Chapter 2: Supplemental materials

The neural signature of emotional false memories

Methods

Questionnaires

The Becks Depression lnventory (BDI-II; Beck et al., 1996)) was administered to measure self-reported depressive symptoms. The BDI-II consists of 21 items, on a scale ranging from 0 to 3. Therefore the sum score can range between 0 and 63. The State-Trait Anxiety Inventory (STAI-trait) to assess individual trait anxiety (Spielberger et al., 1983). The STAI-trait consists of 20 statements, describing participants' general level of anxiety on a 4-point Likert scale. A total score can range from 20 to 80, where higher scores represent higher trait anxiety. The NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) was utilized to assess participants' personality traits, specifically levels of neuroticism. The NEO-FFI consists of 60 items, examining the Big Five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism), containing 12 items per domain. The Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) was administered to assess the general sleep quality of the previous month. Here the total score can range between 0 and 21, where higher scores represent poorer sleep quality. The Morningness–eveningness questionnaire (MEQ; Horne & Östberg, 1976) to check for chronotype, i.e. whether a person's circadian rhythm (biological clock) produces peak alertness in the morning, in the evening, or in between. Scores can range between 16 and 86, where scores of 41 and below indicate "evening types", scores between 42 and 58 indicate "intermediate types" and scored of 59 and above indicate "morning types". Lastly, the St. Mary's Hospital sleep questionnaire (SMH; Ellis et al., 1981) to assess the sleep quality of the preceding night. Here each question is evaluated separately. Of most interest were the questions on subjective sleep quality and sleep depth.

Analysis

fMRIprep pipeline

The following description of the fMRI pipeline is copied directly from the fmriprep output and represents all the individual steps in detail.

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin6Asyc) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following templates were selected for spatial normalization: FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym].

Functional data preprocessing

For each of the BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A B0-nonuniformity map (or fieldmap) was estimated based on a phase-difference map calculated with a dual-echo GRE (gradient-recall echo) sequence, processed with a custom workflow of SDCFlows inspired by the epidewarp.fsl script and further improvements in HCP Pipelines (Glasser et al. 2013). The fieldmap was then co-registered to the target EPI (echo-planar imaging) reference run and converted to a displacements field map (amenable to registration tools such as ANTs) with FSL's fugue and other SDCflows tools. Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al. 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox and Hyde 1997, RRID:SCR_005927). The BOLD time-series, were resampled to surfaces on the following spaces: fsaverage5. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled into a standard space, correspondingly generating the following spatially-normalized, preprocessed BOLD runs: MNI152NLin6Asym.

First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA, Pruim et al. 2015) was performed on the preprocessed BOLD on MNI space time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). Corresponding "non-aggresively" denoised runs were produced after such smoothing. Additionally, the "aggressive" noise-regressors were collected and placed in the corresponding confounds file. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.6.1 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Results

Behavioural results

	Critical intrusion	Hit rate
Total	0.18 ± 0.013	0.186 ± 0.016
Negative	0.174 ± 0.018	0.172 ± 0.015
Neutral	0.134 ± 0.017	0.154 ± 0.018
Positive	0.232 ± 0.023	0.232 ± 0.021

Table S1a. Recall results: Mean proportions and d' values for recall performance and each valence category with standard error of the mean.

	Amount intrusions	Amount of hits
Total	5.4 ± 0.382	55.8 ± 4.72
Negative	1.74 ± 0.18	17.2 ± 1.49
Neutral	1.34 ± 0.166	15.4 ± 1.85
Positive	2.32 ± 0.227	23.2 ± 2.12

Table S1b. Recall results: Average amount of intrusions and hits for recall performance and each valence category with standard error of the mean.

Recall – 38 participants

Participants freely recalled on average 72.8 (± 4.49 SE) words, ranging from 10 to 149 words. On average, participants recalled 55.8 words correct (hit rate = 0.186 ± .016 SE), and falsely recalled 5.4 critical lures (false alarm rate = .18 ± 0.013 SE).

The differences in recall performance between the three different valence categories largely supported the recognition results, by similarly indicating a general positive bias in our sample.

The amount of falsely recalled lures differed per valence category, $F(2,74) = 7.684$ $p < .001$, $\eta_p^2 = .105$. A follow-up pairwise comparisons revealed a higher false alarm rate for critical items in the positive compared to the neutral category, (.232 vs. .134; $t(37) = 3.91$, $p = .001$). Similarly, a difference in hit rates of the veridical memory was found, $F(2,74) = 12.318$, $p < .001$, $\eta_p^2 = .081$, where the follow-up pairwise comparisons revealed both a higher hit rate for positive items compared to negative items (.232 vs .174, $t(37) = 6.69$, $p = .002$) as well as a higher hit rate for positive items compared to neutral items (.232 vs .134, $t(37) = 4.45$, $p < .001$).

