

Kent Academic Repository

Full text document (pdf)

Citation for published version

Michelmann, Sebastian and Staresina, Bernhard and Bowman, Howard and Hanslmayr, Simon (2018) Speed of time-compressed forward replay flexibly changes in human episodic memory. *Nature Human Behavior* .

DOI

<https://doi.org/10.1038/s41562-018-0491-4>

Link to record in KAR

<https://kar.kent.ac.uk/71430/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

UNIVERSITY OF BIRMINGHAM

Research at Birmingham

Speed of time-compressed forward replay flexibly changes in human episodic memory

Michelmann, Sebastian; Staresina, Bernhard; Bowman, Howard; Hanslmayr, Simon

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Michelmann, S, Staresina, B, Bowman, H & Hanslmayr, S 2018, 'Speed of time-compressed forward replay flexibly changes in human episodic memory' Nature Human Behaviour.

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Checked for eligibility: 14/11/2018

This is the accepted manuscript for a forthcoming publication in Nature Human Behaviour.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

1 Speed of time-compressed forward replay flexibly
2 changes in human episodic memory

3 Sebastian Michelmann¹, Bernhard P. Staresina¹, Howard Bowman^{1,2}, Simon
4 Hanslmayr^{1*}

5 1. University of Birmingham, School of Psychology, Centre for Human Brain Health; 2.
6 University of Kent, School of Computing

7 *Email: s.hanslmayr@bham.ac.uk

8

1 Remembering information from continuous past episodes is a complex task ¹. On the one hand, we
2 must be able to recall events in a highly accurate way that often includes exact timing; on the other
3 hand, we can ignore irrelevant details and skip to events of interest. We here track continuous episodes,
4 consisting of different sub-events, as they are recalled from memory. In behavioral and MEG data, we
5 show that memory replay is temporally compressed and proceeds in a forward direction. Neural replay
6 is characterized by the reinstatement of temporal patterns from encoding ^{2,3}. These fragments of
7 activity reappear on a compressed timescale. Herein, the replay of sub-events takes longer than the
8 transition from one sub-event to another. This identifies episodic memory replay as a dynamic process
9 in which participants replay fragments of fine-grained temporal patterns and are able to skip flexibly
10 across sub-events.

11 Episodic memory retrieval is a flexible process that operates at different timescales ¹. In some instances,
12 it is crucial for our behavior to mentally replay events at the same speed as the initial experience: Re-
13 enacting a classic movie scene relies on a temporally accurate representation of dialogue and events.
14 In other instances, it would be highly dysfunctional to recall our memories at the same speed they
15 originally unfolded: We have to be able to reconstruct how we came to work today without zoning out
16 at our desk for thirty minutes and must therefore be able to flexibly adjust the speed of our memory
17 replay.

18 Previous work has already put forward that memory replay in humans could be forward and
19 compressed: Studies that related the timescale between retrieval and perception of a particular event
20 asked participants to mentally navigate routes based on their memories. The duration of memory
21 replay (i.e. mental navigation) was found to be faster than the real navigation, but varied substantially
22 between participants ^{4,5}. Interestingly, this compression mirrors earlier findings of neural replay in
23 rodents, showing that hippocampal place cells, which correspond to certain positions along the animal's
24 path, later fire again on a faster timescale than during navigation. This is interpreted as reflecting
25 compressed replay of past trajectories ^{6,7}. One recent study in humans observed the reactivation of
26 static representations in electrocorticography (ECoG) and found patterns of oscillatory gamma power
27 reappearing faster than during perception ⁸. On the other hand, several studies that investigate neural
28 correlates of memory, find reinstatement of temporal patterns from perception in memory,
29 demonstrating that some patterns are replayed at the same speed ^{2,9-11}. Notably, a recent study that
30 applied functional magnetic resonance imaging (fMRI), managed to track continuous memory
31 reinstatement over long episodes (50min). Spatial patterns reappeared during the free recall of
32 narratives; recall was temporally compressed, but varied between participants ¹².

