nature neuroscience

Memory hierarchies map onto the hippocampal long axis in humans

Silvy H P Collin, Branka Milivojevic & Christian F Doeller

Memories, similar to the internal representation of space, can be recalled at different resolutions ranging from detailed events to more comprehensive, multi-event narratives. Single-cell recordings in rodents have suggested that different spatial scales are represented as a gradient along the hippocampal axis. We found that a similar organization holds for human episodic memory: memory representations systematically vary in scale along the hippocampal long axis, which may enable the formation of mnemonic hierarchies.

The hippocampus, a region that is critical for memory as well as internal representation of space[1], differs in structure and function along its long axis (dorsal-ventral in rodents and posterior-anterior in humans)[2–4]. More specifically, encoded space increases in scale along the long axis of the hippocampus, as evidenced by an increase in place-field size in the rodent hippocampus from dorsal to ventral hippocampus[5]. Such a gradient may provide a mechanism that enables multiple scales of episodic memories, ranging from detailed individual events to more comprehensive multi-event narratives, to be concurrently represented along the long axis of the hippocampus and may underpin hierarchical memory representations[6,7]. However, whether a similar neural gradient indeed underlies organization of episodic memories in humans remains unclear[5,8,9].

To determine whether memory scale is differently represented along the hippocampal long axis, we examined the emergence of new, multi-event representations by combining fMRI and multivoxel pattern analysis with a narrative-insight task (**Fig. 1** and **Supplementary Fig. 1**) (ref. 10). We used realistic, life-like videos showing individual events that could be integrated into narratives[10] and thus experimentally simulate processes involved in episodic memory formation. We gradually built up four separate narratives by initially presenting seemingly unrelated events (A, B, C and X; **Fig. 1**), before sequentially introducing two different 'linking' events (first L1, then L2), which provided insight into direct (A-B, and B-C, respectively) and inferred (A-C) event associations in the narratives. Memories of such narratives can be recalled at different resolutions ranging from detailed events to more comprehensive narratives. Thus, although all of these events were part of the same narrative, we hypothesized that the representation of such a multi-event narrative may differ along the long axis of the hippocampus depending on the 'narrative scale' of the ensuing representation. We propose that there are different resolutions in which these narratives can be represented, which we refer to as small-, medium- and large-scale networks (**Fig. 2a**). Here, a representation at the largest narrative scale would reflect the complete narrative (large-scale network), including all possible event associations between both directly (A-B and B-C) and indirectly (A-C) related events in phase 3 of the task, and would be represented in the anterior hippocampus. At the smallest narrative scale, the representations would contain only individual event-pair associations, which would, in this context, reflect only the most recent directly associated event pairs (small-scale network) and would be represented in the posterior hippocampus. An intermediately scaled representation might concurrently contain multiple event-pair associations in phase 3 of the task, but would not bridge between them (medium-scale network) and would be represented in the mid-portion of the hippocampus.

We employed representational similarity analysis (RSA), which uses correlations of across-voxel activation patterns as a proxy of neural similarity, to quantify memory representations along the long axis of the hippocampus. We split the hippocampus into three regions of interest (ROIs): an anterior portion, a mid-portion and a posterior portion (Online Methods), and computed correlation coefficients between across-voxel activation patterns in each ROI for event pairs of interest (A-B, B-C and A-C) in each of the three experimental phases separately and averaged effects across the four runs (Online Methods and **Supplementary Fig. 1**). B-X served as a control (**Supplementary Fig. 2**). We tested the predicted interaction effect using a reduced a priori model, referred to as mnemonic-scaling contrast. Using this contrast (with contrast weights for small scale, medium scale and large scale: posterior: 2 −1 −1, mid-portion: −1 2 −1, anterior: −1 −1 2; Online Methods and **Fig. 2a**) in a repeated-measure ANOVA, with narrative scale (small, medium and large), ROI (posterior, mid-portion and anterior) and hemisphere (left and right) as within-subject factors, we found a significant interaction effect between narrative scale and ROI ($F_{1,28} = 11$, $P < 0.01$; **Fig. 2**). Thus, the small-scale network was indeed represented in the posterior portion, the medium-scale network in the mid-portion and the large-scale network in the anterior portion of the hippocampus. There was no difference between hemispheres (**Supplementary Fig. 3**). Additional control analyses revealed that these results were unlikely to be a result of an increasing amount of information or passing of time throughout the task (**Supplementary Fig. 2**) or BOLD-signal fluctuations (**Supplementary Fig. 4**).

