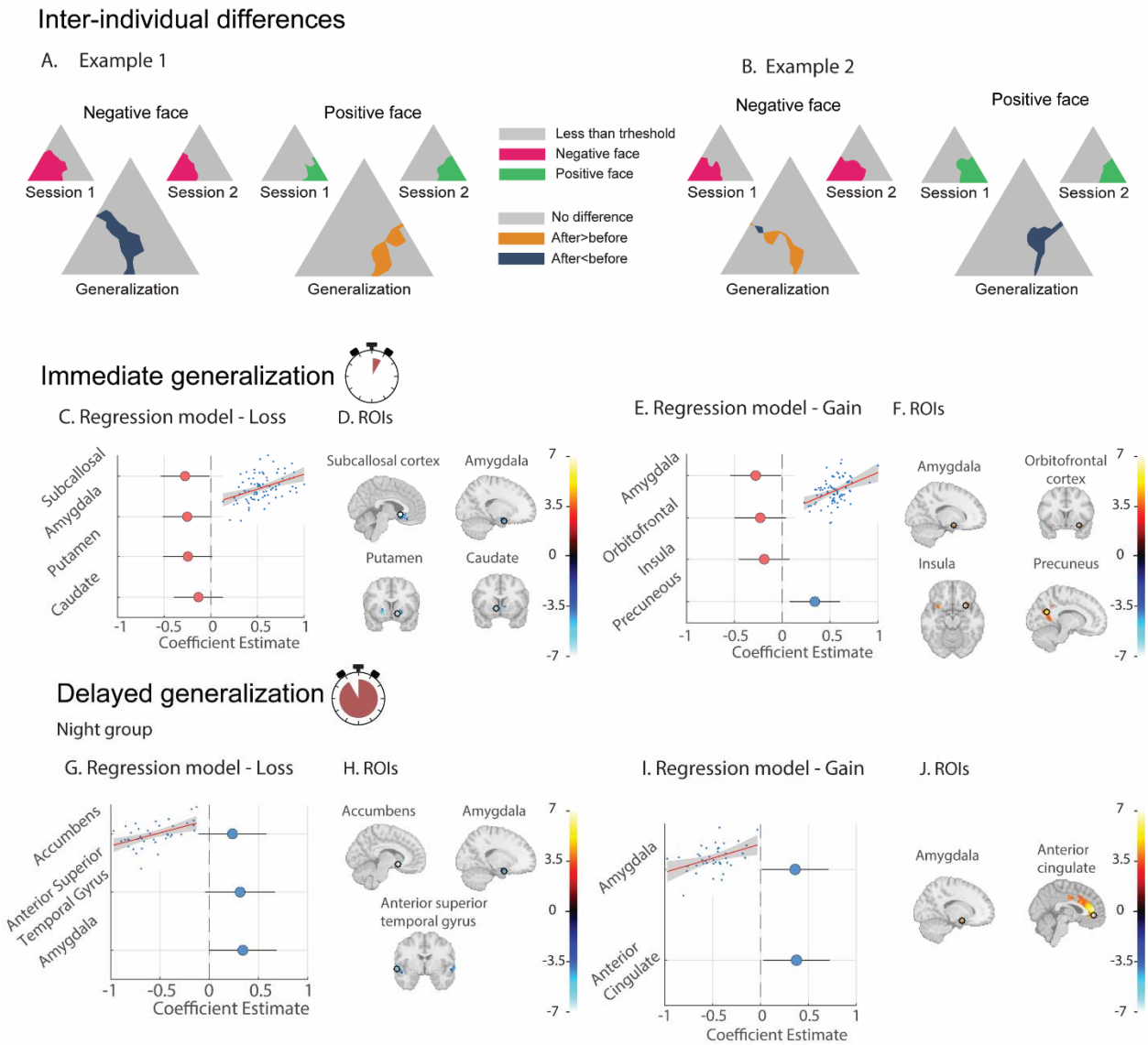


Figure 3. Immediate and late generalization are correlated with Amygdala and limbic activity. A–B. Individual variability in immediate generalization. **A.** A participant who generalized the positive face immediately after learning. Left: areas classified as the negative face during Sessions 1 and 2. Right: areas classified as the positive face. Larger triangles represent changes in classification between sessions, reflecting immediate generalization. **B.** regions classified as the negative face during Sessions 1 and 2. Right: regions classified as the positive face. **C-D.** Regression results for predicting immediate generalization from neural activity during loss-related feedback. **C.** Coefficient estimates for ROIs selected via Lasso regression: Subcallosal Cortex, Amygdala, Putamen, and Caudate. All coefficients are negative, indicating that greater activity in these limbic regions was associated with reduced generalization of the positive face and increased generalization of the negative face. Inset (top right): model fit (x-axis=predicted generalization from ROI activity; y-axis=observed generalization). Red line=regression fit; gray shading=95% confidence intervals. **D**. Mean ROI activity during loss feedback for regions in (C). Black circle marks a 3-voxel-radius sphere centered on the peak voxel within each anatomical ROI. **E-F**. Immediate generalization: gain-