33 Despite these indications about the speed of replay in behavioral and neural data, no study so far has
34 tried to directly read out the temporal dynamics of memory replay in humans on a fine-grained
35 temporal scale. Importantly, the recent advent of multivariate methods in neuroscience has now
36 opened new avenues for the investigation of these processes: By leveraging multivariate patterns in
37 combination with electrophysiology, it is now possible to track representations from perception in a
38 time-resolved manner, as they reappear during memory retrieval ^{2,8,9,13-16}. Importantly, simultaneous
39 EEG and multi-unit recordings in primates demonstrate an intimate relation between neural firing and
40 the phase of slow oscillations in the EEG ¹⁷. Therefore, information about neural patterns can be
41 captured and tracked in human electrophysiology via oscillatory phase ^{2,17,18}.

1 Capitalizing on these methodological advances, we here investigate the flexible dynamics of episodic
2 memory replay in continuous mnemonic representations. Studying these trajectories during memory
3 replay requires a paradigm that prompts participants to evoke continuous representations with distinct
4 subevents from memory. This will make it possible to track fragments of these representations in
5 episodic memory via multivariate analysis methods. To this end, we asked subjects to associate static
6 word-cues with 'video-episodes' consisting of a sequence of three distinct scenes (scenes were short
7 videos of 2 seconds duration each). The three dynamic scenes thus formed a continuous six-second-
8 long video. In encoding-trials, we presented a word-cue during one of the scenes. This allowed us to
9 prompt memory replay in a natural way, i.e. we asked participants to recall in which of the three scene-
10 positions they had learned an association during encoding. After completing this part of the task, we
11 asked about the video-episode itself and confirmed memory accuracy. In a behavioral experiment, we
12 investigated direction and speed of replay via measuring reaction times to the scene-position response.
13 In a separate MEG study, we leveraged the content-specific phase patterns that each scene elicited and
14 used them as handles to track the direction and speed of replay of the video-episodes. If memory replay
15 were indeed compressed, we expected to find evidence for this compression in reaction times and in
16 the reinstatement of neural patterns. This replay could either be forward or backward. In line with
17 previous findings, we expected to find evidence for reactivation of temporal patterns, signifying replay
18 at the same speed for fragments of neural activity^{2,9-11}. We further hypothesized that the disparity
19 between accurate representations and overall compression would be due to a flexible mechanism that
20 allows subjects to skip between temporally accurate patterns (i.e. omit information between
21 fragments). Skipping between sub-events (i.e. scenes of the video-episode) should furthermore
22 manifest itself in a slower average replay within sub-events compared to the overall compression level,
23 if replay is initiated more often from the beginning of a scene than within a scene.

24 In the behavioral experiment, participants associated word-cues with one of three scenes within video-
25 episodes (Figure 1a). We used four continuous video-episodes, each consisting of three individual
26 dynamic scenes. A trial-unique word-cue appeared in one scene during a video-episode. After a brief
27 distractor task (Figure 1b) subjects performed, in alternation, either a cued-recall (CR) retrieval task or
28 an associative-recognition (AR) task (Figure 1d, top). The AR task was included as a control condition,
29 because active replay is arguably not required for recognition. In the CR blocks, we presented
30 participants with the word-cues (Figure 1d, top-left). Their task was to recall the scene-position that
31 was associated with the word-cue as quickly as possible. In AR blocks, subjects successively saw the
32 word-cues superimposed on screenshots from encoding and were asked to decide as quickly as possible
33 whether this association was intact or rearranged (Figure 1d, top-right).

34 To address the direction and speed of memory replay, reaction times (RTs) at retrieval were compared
35 between associations that were learned in the first, second and third scene-position of a video-episode
36 (Figure 1d, bottom). We only used RTs for correct hit trials (correct recall in CR and correctly recognized
37 intact associations in AR blocks) and excluded trials in which the subjects were wrong or guessed (see
38 Supplementary Information for the same analysis including correct guesses). If the CR task indeed
39 elicited replay in the forward direction, we should observe faster reaction times with CR, but not AR,
40 for associations that were learned earlier during encoding. Furthermore, if replay was compressed, the
41 delay between reaction times to different scene-positions should be smaller than the duration of the
42 scenes segments themselves. The resulting 3x2 repeated-measures ANOVA tested the factors position
43 and task. A significant main effect of position ($F_{1.85, 42.48} = 5.884, p = 0.007, \text{partial } \eta^2 = 0.204, 95\% \text{ CI}[0.081, 0.414]$), log-RT: $F_{1.79, 41.26} = 3.375, p = 0.049, \text{partial } \eta^2 = 0.128, 95\% \text{ CI}[0.028, 0.352]$) and a