Post hoc analyses revealed evidence for the smallest scale of narrative representation only in the posterior portion of the hippocampus (**Fig. 2** and **Supplementary Fig. 3**), which is consistent with our previous report[10]. In addition, the medium-scale network was only represented in the mid-portion of the hippocampus, suggesting co-existence of the two directly integrated event-pair associations, without bridging across the inferred associations[11,12]. Finally, we found evidence for large-scale networks only in the anterior

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands. Correspondence should be addressed to S.H.P.C. (s.collin@donders.ru.nl) or C.F.D. (christian.doeller@donders.ru.nl).

Received 23 June; accepted 15 September; published online 19 October 2015; doi:10.1038/nn.4138

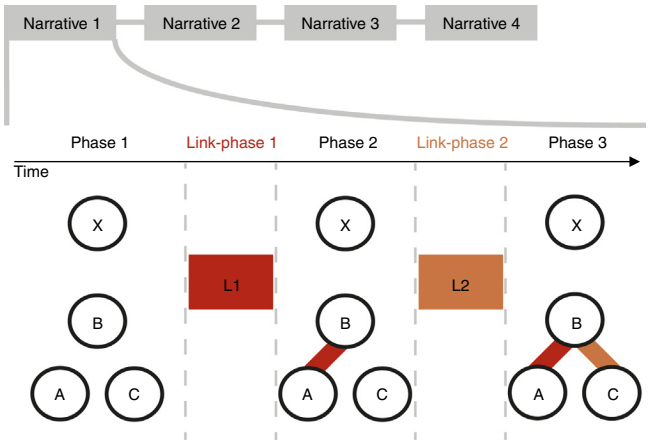

Figure 1 Schematic overview of the narrative-insight task. Participants were presented with animated videos of life-like events. Videos of one narrative were presented in five phases. Phases 1, 2 and 3 contained events A, B and C, and control-event X. Link-phases 1 and 2 were interleaved, during which events L1 and L2 were presented and provided links between events A and B and events B and C, respectively. Thus, some events were directly linked (first A and B via L1, then B and C via L2), whereas other associations had to be inferred (A and C were associated via their shared association with B). Participants performed four runs; in each run, a different narrative was presented (narratives 1–4).

portion of the hippocampus, which included the two directly integrated event-pair associations and a bridge across the inferred association, which were not directly experienced. The crucial difference between the large-scale network and the medium-scale network was that the former predicted this inferred association between A and C, an effect that was restricted to the anterior hippocampus (**Supplementary Fig. 5**).

These data suggest that multi-event narratives are simultaneously represented at multiple narrative scales along the hippocampal long axis, but is this gradient relevant for behavior? Following scanning, participants were asked to report all of the narratives they had seen during the experiment. Some participants remembered four unified narratives (12 participants), whereas others failed to integrate some events into unified narratives (**Supplementary Fig. 6**). We used this difference in performance to split the participants into full-integration and partial-integration groups, and investigated whether the gradient was expressed differently between those groups (Online Methods). We observed that the long-axis gradient was only present in the full-integration group (group × scale × ROI interaction: $F_{4,104} = 2.7$, $P < 0.05$; **Supplementary Figs. 6** and **7**). These results suggest that representing event associations at multiple scales simultaneously supports memory recall of accurate integrated narratives.