related feedback. **E.** ROIs selected via Lasso included the amygdala, orbitofrontal cortex, insula, and precuneus. Limbic regions had negative coefficients, linking greater activity with reduced positive-face generalization. Precuneus (DMN) activity was positively associated with the generalization of the positive face. Inset as in C. **F.** Mean ROI activity during gain feedback for regions in (E). **G-H.** Delayed generalization in the night group: loss-related feedback. **G.** Coefficient estimates for ROIs selected via Lasso regression: Accumbens, Anterior Superior Temporal Gyrus, and Amygdala. All coefficients are positive, indicating that increased activity in these regions was associated with greater generalization of the positive face and less generalization of the negative face. Inset (top left) as in C. **H.** Mean ROI activity during loss feedback for regions in (G). I-J. Delayed generalization: gain-related feedback. **I.** The amygdala and anterior cingulate cortex were selected via LASSO. Both coefficients are positive, suggesting that increased activity in these regions was associated with greater positive-face generalization and decreased negative-face generalization. Inset (top left) as in C. **J.** ROI activity during *gain* trials for regions shown in panel I.

Amygdala activity consistently emerged as an important predictor for immediate and delayed generalization when activated for both positive and negative emotional valence stimuli, being the only ROI repeatedly selected by LASSO in all prediction models. Specifically, amygdala activation during losses was significantly correlated with immediate generalization of the negative face (negative correlation to the composite score) across both experimental groups (N=80, r_p=-0.26, p=0.02, Figure 4A). However, for delayed generalization following overnight sleep, this activity predicted the positive face generalization (N=40, r_p=0.36, *p*=0.02; Figure 4B), a phenomenon that did not occur after daytime wakefulness (Figure 4B, bottom-right, r_p=0.04, p=0.8).

To quantify this effect, we computed a renormalization index for each participant by subtracting the immediate generalization score from the delayed generalization score for each participant (Methods). Positive renormalization indices indicate that, after sleep, the generalization is shifted towards more positive values than those at baseline. In the night group, greater amygdala activation during loss-related feedback predicted stronger renormalization following sleep. Participants with higher amygdala activity not only exhibited greater immediate generalization of the negative face but also demonstrated a clear shift toward positive face generalization after sleep, as reflected by a significant correlation between the renormalization index and amygdala activity during loss (N=40, r_p=0.43, *p*=0.005; Figure 4C). No such correlation was observed in the daytime group (Figure 4C, bottom-right, r=0.12, p=0.45).

A similar pattern emerged when examining amygdala activity during reward-related feedback. Higher amygdala activation while gaining points was associated with greater immediate generalization of the negative face across both groups (N=80, r_p=-0.25, *p*=0.02; Figure 4F). However, the amygdala activity showed trend toward a significant correlation with positive face

generalization following sleep (N=40, r_p=0.28, p=0.08; Figure 4G). Although the correlation was not significant, it was significantly different than the day group that showed no correlation between the amygdala activity and the generalization (Figures 4G bottom-right panel, r_p=-0.16, p=0.31, fisher test to compare night and day correlations: Z=1.9, p=0.03).

Furthermore, the renormalization index showed a strong correlation with Amygdala activity during gains (N=40, r_p=0.4, p=0.009; Figure 4H), indicating that individuals with higher initial amygdala engagement for gain-related feedback also underwent a greater shift from negative to positive generalization after sleep. As was found for loss-related activity, renormalization effects were specific to overnight sleep, no correlation to the renormalization index in the daytime wake group (4H, bottom, right panel, r_p=-0.01, p=0.95).

Sleep-dependent renormalization was consistently observed for Amygdala activity, but not unique to this brain region. Direct correlations with other limbic ROIs revealed a consistent pattern whereby higher activation during learning (either reward or loss) was associated with increased immediate generalization of the negative face, followed by a renormalization toward the positive face after sleep (Supplementary Figures 5 and 6).

In addition, we examined overgeneralization, defined as generalization that is no longer adaptive (i.e., leads to point loss). We found significant correlations between amygdala activity during both loss- and gain-related feedback and two behavioral outcomes: immediate overgeneralization of the negative face and delayed overgeneralization of the positive face following sleep. Notably, overgeneralization scores were highly correlated with overall generalization scores, and the observed associations with amygdala activity did not remain significant after controlling for this shared variance (Supplementary Figure 7).

