Supporting information

van Kesteren et al. 10.1073/pnas.0914892107

SI Experimental Procedures

Design and General Procedure. Participants were tested in a mixed factorial design with prior schema manipulation as between subjects factor (Fig. 1). One day before fMRI scanning, participants watched the first part of the movie (80 min) either in normal (consistent schema group) or temporally scrambled (inconsistent schema group) order. They were instructed to pay attention because they would get detailed questions about the movie on the next day. Procedures on the second day were equal for both groups and lasted ≈1.5 h. Participants were first tested on their memory for the first part of the movie by means of an item recognition memory test (test 1a; Fig. 1) and a test with open questions about the schematic content of the movie (test 1b). They were then placed in the MRI scanner and were instructed to watch the second part of the movie (15 min, in normal order) and again pay attention because they would get questions about this part as well. Subsequently, while still in the scanner, they were asked to complete an item recognition memory test (test 2a) and a multiple choice test on the content of the movie (test 2b). After this, fMRI scanning resumed with a resting period of the same length as the final part of the movie. During this rest period participants were instructed to lie still, close their eyes, think of nothing in particular, and try not to fall asleep. Finally, participants completed another item recognition (test 3a) and multiple choice memory test (test 3b). Functional scans were obtained only during watching of the movie and the rest period.

We controlled for two potential confounds that could influence performance. First, to control for consolidation time, participants were always tested in the MRI scanner 21–26 h (average 23.4 h; no significant group difference) after viewing the first part of the movie. Second, to verify that participants could optimally perceive the sound of the movie against a background of scanner noise, we employed a quieter EPI sequence (for details, see *fMRI Scanning Parameters*), and supplied the participants with earplugs and headphones (Commander XG, Magnetic Resonance Technology). Before starting the movie, we performed a sound test to verify whether participants could easily discriminate movie-related sounds from the scanner noise. After the experiment, participants were once more asked whether they had had difficulty hearing the movie (on a scale from 1 to 5), and they reported little difficulty (1.9). This score did not differ between groups, and subsequent memory performance was not significantly related to these ratings.

Movie and Schema Manipulation. To manipulate schema knowledge while not altering perceptual input, we temporally scrambled the first part of a movie by using Windows Movie Maker version 5.1 (Microsoft) using scenes of minimally 20 seconds and maximally 144 seconds of length. The movie that was used was named *Go* (Banner Entertainment, Columbia Pictures and Saratoga Entertainment, 1999). This movie was chosen because it contains three different story lines that merge together in the last 15 min (shown during scanning). When properly understood, the first part therefore provides a "schema" that facilitates integration of the last 15 min into a coherent story. We chose temporal scrambling of the first part of the movie as a method of manipulating schema knowledge over simply not showing it to avoid group differences in familiarity of the scenes and actors in the movie. No subtitles were shown in any of the movies.

Memory Tests and Analyses. Before MRI scanning, memory for the first part of the movie (shown either in scrambled or correct order) was tested by using an item recognition memory test and open questions. The first item recognition memory test (test 1a) consisted of 60 still frames. Of these, 30 were extracted from the movie that was shown, and 30 were taken from other movies similar in setting, actors, or type of scenery. The stimuli were all equal in size. Contrast and luminance was equalized by using Photoshop 7.0 (Adobe). Pictures were presented for 500 ms by using Presentation 10.2 (Neurobehavioral Systems). Participants were instructed to indicate whether the still frame was taken from the movie they saw the day before (yes or no). Performance on this test was expressed as the percentage of hits minus the percentage of false alarms. The 20 open questions (test 1b) were constructed to reflect comprehension of the storyline of the movie. Names of characters and objects were explicitly named and not explained and the questions did not contain any clues related to events in the second part of the movie. Participants were instructed to write the answers to the questions on a paper answer sheet and to recall as much as they could remember. Answers to these questions were scored, blind for condition, as either correct or incorrect, and performance was expressed as a percentage of correct answers.

Memory of the final part of the movie, shown in the correct order for all participants during scanning, was tested inside the scanner by using item recognition memory tests and multiple choice questions probing content-related knowledge. The order of the tests was counterbalanced across subjects, and they were performed directly after the movie (tests 2a and 2b), and after a 15-min resting state fMRI scan (tests 3a and 3b). The item recognition memory tests (tests 2a and 3a) had the same setup as described above, but contained only 40 still frames. Half of these were extracted from the final part of the movie, and the other half were again taken from other movies. Since these questions had to be answered in the MR scanner, multiple choice questionnaires (tests 2b and 3b) were used to test schematic memory of the movie. These multiple choice questionnaires consisted of 35 questions each. The questions were created in accordance with the constraints described in a previous study testing memory of movies (1). All questions targeted distinct events and all questions together covered the whole content of the second part of the movie. With 70 questions in total, this means that approximately every 13 seconds of the movie was covered by a question. Additionally, we applied constraints particularly related to this study: Questions could not be answered based only on the first part of the movie, but could be answered based only on the second part of the movie. Each multiple choice question was displayed together with a still frame of the movie and three answer options by using Presentation 10.2. Participants were instructed to indicate the answer (a, b, or c). Questions were divided over the two tests (tests 2b and 3b) in such a way that each test covered events from the entire duration of the second part of the movie. Performance on the multiple choice tests was expressed as the percentage of correct answers.