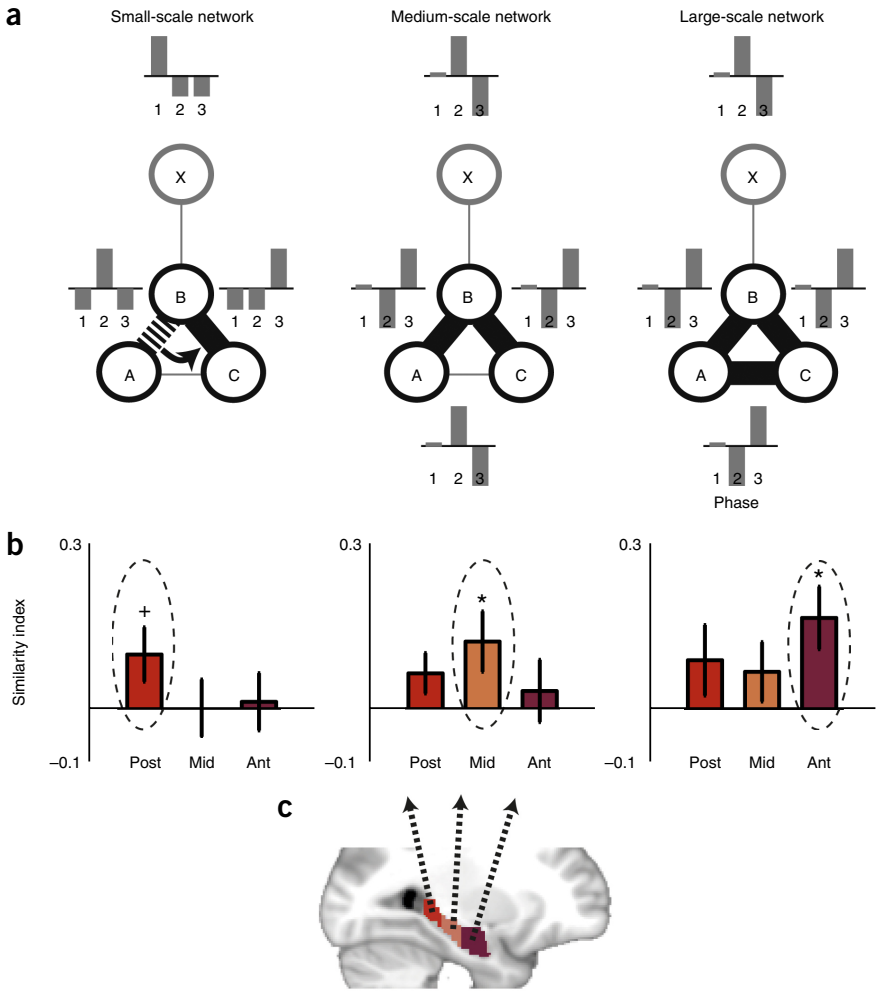

Figure 2 Increasing memory scale along the hippocampal long axis. (**a**) Depiction of the three network scales. Contrast weights for the small-scale network, separately for phases 1, 2 and 3 (A-B: –1 2 –1/B-C: –1 –1 2/B-X: 2 –1 –1), the medium-scale network for phases 2 and 3 (A-B: –1 1/B-C: –1 1/A-C: 1 –1/B-X: 1 –1) and the large-scale network for phases 2 and 3 (A-B: –1 1/B-C: –1 1/A-C: –1 1/B-X: 3 –3) are shown. Note that these three networks do not correspond to experimental phases. We predict a mnemonic-scaling contrast that reflects an interaction between narrative scale and hippocampal ROI: (small, medium, large: pos: 2 –1 –1, mid: –1 2 –1, ant: –1 –1 2). (**b**) Model evidence (parameter estimates) of left and right hippocampus (mean ± s.e.m.) separately for the three ROIs and scales ($N = 29$). Small-scale network: a representation of the narrative sensitive only to the directly linked events immediately after linking (link between event A and B replaced by re-linking B to C later in time) was observed in the posterior hippocampus only (posterior: $F_{1,28} = 4.1$, approaching significance at $^{+}P = 0.053$; mid-portion: $F_{1,28} = 0.002$, $P = 0.96$; anterior: $F_{1,28} = 0.05$, $P = 0.82$). Medium-scale network: increase in neural similarity between both pairs of directly linked events simultaneously, relative to inferred link and control-event X, was present only in the mid-hippocampus (mid-portion: $F_{1,28} = 4.99$, $P < 0.05$; anterior: $F_{1,28} = 0.34$, $P = 0.56$; posterior: $F_{1,28} = 3.14$, $P = 0.09$). Large-scale network: the anterior hippocampus showed a selective increase in neural similarity between all three events (A-B, B-C, A-C) in each narrative, in contrast with X (anterior: $F_{1,28} = 8.6$, $P < 0.01$; posterior: $F_{1,28} = 1.96$, $P = 0.17$; mid-portion: $F_{1,28} = 1.63$, $P = 0.21$). $^{*}P < 0.05$. See also **Supplementary Table 1**. (**c**) Depiction of the three ROIs.