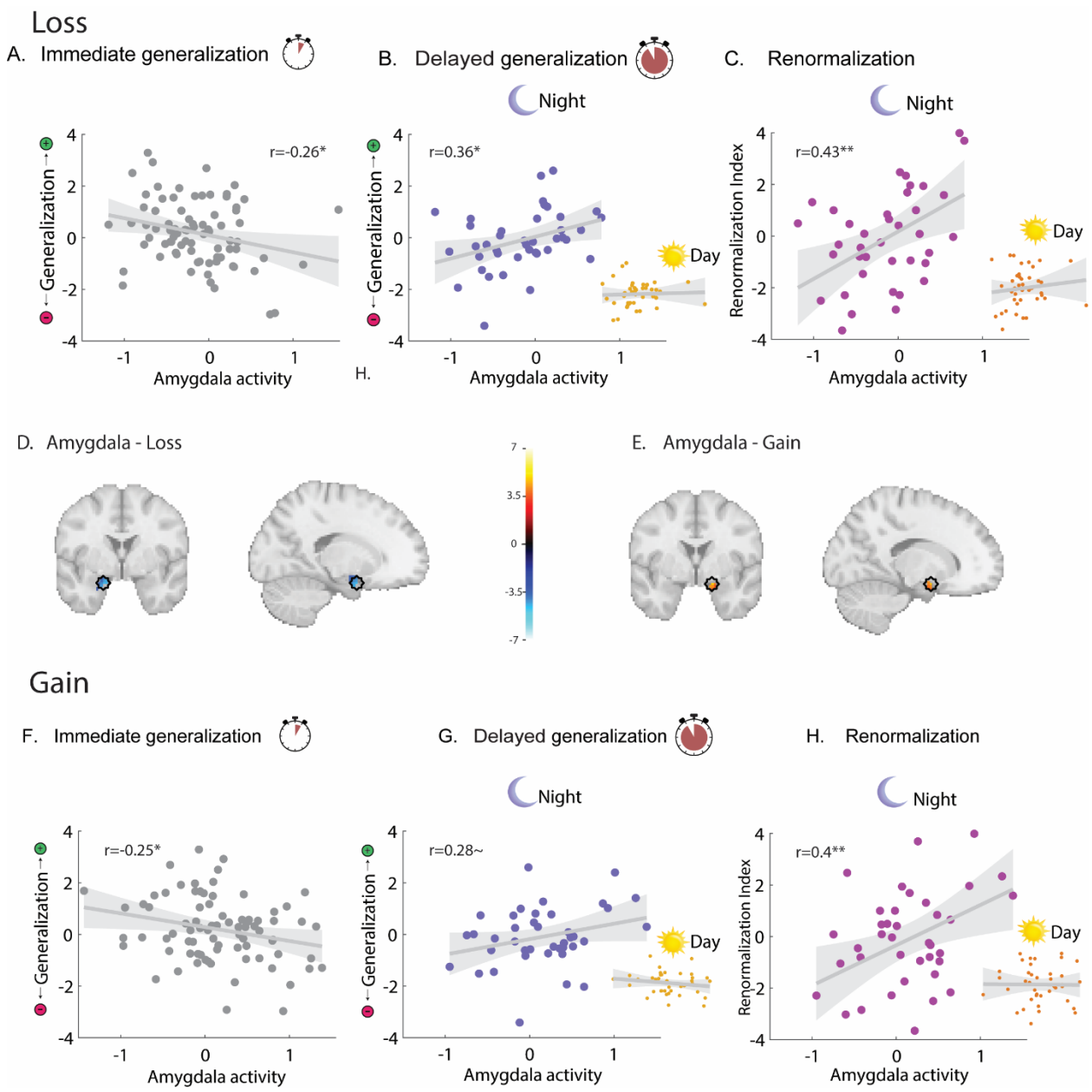


Figure 4. Direct correlations between amygdala activity during learning and generalization behavior. A. Immediate generalization**:** Amygdala activity during loss feedback in the learning phase is negatively correlated with immediate generalization across both groups. Higher amygdala activity is associated with increased generalization of the negative face. The grey line represents the fitted regression line; the gray shaded area indicates 95% confidence intervals. **B.** Delayed generalization**.** Left: In the night group, higher amygdala activity during loss is associated with positive face generalization after sleep. *Bottom right*: No such correlation is observed in the day group, suggesting the effect is sleep-specific. **C.** Correlation with the renormalization index**.** Left: In the night group, higher amygdala activity during loss is significantly correlated with a greater renormalization index, indicating a stronger shift from negative to positive generalization after sleep. Bottom right: No such relationship is observed in the day group. **D.** Activation map showing the amygdala response to loss feedback. The black circle marks a sphere (radius=3 voxels) centered on the peak activation voxel. Activity used for analysis is the average across voxels within this sphere that fall within the anatomical ROI **E.** Same as D for gain feedback. **F–H**. Same analyses as in panels A–C, but for amygdala activity during gain feedback. **F.** Higher amygdala activity during gains is associated with increased immediate generalization of the negative face. **G.** Left: Amygdala activity during gains predicts a renormalization toward positive-face generalization after sleep in the night group. Bottom right: No such effect is observed in the day

group. **H.** Correlation between amygdala activity during gains and the renormalization index: significant in the night group (*left*), absent in the day group (*bottom right*).

Sleep spindle activity correlates with renormalization of negative-to-positive emotional generalization

To further investigate the role of sleep in promoting positive generalization and reversing negative emotional bias, we conducted an overnight sleep study with polysomnography and high-density (256-channel) EEG in a separate cohort of healthy volunteers (n=42, Methods, Figure 5A, B). Participants completed the same behavioral paradigm used in prior experiments, both before and after sleep. Sleep scoring was performed to assess whether sleep architecture was related to generalization outcomes. On average, participants exhibited a typical profile for overnight sleep in healthy young adults, with a mean (SD) total sleep time of 400.2 (69.7) min, sleep efficiency (sleep time/time in bed) of 92.7% (5.2) and a mean sleep latency of 18.3 (15.5) min. Non-rapid eye movement (NREM) sleep, REM sleep, and wake after sleep onset (WASO) constituted 302.2 (32.3), 96.8 (32.3), and 19.7 (17.7) minutes, respectively. We tested for associations between time spent in specific sleep stages and generalization, but no significant relationships emerged (Supplementary Table 1).