Also for the recall results, we calculated a memory bias score by taking the difference between the amount of positive and negative falsely recalled critical lures. A one-samples t-test revealed a significant positive bias (.58 vs no bias = 0, *t*(37) = 2.44, *p* = .0197, *d* = .396) for false memory as well as for the amount of correctly recalled studied items (7.79 vs no bias = 0, *t*(37)= 3.69, *p* < .001, *d* = .598). Again, a second difference score between positive and neutral items was calculated as well which revealed a significant positive bias for the critical lures (.97 vs no bias = 0, *t*(37)= 3.91, *p* <. 001, *d* = .635) as well as for the studied items (7.79 vs no bias = 0, *t*(37)= 4.45, *p* <. 001, *d* = .722).

Questionnaires

We correlated the recognition memory bias outcome with the mood questionnaires that were administered. There was no correlation between BDI-II (*r*(36) = .023, *p* = .893), STAI (*r*(36) = -.08, *p* = .616) nor neuroticism (NEO) scores (*r*(36) = .05, *p* = .776) and memory bias (pos-neg) of critical lures. Similarly, no correlation between BDI-II (*r*(36) = -.17, *p* = .3), STAI (*r*(36) = -.29, *p* = .076) nor neuroticism (NEO) scores (*r*(36) = -.26, *p* = .112) and memory bias (pos-neg) of veridical memory (old items) was found.

Next, we correlated the *recall* memory bias scores with the mood questionnaires that were administered. There was no correlation between BDI-II (*r*(36) = -0.23, *p* = .17), STAI (*r*(36) = -.24, *p* = .144) nor neuroticism (NEO) scores (*r*(36) = -.24, *p* = .139) and memory bias (pos-neg) of critical lures. Similarly, no correlation between BDI-II (*r*(36) = -.2, *p* = .225) nor STAI (*r*(36) = -.28, *p* = .089) and memory bias (pos-neg) of veridical memory (old items) was found. However, a significant negative correlation between neuroticism (NEO) scores and veridical memory (old items) was found (*r*(36) = -.34, *p* = .035). This suggests that those individuals who score higher on the neuroticism scale, show a lower positive bias.

Recall – 60 participants

As a control analysis, we performed the same analysis on the full sample of 60 participants available for the recall data.

Participants freely recalled on average 73 (± 24.7) words, ranging from 10 to 149 words. On average, participants recalled 47.7 words correct (hit rate = 0.159, SE = 0.013), and falsely recalled 4.9 critical lures (false alarm rate = 0.164, SE = 0.01).

The differences in recall performance between the three different valence categories largely supported the recognition results, by similarly indicating a general positive bias in our sample.

The amount of falsely recalled lures differed per valence category, *F*(2,118) = 6.576, *p* = 0.002, η_p^2= 0.062. A follow-up pairwise comparisons revealed a higher false alarm rate for critical items in the positive compared to the neutral category, (0.202 vs. 0.128; *t*(59) = 3.59, p = 0.002). Similarly, a difference in hit rates of the veridical memory was found, *F*(2,118) = 16.901, *p* < .001, η_p^2= 0.055, where the follow-up pairwise comparisons revealed both a higher hit rate for positive items compared to negative items (0.197 vs 0.144, *t*(59) = 4.77, *p* <.001) as well as a higher hit rate for positive items compared to neutral items (0.197 vs 0.137, *t*(59) = 4.95, *p* < .001).

Also for the recall results, we calculated a memory bias score by taking the difference between the amount of positive and negative falsely recalled critical lures. A one-samples t-test revealed no significant bias (0.4 vs no bias = 0, *t*(59)=1.89, *p* = 0.064, *d* = 0.244). However, a significant positive bias was found for the amount of correctly recalled studied items (5.367 vs no bias = 0, *t*(59)= 4.77, *p* < .001, *d* = 0.616). Again, a second difference score between positive and neutral items was calculated as well which revealed a significant positive bias for the critical lures (0.733 vs no bias = 0, *t*(59)= 3.59, *p* <. 001, *d* = 0.463) as well as for the studied items (6.07 vs no bias = 0, t(59)= 4.95, *p* <. 001, *d* = 0.639).

Imaging results

General FM contrast - recall:

Similar as with the recognition data, we contrasted the encoding trials related to a subsequent “correct memory” during recall with the trials related to a subsequent “false memory” of all three valence categories in a so-called general FM contrast. There were no significant clusters that survived whole brain correction, nor small volume correction (SVC) for the mPFC, amygdala and hippocampus in the average group mean nor negative group mean. Results did not change after including the complete sample of 60 participants.