In sum, our results provide, to the best of our knowledge, the first evidence in humans that event associations are represented as memory hierarchies with multiple associative networks increasing in scale along the hippocampal long axis: small-, medium- and large-scale networks were represented in posterior, mid-portion and anterior hippocampus, respectively. Moreover, this hierarchical memory gradient was related to accurate recall or construction of integrated narratives. These results demonstrate that a mnemonic gradient underlies the organization of human episodic memory, which may relate to the gradient of the scale of encoded space[5].

Previous research showed involvement of the mid-portion and anterior hippocampus during inference that could be driven by bridging between unseen associations or the formation of more complex networks, potentially with more complex networks represented anteriorly and less complex networks represented in mid-portion[11,13] (**Supplementary Fig. 8**). One possibility is that the large-scale network effect in the anterior hippocampus reflects such a bridging function. Alternatively, it might index a complete and integrated representation akin to a relational memory network[1,7] or cognitive map[14]. Notably, these two explanations are not mutually exclusive; making inferences about unseen connections is crucial for the creation of large-scale mnemonic networks. This dovetails with previous research that showed a functional dissociation within the rodent hippocampus: ventral hippocampal neurons in rats represent global event context[15] while neurons in dorsal hippocampus encode more specific event information[15,16].

Our results accord with previous findings on the role of the hippocampus in memory generalization. For example, new conceptual knowledge[17] and the formation of schemas[18] require a certain degree of abstraction from individual events and seem to preferentially involve anterior hippocampus. In contrast, smaller networks consisting of few elements seem to engage more posterior regions[10]. There are many different proposals about hippocampal long axis functional dissociation (**Supplementary Fig. 9**). The hippocampal memory gradient may provide a mechanism that enables multiple scales of episodic memories, ranging from individual events to more comprehensive multi-event narratives, to be represented by the brain simultaneously as different levels of mnemonic hierarchies[6,7]. An interesting question for future research is whether each level of the memory gradient is specific to one scale of the narrative representation or, alternatively, whether anterior hippocampal subregions extend the more posterior, lower-scale narrative representations.

Although it is clear from our data that this scaled coding mechanism supports memory performance, the question of how the hierarchical representation relates to performance remains unanswered. One possibility is that overall memory benefits from both maintaining the ability to individuate memories of separate events[19] and to integrate multiple experiences for the purpose of generalization or abstraction of knowledge[17,20]. Representing events at multiple scales may provide an effective way of providing a context or schema for individual events, which is known to improve memory performance and may protect against loss of individual event details.

In conclusion, we provide evidence for a mnemonic gradient along the hippocampal longitudinal axis, which enables the concurrent representation of multiple memories in hierarchies, a finding with potential implications for the classroom.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors would like to thank A. Backus, S. Bosch and L. Deuker for useful discussions, N. Müller for help during stimulus development, and K. Georgelou for help during behavioral pilot work. This work was supported by the European Research Council (ERC-StG RECONTEXT 261177) and the Netherlands Organization for Scientific Research (NWO-Vidi 452-12-009).