Next, we examined whether specific electrophysiological activities during sleep may be correlated with generalization. We focused on sleep spindle activities over centro-occipital regions occurring during NREM sleep in the EEG sigma frequency band (12-15Hz, Methods; Figure 5C)[58] as such activities are implicated in memory consolidation and cortical plasticity[59–63] We found that greater sigma power during NREM sleep was significantly correlated with an increase in delayed generalization of the *positive* face following sleep (r_p=0.38, p=0.01; Figure 5E). We also observed a trend toward positive correlation between *immediate* generalization of the negative face and subsequent sigma power (r_p=0.25, p=0.09), as well as a trend toward negative correlation between immediate generalization of the positive face and sigma power (r_p=–0.26, p=0.09; Figure 5D). These trends suggest that individuals who initially generalized the negative face more strongly than the positive face exhibited increased spindle activity during subsequent sleep. Could this pattern reflect a compensatory mechanism, wherein enhanced spindle activity supports a renormalization of negative emotional generalization? Consistent with this notion, the *renormalization* in positive face generalization (i.e., the increase from immediate to delayed generalization) was also significantly correlated with sigma power during sleep (r_p=0.40, p=0.009; Figure 5F).

To extend this analysis, we went beyond power spectral analysis to detect and quantify individual spindle events using SleepEEGpy[50], incorporating YASA[53] detection (Methods, Figure 5G). We found that the number of fast (13–15 Hz) spindles during NREM sleep was positively correlated with post-sleep generalization of the positive face (r=0.33, p=0.03; Figure 5I) and with the renormalization in positive generalization (r=0.031, p=0.04; Figure 5J). Notably, we also found that a higher number of fast spindles was associated with reduced generalization of the neutral face after sleep (r=–0.38, p=0.013; figure 5H). Since neutral faces were not associated with reward or punishment, this finding suggests that sleep spindles may help suppress the generalization of emotionally irrelevant or neutral stimuli, highlighting a possible role of spindles in valence-selective memory processing (see Discussion).

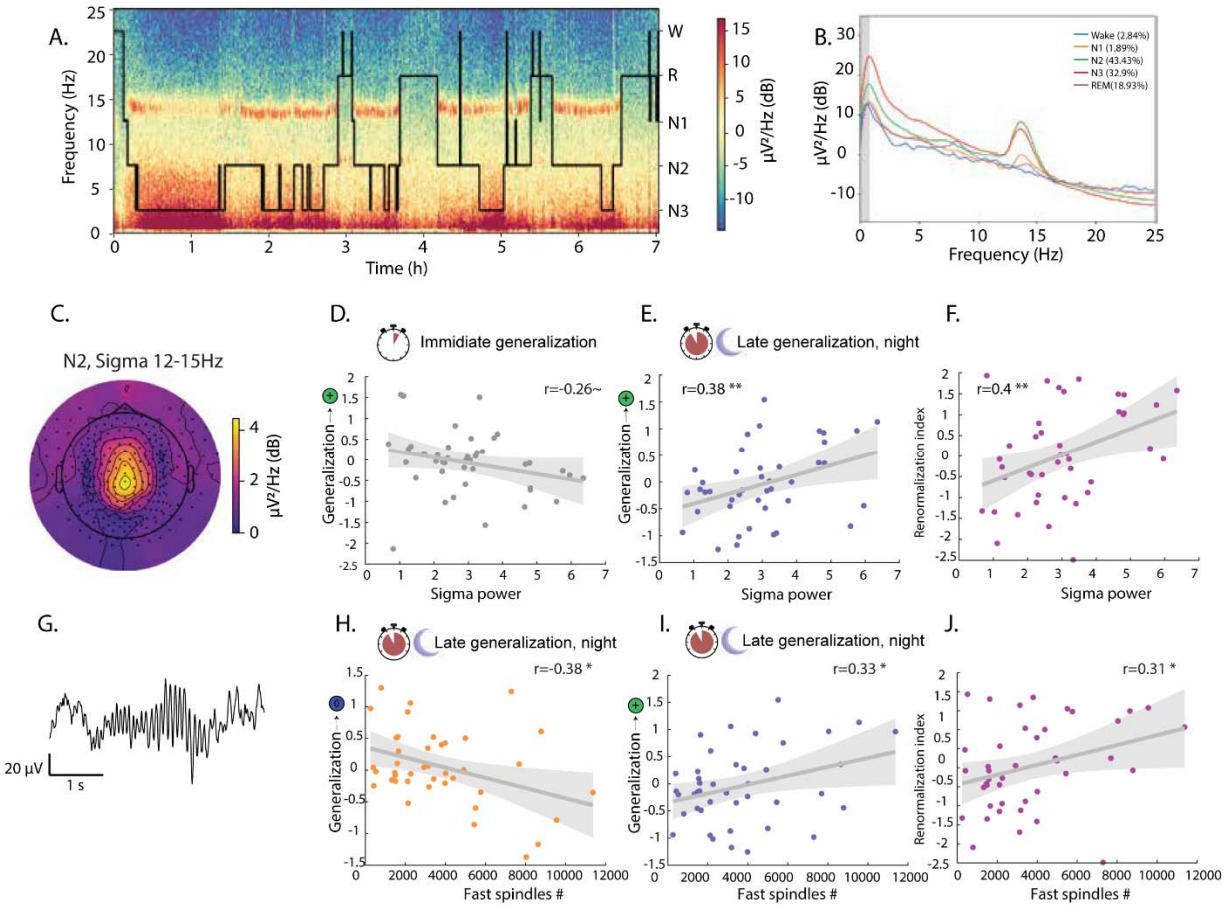


Figure 5: Sleep spindle activity facilitates positive generalization A. Time–frequency spectrogram (log power) from a representative participant recorded at electrode Pz, illustrating spectral dynamics across the night. The overlaid hypnogram depicts transitions between sleep stages. **B.** Power Spectral Density (PSD) plots of preprocessed data from electrode Pz from a representative participant, separated by sleep stage. Colored lines