Emotional FM contrast - recall

Also for the recall data, a second contrast was constructed to test the involvement of the mPFC and amygdala during emotional FM memory formation. Here, the activation during the negative and positive intrusion-related trials were compared to the neutral intrusion-related trials. Significant clusters in both left and right lateral occipital cortex were found (cluster-level FEW corrected; p < .05; see table S2). No significant clusters were found after including the complete sample of 60 participants.

Table S2: fMRI clusters of emotional FM contrast using recall data, group mean of emotional > neutral

	Cluster size	*p*	Z-value	x	y	z
Lateral occipital cortex (L)	70	< .001	4.25	48	-80	-12
Lateral occipital cortex (R)	35	.0335	4.06	-36	-86	-8

Clusters of voxels where activity is higher during the encoding of emotional (positive and negative) trials subsequently related to an intrusion vs neutral trials subsequently related to an intrusion during the free recall. For each cluster, the local maximum is reported. Cluster p-value is whole brain corrected at the cluster level (FWE, p<.05), or small-volume corrected (SVC) for an anatomical mask based on the Brainnetome atlas. All coordinates are in MNI space. L = left, R = right.

Eye tracking

During the encoding task, pupil dilation was monitored as a proxy for arousal (Bradley et al., 2008; Partala & Surakka, 2003). Pupil dilation was recorded with a sampling rate of 250 Hz using the high precision Eyelink 1000 plus eye tracker (SRResearch). To ensure there was no interference in pupil size due to word length and thus more brighter pixels on the screen, each word was flanked by meaningless symbols (dash "-" and pipe "|") to ensure each presented word had a similar length but was still easily readable. In addition, words were presented in a purple hue (RGB 190,66,105) on a light grey background (RGB 95,95,95) to decrease the difference in brightness between the word and background, while still remaining easily distinguishable from the background. Pupil dilation changes during the task was recorded in one pilot subject prior to the onset of the study, who did not speak Dutch, to ensure we would measure only pupil response to the word presentation without a response to the content. The pupil dilation did not show any average differences between the valence conditions.

For each individual, pupil size responses were calculated for each trial. First, the recorded data was downsampled to 50 Hz. Pupil size responses were baseline corrected by subtracting the average individual pupil diameter of 1 minute following the task where participants were shown a fixation cross on the same background as during the task. Linear interpolation was performed using in-house built scripts (Hermans et al., 2013) to remove eye blinks from the pupil data. The trial window of the pupil response was set to 3.25 seconds as this was the maximum same-sized window for each trial and ITI without crossing over into the next trial (750ms word presentation + 2.5 seconds ITI). A per participant average per valence condition was calculated and compared using a repeated measures ANOVA test.

It should be noted that the design of the current study was optimized to ensure maximal false memory using shorter presentation speeds (McDermott & Watson, 2001), therefore limiting the time window length to assess pupil dilation responses. Any later occurring changes could therefore not be detected. In addition, the time windows contained both the word presentation as well as the subsequent ITI. This switch in screen presentation could possibly confound the results. However, since this was identical for each trial and we were only interested in differences between valence conditions, it should not affect the overall interpretation. Lastly, actual dilation values used to compare between conditions are mostly arbitrary, as these depend on the resolution of the lens and the distance between the eye tracker and the pupil. These values should be regarded as relative changes in dilation.

No group differences in average pupil dilation between the valence categories within each subject were found, $F(2,104) = .925$, $p = .4$, $\eta_p^2 = .0001$. This suggests that arousal was not driving the main fMRI findings in the emotional FM contrast.

References

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, *78*(2), 490–498.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Ellis, B. W., Johns, M. W., Lancaster, R., Raptopoulos, P., Angelopoulos, N., & Priest, R. G. (1981). The St. Mary’s Hospital Sleep Questionnaire: A Study of Reliability. *Sleep*, 4(1), 93–97. https://doi.org/10.1093/sleep/4.1.93

Hermans, E. J., Henckens, M. J. A. G., Roelofs, K., & Fernández, G. (2013). Fear bradycardia and activation of the human periaqueductal grey. *NeuroImage*, *66*, 278–287. https://doi.org/10.1016/j.neuroimage.2012.10.063

Horne, J. A., & Östberg, O. (1976). *A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms*. 4, 97–110.

McDermott, K. B., & Watson, J. M. (2001). The Rise and Fall of False Recall: The Impact of Presentation Duration. *Journal of Memory and Language*, *45*(1), 160–176. https://doi.org/10.1006/jmla.2000.2771

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, *59*(1), 185–198. https://doi.org/10.1016/S1071-5819(03)00017-X

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg P., & Jacobs, G. A. (1983). *State-trait anxiety inventory for adults*. Palo Alto.