AUTHOR CONTRIBUTIONS

S.H.P.C. designed the experiment, collected the data, analyzed the data and wrote the manuscript. B.M. designed the experiment, supervised data analysis and wrote the manuscript. C.F.D. conceived and designed the experiment, supervised the project, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M. & Tanila, H. *Neuron* **23**, 209–226 (1999).
2. Fanselow, M.S. & Dong, H. *Neuron* **65**, 7–19 (2010).
3. Ritchey, M., Montchal, M., Yonelinas, A. & Ranganath, C. *Elife* **4**, 1–19 (2015).
4. Liang, J.C., Wagner, A.D. & Preston, A.R. *Cereb. Cortex* **23**, 80–96 (2013).
5. Strange, B.A., Witter, M.P., Lein, E.S. & Moser, E.I. *Nat. Rev. Neurosci.* **15**, 655–669 (2014).
6. Horner, A.J. & Burgess, N. *Curr. Biol.* **24**, 988–992 (2014).
7. McKenzie, S. *et al. Neuron* **83**, 202–215 (2014).
8. Poppenk, J., Evensmoen, H.R., Moscovitch, M. & Nadel, L. *Trends Cogn. Sci.* **17**, 230–240 (2013).
9. Milivojevic, B. & Doeller, C.F. *J. Exp. Psychol.* **142**, 1231–1241 (2013).
10. Milivojevic, B., Vicente-Grabovetsky, A. & Doeller, C.F. *Curr. Biol.* **25**, 821–830 (2015).
11. Shohamy, D. & Wagner, A. *Neuron* **60**, 378–389 (2008).
12. Zeithamova, D., Dominick, A.L. & Preston, A.R. *Neuron* **75**, 168–179 (2012).
13. Staresina, B.P. & Davachi, L. *Neuron* **63**, 267–276 (2009).
14. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford University Press, 1978).
15. Komorowski, R.W. *et al. J. Neurosci.* **33**, 8079–8087 (2013).
16. Komorowski, R.W., Manns, J.R. & Eichenbaum, H. *J. Neurosci.* **29**, 9918–9929 (2009).
17. Kumaran, D., Summerfield, J.J., Hassabis, D. & Maguire, E. *Neuron* **63**, 889–901 (2009).
18. van Kesteren, M.T., Fernández, G., Norris, D.G. & Hermans, E.J. *Proc. Natl. Acad. Sci. USA* **107**, 7550–7555 (2010).
19. Staresina, B.P., Fell, J., Do Lam, A.T.A., Axmacher, N. & Henson, R.N. *Nat. Neurosci.* **15**, 1167–1173 (2012).
20. Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B. & Botvinick, M.M. *Nat. Neurosci.* **16**, 486–492 (2013).

ONLINE METHODS

Participants. 35 healthy students from the Radboud University campus (17 males) participated in this study. All participants were right-handed. Six participants were excluded from further analyses: four due to excessive head motion (>2mm); and further two due to technical problems leading to incomplete data sets. The final group consisted of 29 participants (14 males, aged 18–33 years, mean age 22.8) who all had normal or corrected-to-normal vision. All participants gave written informed consent. The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, the Netherlands).

Study design. *Narrative-insight task (scanning).* First, participants completed the narrative-insight task[10] in the MRI scanner. Stimuli consisted of four animated narratives generated using The Sims 3 game (http://www.thesims3.com). Each narrative consisted of five separate events of 5 s in duration. Notably, participants were presented with only one narrative per scanning run, with four scanning runs in total. In addition, each run contained one control event (event X). New information was introduced twice in each run so that the events were gradually integrated into one narrative (**Fig. 1**).