denote individual stages. The gray-shaded region indicates frequencies removed by the high-pass filter. **C.** Topographical map of sigma power (12–15 Hz) during N2 sleep from a representative participant, demonstrating typical centroparietal distribution of sleep spindles. **D.** Correlation between sigma power during NREM sleep (from each participant's peak-sigma channel) and immediate generalization of the positive face. Participants with lower immediate positive generalization exhibited greater subsequent sigma power. **E.** Higher sigma power during NREM sleep is positively correlated with delayed generalization of the positive face, measured after sleep. **F.** Sigma power during NREM sleep is significantly correlated with the renormalization of the positive face generalization from the pre-sleep (immediate) to post-sleep (delayed) session. **G.** A representative example of a fast spindle (13–15 Hz) recorded during NREM sleep, shown in the time domain. Scale bars indicate 1 second (horizontal) and 20µV (vertical). **H.** Greater spindle count during NREM sleep (from the peak-sigma channel) is associated with reduced delayed generalization of the neutral faces, suggesting selective suppression of non-valenced stimuli generalization. **I.** The number of fast spindles during NREM sleep is positively correlated with delayed generalization of the positive face. **J.** Fast spindle count is also positively correlated with the renormalization in positive generalization from immediate to delayed sessions.

Discussion

In this study, we tested the hypothesis that sleep reduces emotional generalization of negative stimuli while boosting the generalization of positive stimuli. Using a face discrimination task with morphed faces, we found that overnight sleep selectively promoted generalization of positive associations, whereas daytime wakefulness was linked to generalization of negative associations (Figure 2). fMRI analysis revealed that greater limbic responses to emotional feedback during learning (gains or losses) were associated with broader immediate generalization of the negative face. This activity was associated with a reversal of this pattern after sleep: participants exhibited greater generalization of positive faces following overnight sleep, but not after a day awake (Figure 3). Amygdala activity consistently emerged as a main predictor of both immediate and delayed generalization, for both positive and negative emotional valence (Figure 4). Participants with high amygdala activation during learning exhibited the strongest immediate generalization of the negative face and showed the greatest sleep-associated renormalization. Finally, we found that sleep spindle activity was significantly correlated with a shift from negative to positive generalization (Figure 5).

Data from the behavioral experiment revealed a clear divergence in delayed generalization profiles following overnight sleep or daytime wakefulness. Participants who remained awake after learning tended to generalize the negatively associated face, reinforcing a negative bias, while those who slept were more likely to generalize the positively associated face. These findings align with previous research showing that negative stimuli often elicit broader generalization curves than neutral or positive stimuli [1,3,4], a pattern thought to reflect an adaptive threat-detection mechanism. Our results extend this literature by showing that sleep does not

merely preserve these biases but actively reshapes the generalization landscape, attenuating negative overgeneralization and enhancing positive generalization

fMRI data showed that greater amygdala activity during processing of either positive or negative outcomes was linked to broader generalization of negative stimuli immediately after learning. After sleep, this pattern reversed: higher amygdala activity was now associated with greater generalization of positive faces, suggesting that sleep supports a renormalization of amygdala-related emotional memory. This effect was absent in the wake group, reinforcing its sleep dependency. Notably, participants with high amygdala activation showed, after sleep, even greater positive generalization than at baseline, extending beyond simple renormalization. These findings align with previous research linking amygdala activation to loss encoding and the broadening of generalization curves for negatively associated stimuli[1,4]. Moreover, sleep deprivation has been shown to amplify amygdala reactivity to emotional stimuli and impair discrimination between threat and safety cues [9,64], potentially contributing to overgeneralization in anxiety disorders. Our results suggest that sleep may help restore the negative bias by shifting amygdala-linked generalization away from negative and toward positive stimuli.

To investigate the role of sleep-specific neurophysiology in this reorganization, we recorded high-density EEG during overnight sleep. We found that both sigma power and the number of fast spindle events during NREM sleep were positively correlated with the generalization of positively associated faces. Additionally, we observed a trend suggesting that individuals who showed greater negative generalization immediately after learning exhibited increased spindle activity during subsequent sleep, pointing to a possible compensatory or recovery-related renormalization mechanism. Both measures of spindle activity were significantly correlated with the renormalization in positive face generalization, reflecting a shift from negative to positive bias after sleep. These findings extend our understanding of the association between sleep spindles and memory consolidation to show that spindles play a role in emotionally adaptive reorganization of generalized learned associations.

In addition to modulating emotional memory, our results suggest that sleep spindles may contribute to the selective suppression of emotionally neutral information. Specifically, greater fast spindle activity was associated with reduced generalization of neutral faces, which were not linked to reward or punishment. This pattern suggests that spindles may serve a dual function, supporting the generalization of emotionally salient memories while promoting forgetting of

behaviorally irrelevant content. This interpretation extends prior work proposing that sleep supports the prioritization of salient or goal-relevant information during consolidation[12,16,65].