All performance data were analyzed using SPSS 15.0. The memory tests on the first part of the movie (tests 1a and 1b) were analyzed using independent samples t tests with GROUP (consistent versus inconsistent schema) as between-subjects factor. Statistical analyses on the tests on the final part of the movie (tests 2a, 2b, 3a, and 3b) were performed using repeated-measures ANOVAs with GROUP (consistent versus inconsistent schema) as between-subjects factor and TIME (after movie versus after rest) as within-subject factor. Alpha was set at 0.05 throughout.

MRI Scanning Parameters. Participants were scanned using a 1.5 Tesla Siemens Magnetom Avanto system equipped with an eight-

channel phased array head coil (MRI Devices). For BOLD fMRI images, we used a T2* weighted gradient echo EPI sequence with the following parameters: repetition time (TR), 2.31 s; echo time (TE), 35 ms; 34 slices, ascending slice order; 3.5-mm slice thickness; 0.35 mm slice gap; matrix size: 64 × 64; field of vision (FOV), 212 × 212 mm; flip angle, 90°; voxel size, 3.3 × 3.3 × 3.85. Slices were angulated in an oblique axial manner to reach whole brain coverage. To reduce the gradient acoustic noise, we used a relatively low readout bandwidth of 1,396 Hz per pixel, which halves the amplitude of the readout gradient (2), in combination with a generalized autocalibrating partially parallel acquisitions (GRAPPA) parallel acceleration factor of 2 (3). To ensure reaching a steady-state condition and to let participants become accustomed to the scanner noise, the first 11 scans were discarded. Additionally, T1 weighted anatomical scans at 1 mm isotropic resolution were acquired by using an MP-RAGE sequence with TR of 2,250 ms, time for inversion (TI) of 850 ms, flip angle of 15°, and FOV of 256 × 256 × 176 mm. Acquisition time was again reduced by using GRAPPA with acceleration factor 2 and 24 reference lines.

Interregional Partial Correlations Analysis. We implemented interregional partial correlation analyses as follows. First, functional images were parcellated anatomically based on the Automatic Anatomical Labeling (AAL) template (4), which consists of 116 regions. We merged right and left hippocampi into a single area reflecting the bilateral hippocampus to prevent the known strong interhemispheric correlation of both hippocampi (5) from reducing partial correlations with other regions. Time courses of the remaining 115 regions were calculated by averaging the signal over constituent voxels. Then, a 115 × 115 partial correlation matrix was calculated, which contained, in each of the off-diagonal cells, pairwise interregional correlation coefficients after partialling out any variance explained by time courses of any of the other regions. Such partial correlation matrices were calculated for each participant and each condition (natural viewing of the movie and rest). Subsequently, we defined the vmPFC as a set of eight (four bilateral) regions within the AAL (orbital part of the middle frontal gyrus, orbital part of the superior frontal gyrus, medial orbital part of the superior frontal gyrus, and gyrus rectus; Fig. 3*A*) located around the area previously found to increase its involvement in memory retrieval as a function of remoteness of memory (6).

To allow valid inferences on group differences, we applied a Fisher's z transformation to all individual partial correlation coefficients. Resulting values were then analyzed using SPSS 15.0 by applying a repeated measures ANOVA with TIME (encoding versus postencoding rest) and AREA (eight different regions) as within-subjects factors and GROUP (consistent versus inconsistent schema) as a between-subjects factor. Moreover, we investigated whether performance measures of prior schema strength was predictive of interregional partial correlations by using bivariate Pearson's correlations. Finally, to investigate to what extent any group differences in connectivity between hippocampus and vmPFC would be specific, we compared connectivity of the hippocampus with the vmPFC to connectivity with a control pathway. For this purpose, we selected an equal number of regions within the extrastriate/inferotemporal ventral visual processing stream (bilateral lingual gyrus, inferior temporal gyrus, fusiform gyrus, and parahippocampal gyrus; Fig. 3*B*), shown to be involved in object identification (7). This aspect of perceptual processing is likely independent of schema knowledge and, therefore, unaffected by our prior schema manipulation. Partial correlation coefficients were averaged over the eight subregions for both pathways and then entered into a repeated measures ANOVA with TIME (encoding versus postencoding rest) and PATHWAY (vmPFC and ventral visual) as within-subjects factors, and GROUP (consistent versus inconsistent schema) as between-subjects factor. Alpha was set at 0.05 throughout.