Each run consisted of five different phases (see **Fig. 1**). During phase 1, 2 and 3, participants saw 4 seemingly unrelated events: A (for example, a man eating soup), B (for example, a child playing on the floor), C (for example, a man watching TV) and X (for example, a man cooking). Between phase 1 and 2, a new event was presented (event L1), which linked two of the seemingly unrelated events: A and B (for example, the man from event A brings the child from event B to bed). A similar linking event (event L2) was presented between phases 2 and 3, and linked another two of the seemingly unrelated events: B and C (for example, the man from event C gives a bottle to the child from event B). Therefore, by the third phase, it was clear to the participants that events A, B and C were all part of the same narrative with direct links between A and B, and B and C, and an indirect link between A and C (via B). The content of L1 and L2 was counterbalanced between subjects. Since the X-event was never linked to any of the other events, it served as a control.

A pseudo-randomized order was used to present each event in phase 1, 2 and 3, that is, all four events were shown before an event was repeated and an event was never shown twice in a row; this was done to avoid temporal confounds in the representational similarity analysis (RSA). Each event was shown six times per phase, with an inter-trial interval of 5.3 s on average (1, 4 or 11 s, uniform distribution). Thus, each original event (A, B, C and X) was presented 18 times in the task in total. The link-events were repeated six times as well, interspersed with inter-trial intervals of 12 s on average (6, 12 or 18 s, uniform distribution), see **Supplementary Figure 1** for a schematic overview of the task structure. Participants finished the entire task structure for one narrative (Phase 1, link-event 1, Phase 2, link-event 2, Phase 3) before continuing with the same task structure for the following narrative.

Additionally, there was a target event (a girl riding a scooter), not related to any of the narratives, which was presented at random moments during the experiment (in 8% of all trials). Participants were instructed to press a button whenever they saw this target event. The purpose of the target event was to make sure that participants were attending the stimuli. Before the first run, participants were presented with an example narrative (shown in **Fig. 1**). All example events were shown twice following the same procedure as in the actual task narratives to ensure that participants understood the task well.

The task was presented using Presentation software (Neurobehavioral Systems, version 16.2). Afterwards, participants were taken out of the MRI scanner and completed a narrative-recall task, in which they were asked to write down concisely all narratives they have seen during the task (20 min). In this within-subject design, no blinding procedures were applied for data collection and analysis.

Additional behavioral experiment. Since behavioral testing was done after the Narrative-insight task was completed, we ran a separate behavioral experiment to test participants' memory performance immediately after the first link (see **Supplementary Fig. 10** for more details).

MRI acquisition. All images were acquired using a 3T TIM Trio scanner equipped with a 32 channel head coil (Siemens, Erlangen, Germany). For the functional images, a 3D Echo Planar Imaging (EPI) sequence was used[21], with the following parameters: 56 axial slices, voxel size 1.5 mm isotropic, TR = 1,888 ms, TE = 26 ms, flip angle = 16 deg, GRAPPA acceleration factor = 2, FOV = 222 × 222 × 84 mm. In addition, a structural T1 sequence (MPRAGE, 1mm isotropic, TE = 3.03 ms, TR = 2,300 ms, flip angle = 8 deg, FOV = 256 × 256 × 192 mm) was acquired. A dual echo two-dimensional gradient echo sequence with voxel size of 3.5 × 3.5 × 2.0 mm, TR = 1,020 ms, dual echo (10 ms, 12.46 ms), flip angle = 90 deg, and separate magnitude and phase images was used to create a gradient field map to correct for distortions.

MRI data analysis. *Preprocessing.* Image preprocessing was performed using the Automatic Analysis Toolbox[22], which uses custom scripts combined with core functions from SPM8 (http://www.fil.ion.ucl.ac.uk/spm), FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) and FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). The structural images were bias-corrected and de-noised using an optimized non-local means filter to improve image quality[23]. The unified realign and unwarp procedure, as implemented in SPM8 (ref. 24), was used to correct for head motion and voxel displacement due to magnetic-field inhomogeneity. Co-registration of the functional images with the structural image was performed with the following procedure: the structural image was co-registered to the T1 template, and the mean EPI was co-registered to the EPI template. This co-registered mean EPI was then co-registered to the structural image. The co-registration parameters of the mean EPI were applied to all functional volumes. Functional images were corrected for physiological noise with RETROICOR[25]. RETROICOR uses ten cardiac phase regressors (fifth order fourier set), ten respiratory phase regressors (fifth order fourier set) and six other nuisance regressors (heart rate frequency, heart rate variability, raw respiration data, respiratory amplitude, respiratory frequency and frequency times amplitude of respiration). The FSL brain extraction toolbox was used to create a skull-stripped structural image. This structural image was segmented into gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) with SPM8 (ref. 26). Mean-intensity values at each time point were extracted for WM and CSF and used as nuisance regressors in the general linear model (GLM) analysis (see below).