Some limitations of this study should be explicitly acknowledged. First, circadian factors may influence the observed group differences. Because testing sessions occurred at opposite times of day (morning vs. evening), it is difficult to fully disentangle the effects of sleep from those of time of day. In the behavioral experiment, participants in the day group reported higher subjective sleepiness in the evening session than the night group did in the morning, although objective vigilance (PVT) scores did not differ. Previous research has linked fatigue to more negative interpretation biases[9,64], which may in principle partly explain the negative generalization in the day group. However, neither SSS nor PVT scores were significantly correlated with generalization measures, arguing against a major role for circadian factors. Further arguing against circadian confounds, robust correlations between individual sleep spindle activity in the sleep EEG experiment and shifts in emotional generalization support the interpretation that sleep plays an active, corrective role in modulating emotional overgeneralization. Such renormalization may be mediated in part via spindle-related mechanisms operating during NREM sleep. Second, group-level generalization differences in the behavioral experiment were not replicated in the fMRI experiment, where no group differences in subjective sleepiness were found. While we can only speculate about the source of differences between the two experiments, the learning in the fMRI experiment was much longer with extended inter-stimulus and inter-trial intervals (a total of over 1 hour compared with 15 minutes in the behavioral task), and this could potentially have influenced the delayed generalization effect[66].

These findings carry important translational implications. While both negative overgeneralization[1,4,67–69] and disrupted sleep[6,8,9,70–72] have been linked to anxiety and PTSD, few studies have examined how they interact to shape emotional memory and vulnerability to psychopathology [17]. Our results suggest that sleep, particularly NREM spindle activity, may serve as a protective mechanism. Sleep spindles may not only reduce negative amygdala-related overgeneralization but also promote adaptive generalization of positive content. Previous studies have shown that sleep spindle density increases following stress[73] and is elevated in individuals with PTSD[74–76]. While trauma may trigger these increases, our data suggests that spindles may also play a role in emotional recovery. Indeed, sleep spindles in individuals with PTSD have been linked to healthier sleep**,** reduced anxiety, and improved emotional recovery [73,77–79]. In contrast, lower spindle density has been associated with greater

vulnerability to sleep disturbances following stress[67]. Future work could build on these findings, testing whether enhancing sleep, and particularly sleep spindles, can reduce emotional overgeneralization in clinical populations. Promising interventions could include targeted memory reactivation (TMR), which reintroduces learning cues during sleep to reactivate specific memories [63,80–82], brain stimulation techniques designed to boost sleep spindle activity[83,84], and pharmacological treatments to elevate sleep spindle activity[85,86] Applying these methods to emotional generalization paradigms could help advance sleep-based interventions for anxiety and PTSD.

References

1. Dunsmoor, J. E. & Paz, R. Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biol Psychiatry* **78**, 336–343 (2015).

2. Guttman, N. & Kalish, H. I. Discriminability and stimulus generalization. *J Exp Psychol* **51**, 79–88 (1956).

3. Schechtman, E., Laufer, O. & Paz, R. Negative valence widens generalization of learning. *Journal of Neuroscience* **30**, 10460–10464 (2010).

4. Laufer, O., Israeli, D. & Paz, R. Behavioral and neural mechanisms of overgeneralization in anxiety. *Current Biology* **26**, 713–722 (2016).

5. Barlow, D. H., Allen, L. B. & Choate, M. L. Toward a Unified Treatment for Emotional Disorders – Republished Article. *Behav Ther* **47**, 838–853 (2016).

6. Cox, R. C. & Olatunji, B. O. A systematic review of sleep disturbance in anxiety and related disorders. *J Anxiety Disord* **37**, 104–129 (2016).

7. Breslau, N., Roth, T., Rosenthal, L. & Andreski, P. Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young Adults. *Biol Psychiatry* **39**, 411–418 (1996).

8. Shanahan, L., Copeland, W. E., Angold, A., Bondy, C. L. & Costello, E. J. Sleep problems predict and are predicted by generalized anxiety/depression and oppositional defiant disorder. *J Am Acad Child Adolesc Psychiatry* **53**, 550–558 (2014).

9. Ben Simon, E., Vallat, R., Barnes, C. M. & Walker, M. P. Sleep Loss and the Socio-Emotional Brain. *Trends Cogn Sci* **24**, 435–450 (2020).

10. Staner, L. Sleep and anxiety disorders. *Dialogues Clin Neurosci* **5**, 249–258 (2003).

11. Oudiette, D., Antony, J. W., Creery, J. D. & Paller, K. a. The Role of Memory Reactivation during Wakefulness and Sleep in Determining Which Memories Endure. *Journal of Neuroscience* **33**, 6672–6678 (2013).

12. Igloi, K., Gaggioni, G., Sterpenich, V. & Schwartz, S. A nap to recap or how reward regulates hippocampal-prefrontal memory networks during daytime sleep in humans. *Elife* **4**, 1–16 (2015).

13. Gruber, M. J., Ritchey, M., Wang, S. F., Doss, M. K. & Ranganath, C. Post-learning Hippocampal Dynamics Promote Preferential Retention of Rewarding Events. *Neuron* **89**, 1110–1120 (2016).