Intersubject Synchronization Analysis. Calculations were implemented in Matlab 7.5 (Mathworks) by using custom scripts combined with cluster-based nonparametric randomization tests as applied in the Matlab toolbox FieldTrip (fieldtrip.fcdonders.nl; Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands), a Matlab toolbox for analysis of biological data. First, low frequency confounds (0.01 Hz cut-off discrete cosine transform high pass filter) and movement-correlated (six parameter rigid body transformation-derived translations and rotations) signals were removed from all subjects' functional scan series. Second, data were masked by using a MNI152 space gray matter tissue probability map (see International Consortium for Brain Mapping: www.loni.ucla.edu/ICBM) with a probability threshold of 0.45, allowing for extraction of the nonselective component (8); i.e., the mean BOLD signal time course over all gray matter voxels of a single participant. Fourth, the nonselective component was low-pass filtered (moving average of three timepoints) to remove supra-BOLD frequencies, and this signal was again regressed out of each voxel's time course, resulting in filtered 3D time series data for each participant, thus removing the contribution of global signal fluctuations to voxel-wise time-series.

ISS main effects were tested as follows: For each voxel, each subject's time series was correlated with the mean of all other subjects' time series in the same voxel, and this correlation was expressed in a *t* statistic, resulting in one ISS map for each participant. Subsequently, a one-sample *t* test was run across these subject-specific ISS maps. To accommodate dependencies within these *t* statistics, nonparametric randomization tests were applied to validly test the null hypothesis of zero ISS across the group. Specifically, this procedure tests the null hypothesis (H0) that the time series data of a random set of subjects can be sign-permuted without affecting ISS *t* statistic across the group. This sign permutation procedure destroys any synchronization of time series across subjects without affecting either the autocorrelational structure of the signal or dependencies between the subject-specific ISS maps and can therefore be used to estimate a null distribution. To achieve an accurate approximation of this null distribution, 1,000 randomizations (limited by computational resources) were performed, and the null distribution was pooled across voxels. This null distribution was used to threshold the *t* maps at a $P < 0.001$, uncorrected. Subsequently, weights (i.e., the sum of all *t* values) were calculated for each cluster of adjacent voxels. The same procedure was applied to all 1,000 randomization-derived *t* maps, thus resulting in a null distribution of cluster weights. Clusters within the nonrandomized ISS *t* map exceeding a threshold based on the 5% largest clusters within all randomizations were considered significant. This method implements an alpha = 0.05, one-sided, test for cluster significance corrected for multiple comparisons (9) at the whole-brain level.

Group differences in ISS were tested by using a similar method, now testing H0 that assignments of participants to groups could be permuted without changing the difference in ISS between groups. For each voxel, each participant's time course was correlated with the mean of all time courses of other participants within the same group (consistent versus inconsistent schema). These correlations were expressed in a *t* statistic, resulting in one ISS map for each participant. Group differences across these maps were tested by using voxel-wise independent *t* tests. Voxel and cluster level null distributions were obtained by randomizing group assignments and repeating these calculations (1,000 random permutations). The voxel level null distribution was used to threshold the group-level *t* map at $P < 0.001$, uncorrected. Cluster-level alpha for this group comparison was set at 0.05, corrected for our three a priori reduced search regions (i.e., voxels in aforementioned bilateral hippocampus, vmPFC, or extrastriate/inferotemporal ventral stream regions of interest; Fig. 3 *A* and *B*) by restricting all calculations (including randomizations) to voxels comprising the search regions.

1. Furman O, Dorfman N, Hasson U, Davachi L, Dudai Y (2007) They saw a movie: Long-term memory for an extended audiovisual narrative. *Learn Mem* 14: 457–467.
2. de Zwart JA, van Gelderen P, Golay X, Ikonomidou VN, Duyn JH (2006) Accelerated parallel imaging for functional imaging of the human brain. *NMR Biomed* 19: 342–351.
3. Griswold MA, et al. (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 47:1202–1210.
4. Salvador R, et al. (2005) Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb Cortex* 15:1332–1342.
5. Stein T, et al. (2000) Functional connectivity in the thalamus and hippocampus studied with functional MR imaging. *AJNR Am J Neuroradiol* 21:1397–1401.
6. Takashima A, et al. (2006) Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proc Natl Acad Sci USA* 103:756–761.
7. Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15:20–25.
8. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303:1634–1640.
9. Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.

Table S1. Memory performance on tests regarding the final 15 min of the movie for the two experimental groups

	Test 2		Test 3	
Groups	a: Item recognition memory	b: Content-related memory	a: Item recognition memory	b: Content-related memory
Consistent schema	0.79 (0.12)	0.72 (0.11)	0.81 (0.10)	0.70 (0.11)
Inconsistent schema	0.82 (0.13)	0.69 (0.10)	0.77 (0.13)	0.68 (0.11)

For item recognition memory tests (2a and 3a), values represent means (and SD) of proportions hits minus proportions false alarms. For content-related multiple choice tests (2b and 3b), mean proportions of correct responses (and SD) are shown. SD, standard deviation.