1 position by task interaction ($F_{1.75, 40.34} = 5.9, p = 0.008$, partial $\eta^2 = 0.204$, 95% $CI[0.05, 0.44]$, log-RT: $F_{1.76,$
2 $40.58} = 5.606, p = 0.009$, partial $\eta^2 = 0.196$, 95% $CI[0.03, 0.467]$) were obtained. Both effects were driven
3 by the cued-recall task (repeated-measures ANOVA position only: $F_{1.79, 41.19} = 9.082, p = 0.001$, partial η^2
4 $= 0.283$, 95% $CI[0.102, 0.514]$, log-RT: $F_{1.60, 36.90} = 8.207, p = 0.002$, partial $\eta^2 = 0.263$, 95% $CI[0.079,$
5 $0.504]$): During encoding, individual scenes of each video-episode lasted 2 seconds. During CR retrieval,
6 however, associations that were learned in the first scene-position of a video-episode (mean RT = 2.5s)
7 were recalled on average 116ms faster than associations that were learned in the second scene-
8 position (one-tailed $t_{23} = -1.870, p = 0.037$, Cohen's $d = -0.382$, 95% $CI[-0.793, 0.037]$, log-RT: one-tailed
9 $t_{23} = -2.4, p = 0.012$, Cohen's $d = -0.49$, 95% $CI[-0.909, -0.061]$). Associations that were learned in the
10 second scene-position (mean RT = 2.617s) were recalled on average 176ms faster than associations
11 that were learned in the third scene-position (one-tailed $t_{23} = -2.767, p = 0.006$, Cohen's $d = -0.565$, 95%
12 $CI[-0.991, -0.128]$, log-RT: one-tailed $t_{23} = -2.274, p = 0.016$, Cohen's $d = 0.464$, 95% $CI[-0.881, -0.038]$),
13 (mean RT = 2.793s). The replay of the video-episodes was therefore forward and compressed during
14 CR, which replicated our findings from a behavioral pilot experiment (see Supplementary Information).
15 The average RT difference of 146ms per position corresponds to a compression factor of 13.7 during
16 replay.

17 Might the effects be due to asymmetrical encoding of scene-positions? One could argue that
18 associations have a higher saliency when they are presented in the first scene-position, leading to
19 higher confidence and shorter RTs during retrieval. Additionally, subjects can take more time to
20 rehearse early associations during the remainder of the video-episode, perhaps resulting in the weakest
21 memory trace for the last scene. Importantly, however, if the serial position merely affects the overall
22 strength of the memory trace in our paradigm, we should observe comparable effects on cued recall
23 (CR) and associative recognition (AR). Conversely, if the effect is contingent on the need to mentally
24 replay scene after scene, serial position at encoding should only exert an effect on the CR task.

25 Importantly, no differences in reaction times between scene-positions were evident in the AR task
26 (Figure 1d, right; repeated-measures ANOVA: $F_{1.64, 37.66} = 0.708, p = 0.472$, partial $\eta^2 = 0.03$, 95%
27 $CI[0.003, 0.196]$, log-RT: $F_{1.61, 36.95} = 0.793, p = 0.435$, partial $\eta^2 = 0.033$, 95% $CI[0.002, 0.231]$, pairwise
28 comparisons of positions: all $ps > 0.199$, all $BF_{01} > 2.158$). The Bayesian statistics provide evidence for
29 the specificity of the effect, a significant position by task interaction, further supports that the position
30 effect on RTs is specific to the CR task and contradicts a saliency-based explanation. Finally, we observed
31 a significant main effect of task with unscaled ($F_{1.00, 23.00} = 62.349, p < 0.001$, partial $\eta^2 = 0.731$, 95%
32 $CI[0.633, 0.849]$) and log-transformed ($F_{1.00, 23.00} = 95.036, p < 0.001$, partial $\eta^2 = 0.805$, 95% $CI[0.718,$
33 $0.899]$) reaction times. This was due to faster RTs in associative-recognition blocks (two-tailed $t_{23} = -$
34 $7.896, p < 0.001$, Cohen's $d = -1.612$, 95% $CI[-2.216, -0.992]$, log-RT: two-tailed $t_{23} = -9.749, p < 0.001$,
35 Cohen's $d = -1.99$, 95% $CI[-2.68, -1.285]$). Taken together these results are evidence that successful
36 recall of elements from a continuous video-episode relies on compressed forward replay.