14. Fischer, S. & Born, J. Anticipated Reward Enhances Offline Learning During Sleep. *J Exp Psychol Learn Mem Cogn* **35**, 1586–1593 (2009).

15. Feld, G. B., Besedovsky, L., Kaida, K., Münte, T. F. & Born, J. Dopamine D2-like Receptor Activation Wipes Out Preferential Consolidation of High over Low Reward Memories during Human Sleep Gordon. *J Cogn Neurosci* **26**, 2310–2320 (2014).

16. Payne, J. D., Stickgold, R., Swanberg, K. & Kensinger, E. A. Sleep preferentially enhances memory for emotional components of scenes. *Psychol Sci* **19**, 781–788 (2008).

17. Davidson, P., Carlsson, I., Jönsson, P. & Johansson, M. Sleep and the generalization of fear learning. *J Sleep Res* **25**, 88–95 (2016).

18. Witkowski, S., Schechtman, E. & Paller, K. A. Examining sleep's role in memory generalization and specificity through the lens of targeted memory reactivation. *Curr Opin Behav Sci* **33**, 86–91 (2020).

19. Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D. & Walker, M. P. *Human Relational Memory Requires Time and Sleep*. www.pnas.org/cgi/content/full/ (2007).

20. Davidson, P., Carlsson, I., Jönsson, P. & Johansson, M. A more generalized fear response after a daytime nap. *Neurobiol Learn Mem* **151**, 18–27 (2018).

21. Talamini, L. M. & Juan, E. Sleep as a window to treat affective disorders. *Current Opinion in Behavioral Sciences* vol. 33 99–108 Preprint at https://doi.org/10.1016/j.cobeha.2020.02.002 (2020).

22. Vieira, T. F., Bottino, A., Laurentini, A. & De Simone, M. Detecting siblings in image pairs. *Visual Computer* **30**, 1333–1345 (2014).

23. Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).

24. Roach, G. D., Dawson, D. & Lamond, N. Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiol Int* **23**, 1379–1387 (2006).

25. Drummond, S. P. A. *et al.* The neural basis of the psychomotor vigilance task. *Sleep* **28**, 1059–1068 (2005).

26. Basner, M. & Dinges, D. F. Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss. *Sleep* **34**, 581–591 (2011).

27. Dinges, D. F. & Powell, J. W. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers* **17**, 652–655 (1985).

28. Hoddes, E., Zarcone, V. & Dement, W. Stanford Sleepiness Scale (SSS). *Enzyklopädie der Schlafmedizin* **1184**, (1972).

29. Beck, A. T., Steer, R. A. & Brown, G. Beck Depression Inventory–II. *APA PsycTests* (1996).

30. Marteau, T. M. & Bekker, H. The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology* **31**, 301–306 (1992).

31. Torrubia, R., Ávila, C., Moltó, J. & Caseras, X. The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Pers Individ Dif* **31**, 837–862 (2001).

32. CLONINGER, C. R. THE TRIDIMENSIONAL PERSONALITY QUESTIONNAIRE: U.S. NORMATIVE DATA. *Psychol Rep* **69**, 1047 (1991).

33. Carver, C. S. & White, T. L. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *J Pers Soc Psychol* **67**, 319–333 (1994).

34. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* **31**, 5–17 (2001).

35. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111–116 (2019).

36. Esteban, O. *et al.* fMRIPrep. *Software* (2018).

37. Krzysztof J Gorgolewski *et al.* Nipype. *Software* (2018).

38. Gorgolewski, K. *et al.* Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front Neuroinform* **5**, (2011).

39. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).

40. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging* **29**, 1310–1320 (2010).

41. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* **12**, 26–41 (2008).

42. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* **20**, 45–57 (2001).

43. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* (2009).

44. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage* **17**, 825–841 (2002).

45. Cox, R. W. & Hyde, J. S. Software tools for analysis and visualization of fMRI data. *NMR Biomed* **10**, 171–178 (1997).

46. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

47. Lanczos, C. *EVALUATION OF NOISY DATA*. ANAL. Ser. B* https://epubs.siam.org/terms-privacy (1964).

48. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).

49. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

50. Falach, R. *et al.* SleepEEGpy: a Python-based software integration package to organize preprocessing, analysis, and visualization of sleep EEG data. *Comput Biol Med* **192**, 110232 (2025).

51. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front Neurosci* (2013) doi:10.3389/fnins.2013.00267.

52. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M. & Robbins, K. A. The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front Neuroinform* **9**, 1–19 (2015).

53. Vallat, R. & Walker, M. P. An open-source, high-performance tool for automated sleep staging. *Elife* **10**, 70092 (2021).

54. Donoghue, T. *et al.* Parameterizing neural power spectra into periodic and aperiodic components. *Nat Neurosci* **23**, 1655–1665 (2020).

55. Lee, T. W., Girolami, M. & Sejnowski, T. J. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput* **11**, 417–441 (1999).

56. Rupp, T. L. *Psychomotor Vigilance Performance*. *Encyclopedia of Sleep* (Published by Elsevier Inc., 2013). doi:10.1016/B978-0-12-378610-4.00148-0.

57. Tibshirani, R. *Regression Shrinkage and Selection via the Lasso*. *Source: Journal of the Royal Statistical Society. Series B (Methodological)* vol. 58 (1996).