Physiological measures. To correct for physiological noise (see above for details), heart rate was monitored with a pulse oximeter placed on the ring finger of the left hand using BrainAmp (ExG amplifier, Brain products GmbH). Participants were instructed to keep this hand as still as possible during the experiment. Heart rate data were inspected and corrected for movement-related and other measurement artifacts. Respiration was recorded at a sampling rate of 400 Hz using the respiration belt enclosed by BrainAmp (ExG amplifier, Brain products GmbH).

First-level modeling. For each narrative separately, each event in phase 1, 2 and 3 of the experiment (event A, B, C, X) was modeled with a GLM using two separate regressors: one for the three odd trials and one for the three even trials. These regressors were convolved with the canonical hemodynamic response function (HRF). First-level modeling was performed according to the methods suggested previously[27]. In short, for each regressor of interest, a separate GLM was performed containing the regressor of interest and another regressor including all other events of the experiment. This resulted in 96 GLMs per participant, with two regressors for each event (A, B, C, X) in each phase (1, 2, 3) for each narrative. Additionally, each GLM included: target events, link-events (link 1 and 2), and button presses (all convolved with the HRF), and 6 motion parameters (translations of X, Y, and Z coordinates, pitch, roll, and yaw), mean signal intensity in CSF, mean signal intensity in WM, and 26 regressors for physiological noise (see preprocessing for more detailed explanation). High pass filtering with a cutoff of 128 s was used to remove effects of low-frequency signal drifts.

RSA. We defined a priori regions of interest (ROIs, see below) and examined the correlation between across-voxel activation patterns of first-level beta estimates within these ROIs as a proxy of neural similarity[28]. The regressors modeling odd and even trials for events A, B, C and X were considered as the regressors of interest. We averaged across the correlations for odd and even trials which led to a 48 × 48 matrix of correlations (four events of interest per phase, three phases per narrative and four narratives). Only event-pair correlations for event pairs in the same task phase were analyzed (**Supplementary Fig. 1**). These Pearson's correlation coefficients were normalized using Fisher Z transformation. We then defined contrasts designed to model three different representational networks (**Supplementary Fig. 1**).

Small-scale network. We predicted that the smallest narrative scale would contain only representations of individual event-pair associations, which would,

in this context, reflect only the most recent directly associated event pairs and would depend on the posterior portion of the hippocampus. The contrast we used to test this prediction reflected an increase in A-B similarity from phase 1 to phase 2 followed by a decrease again from phase 2 to phase 3, combined with an increase in B-C similarity from phase 2 to phase 3, relative to B-X similarity. The contrast weights for phase 1, phase 2 and phase 3 were as follows: A-B: −1 2 −1/B-C: −1 −1 2/B-X: 2 −1 −1 (**Fig. 2a**).

Medium-scale network. We predicted that a medium-scale network would concurrently represent multiple event-pair associations, but without bridging between them. The contrast we used to test this prediction reflected an increase in A-B similarity and B-C similarity between phase 2 and phase 3, relative to A-C similarity and B-X similarity, with the following contrast weights for phase 2 and phase 3: A-B: −1 1/B-C: −1 1/A-C: 1 −1/B-X: 1 −1 (**Fig. 2a**).