58. Andrillon, T. *et al.* Sleep Spindles in Humans: Insights from Intracranial EEG and Unit Recordings. *Journal of Neuroscience* **31**, 17821–17834 (2011).

59. Fernandez, L. M. J. & Lüthi, A. Sleep spindles: Mechanisms and functions. *Physiol Rev* **100**, 805–868 (2020).

60. Geva-Sagiv, M. *et al.* Augmenting hippocampal–prefrontal neuronal synchrony during sleep enhances memory consolidation in humans. *Nat Neurosci* **26**, 1100–1110 (2023).

61. Seibt, J. *et al.* Cortical dendritic activity correlates with spindle-rich oscillations during sleep in rodents. *Nat Commun* doi:10.1038/s41467-017-00735-w.

62. Schmidig, F. J. *et al.* A visual paired associate learning (vPAL) paradigm to study memory consolidation during sleep. *J Sleep Res* **33**, (2024).

63. Bar, E. *et al.* Local Targeted Memory Reactivation in Human Sleep. *Current Biology* **30**, (2020).

64. Motomura, Y. *et al.* Sleepiness induced by sleep-debt enhanced amygdala activity for subliminal signals of fear. *BMC Neurosci* **15**, (2014).

65. Stickgold, R. & Walker, M. P. Sleep-dependent memory triage: Evolving generalization through selective processing. *Nat Neurosci* **16**, 139–145 (2013).

66. Vassiliadis, P., Lete, A., Duque, J. & Derosiere, G. Reward timing matters in motor learning. *iScience* **25**, (2022).

67. Lissek, S. *et al.* Overgeneralization of Conditioned Fear as a Pathogenic Marker of Panic Disorder. *American Journal of Psychiatry* **167**, 47–55 (2010).

68. Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B. & Hermans, D. Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behav Ther* **46**, 561–582 (2015).

69. Lissek, S. *et al.* Generalized Anxiety Disorder Is Associated With Overgeneralization of Classically Conditioned Fear. *Biol Psychiatry* **75**, 909–915 (2014).

70. Murkar, A. L. A. & De Koninck, J. Consolidative mechanisms of emotional processing in REM sleep and PTSD. *Sleep Med Rev* (2018) doi:10.1016/j.smrv.2018.03.001.

71. Babson, K. A. & Feldner, M. T. Temporal relations between sleep problems and both traumatic event exposure and PTSD: A critical review of the empirical literature. *J Anxiety Disord* **24**, 1–15 (2010).

72. Vanderheyden, W. M. *et al.* Sleep alterations following exposure to stress predict fear-associated memory impairments in a rodent model of PTSD. *Exp Brain Res* **233**, 2335–2346 (2015).

73. Natraj, N. *et al.* Sleep Spindles Favor Emotion Regulation Over Memory Consolidation of Stressors in Posttraumatic Stress Disorder. *Biol Psychiatry Cogn Neurosci Neuroimaging* **8**, 899–908 (2023).

74. Van Der Heijden, A. C. *et al.* Sleep spindle dynamics suggest over-consolidation in posttraumatic stress disorder. *Sleep* **45**, (2022).

75. Wang, C. *et al.* Increased oscillatory frequency of sleep spindles in combat-exposed veteran men with posttraumatic stress disorder. *Sleep* **43**, (2020).

76. Denis, D. *et al.* Sleep Power Spectral Density and Spindles in PTSD and Their Relationship to Symptom Severity. *Front Psychiatry* **12**, (2021).

77. Azza, Y. *et al.* Sleep's role in updating aversive autobiographical memories. *Transl Psychiatry* **12**, (2022).

78. Dang-Vu, T. T. *et al.* Sleep spindles predict stress-related increases in sleep disturbances. *Front Hum Neurosci* **9**, (2015).

79. Natraj, N. & Richards, A. Sleep spindles, stress and PTSD: The state of the science and future directions. *Neurobiol Stress* **23**, (2023).

80. Cellini, N. & Capuozzo, A. Shaping memory consolidation via targeted memory reactivation during sleep. *Ann N Y Acad Sci* 1–20 (2018) doi:10.1111/nyas.13855.

81. Oudiette, D. & Paller, K. a. Upgrading the sleeping brain with targeted memory reactivation. *Trends Cogn Sci* **17**, 142–9 (2013).

82. Rasch, B., Büchel, C., Gais, S. & Born, J. Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science* **315**, 1426–9 (2007).

83. Lustenberger, C. *et al.* Feedback-Controlled Transcranial Alternating Current Stimulation Reveals a Functional Role of Sleep Spindles in Motor Memory Consolidation. *Current Biology* **26**, 2127–2136 (2016).

84. Choi, J., Kwon, M. & Jun, S. C. A Systematic Review of Closed-Loop Feedback Techniques in Sleep Studies—Related Issues and Future Directions. *Sensors* **20**, 2770 (2020).

85. Carbone, J. *et al.* The effect of zolpidem on targeted memory reactivation during sleep. *Learning & Memory* **28**, 307–318 (2021).

86. Zhang, J., Yetton, B., Whitehurst, L. N., Naji, M. & Mednick, S. C. The effect of zolpidem on memory consolidation over a night of sleep. *Sleep* **43**, (2020).