Large-scale network. Finally, the large-scale network includes all possible event associations between both directly (A-B and B-C) and indirectly (A-C) related events. The contrast we used to test this prediction reflected an increase in A-B similarity, B-C similarity, and A-C similarity, between phase 2 and phase 3 relative to B-X similarity. To determine whether this model indeed reflected presence of indirect associations, we looked at an increase in similarity of the indirect link (A-C) separately, relative to control B-X similarity, with the following contrast weights for phase 2 and phase 3: A-B: −1 1/B-C: −1 1/A-C: −1 1/B-X: 3 −3 (**Fig. 2a**).

Contrasts were normalized by dividing each contrast by the square-root of the sum-of-squares of its contrast weights, which permitted us to directly compare the parameter estimates of the three different models. The sample size was based on our previous study[10] and power analysis was performed with G*power (dz = 1.033, alpha = 0.0001, power = 0.8). Participants were not grouped and therefore no randomization of participants was performed.

ROI definition. A hippocampal mask was constructed using the WFU pick-atlas[29]. We predicted a gradual change along the long axis of the hippocampus, and therefore split the hippocampal mask in approximately equal lengths along the long axis (posterior portion of the hippocampus: from Y = −40 to −30; mid-portion of the hippocampus: from Y = −29 to −19; anterior portion of the hippocampus: from Y = −18 to −4). A unified segmentation procedure (SPM8) was used to estimate parameters relating individual anatomy to MNI space. The inverse normalization parameters were used to create subject specific (gray matter) ROIs in native space based on the MNI masks described above (since first-level modeling was performed in native space).

ROI analyses. A repeated-measures ANOVA with a mnemonic scaling contrast (that is, a reduced a priori model with predicted contrast for small-scale, medium-scale and large-scale network, respectively: posterior: 2 −1 −1, mid-portion: −1 2 −1, anterior: −1 −1 2) with within-subject factors scale (small, medium, large), ROI (posterior portion, mid-portion, anterior portion) and hemisphere (left, right) was used to test the prediction of the increasing gradient (small-scale in posterior portion, medium-scale in mid-portion, large-scale in anterior portion of the hippocampus). To investigate this gradient further, we performed post-hoc repeated-measures ANOVAs for each model (small-scale, medium-scale, large-scale) and each ROI (anterior portion, mid-portion, posterior portion) separately with Hemisphere (left and right) as within-subject factor, see above and **Figure 2a** for details. To examine the interaction between the fMRI and behavioral results, we performed a repeated-measures ANOVA with within-subject factors scale (small scale, medium scale and large scale), ROI (anterior portion, mid-portion and posterior portion) and hemisphere (left, right), and included 'performance group' as between-subjects factor based on the performance in the narrative-recall task (full integration of all 4 narratives versus partial integration).

A **Supplementary Methods Checklist** is available.

21. Poser, B.A., Koopmans, P., Witzel, T., Wald, L. & Barth, M. *Neuroimage* **51**, 261–266 (2010).
22. Cusack, R. *et al. Front. Neuroinform.* **8**, 90 (2015).
23. Coupé, P., Yger, P., Prima, S., Hellier, P. & Kervrann, C. *IEEE Trans. Med. Imaging* **27**, 425–441 (2007).
24. Andersson, J.L., Hutton, C., Ashburner, J., Turner, R. & Friston, K. *Neuroimage* **13**, 903–919 (2001).
25. Glover, G.H., Li, T.Q. & Ress, D. *Magn. Reson. Med.* **44**, 162–167 (2000).
26. Ashburner, J. & Friston, K.J. *Neuroimage* **26**, 839–851 (2005).
27. Mumford, J.A., Turner, B.O., Ashby, F.G. & Poldrack, R. *Neuroimage* **59**, 2636–2643 (2012).
28. Kriegeskorte, N. & Kievit, R. *Trends Cogn. Sci.* **17**, 401–412 (2013).
29. Maldjian, J.A., Laurienti, P.J., Kraft, R. & Burdette, J.H. *Neuroimage* **19**, 1233–1239 (2003).

doi:10.1038/nn.4138