37 In the alternating blocks of the behavioral experiment, participants recalled on average 69.47% ($SD =$
38 23.21%) of the correct word-scene associations in cued-recall (CR) blocks. They further recognized
39 90.27% ($SD = 10.74\%$) of intact associations (Hits) and erroneously named 12.40% ($SD = 14.38\%$) of
40 rearranged associations intact (False Alarms) in an associative-recognition (AR) blocks. Performance in
41 CR (i.e. percent correct responses) and in AR (i.e. percent Hits minus percent False Alarms) was
42 compared with a 2x3 repeated-measures ANOVA. This revealed a significant main effect of task ($F_{1, 23} =$
43 $38.30, p < 0.001$, partial $\eta^2 = 0.625$, 95% $CI[0.408, 0.826]$), driven by a better performance in the

1 associative-recognition blocks (two-tailed $t_{23} = 6.189$, $p < 0.001$, Cohen's $d = 1.263$, 95% $CI[0.716,$
2 $1.796]$) and a significant factor position ($F_{1.72, 39.65} = 7.624$, $p = 0.002$, partial $\eta^2 = 0.249$, 95% $CI[0.125,$
3 $0.515]$, interaction task with position $F_{1.84, 42.24} = 1.145$, $p = 0.324$, partial $\eta^2 = 0.047$, 95% $CI[0.003,$
4 $0.238]$). This was driven by a slightly better performance in the cued-recall task, for associations that
5 were learned in the second position of a video-episode (repeated-measures ANOVA: $F_{1.58, 36.24} = 2.794$,
6 $p = 0.086$, partial $\eta^2 = 0.108$, 95% $CI[0.045, 0.342]$, position 1 vs. 2: two-tailed $t_{23} = -2.804$, $p = 0.01$,
7 Cohen's $d = -0.572$, $CI[-1 -0.135]$, position 2 vs. 3: two-tailed $t_{23} = 1.962$, $p = 0.062$, Cohen's $d = 0.4$, 95%
8 $CI[-0.02, 0.813]$) and a worse performance in associative-recognition for associations that were learned
9 in the third position (repeated-measures ANOVA: $F_{1.86, 42.68} = 5.552$, $p = 0.008$, partial $\eta^2 = 0.194$, 95%
10 $CI[0.091, 0.373]$, position 2 vs. 3: two-tailed $t_{23} = 3.879$, $p < 0.001$, Cohen's $d = 0.792$, 95% $CI[0.325,$
11 $1.246]$, all other $ps > 0.14$). Importantly the better performance for the second position in the CR task
12 cannot explain the RT effects. Firstly, in the behavioral data there is no difference in performance
13 between the 1st and the 3rd position (two-tailed $t_{23} = -0.282$, $p = 0.781$, Cohen's $d = 0.058$, 95% $CI[-$
14 $0.457, 0.344]$), yet the biggest difference in reaction time is between the 1st and the 3rd position.
15 Furthermore, position 2 is remembered better than position 1 in the CR task of the behavioral data
16 (position 1 vs. 2: two-tailed $t_{23} = -2.804$, $p = 0.01$, Cohen's $d = 0.572$, 95% $CI[-1, -0.135]$), yet the average
17 reaction times are faster for the first position.

18 In the MEG experiment subjects remembered on average 63.54% ($SD = 11.768\%$) of associations,
19 excluding guesses. After preprocessing on average 200.348 trials ($SD = 38.645$) remained for known
20 correct associations and an additional 116 trials ($SD = 39.425$) were guessed or incorrect responses. In
21 the MEG experiment subjects completed on average 10.8696 blocks ($SD = 5.8644$). In the behavioral
22 experiment, subjects completed on average 12.3333 blocks ($SD = 6.6833$) of
23 encoding/distractor/retrieval. These alternated between the two tasks.

24 In the MEG experiment, participants performed the same CR task as in the behavioral experiment, with
25 the only difference being that they gave responses after the word-cue disappeared (Figure 1c). In a first
26 step, we asked whether perceptual content could be distinguished based on oscillatory phase. To this
27 end, we compared the inter-trial phase coherence (ITPC) between encoding-trials grouped according
28 to their video-content against the ITPC between trials grouped randomly. This has been used previously
29 to reveal the content specific entrainment of cortical rhythms to naturalistic dynamic stimuli^{2,17}. The
30 four video-episodes showed reliably distinguishable phase patterns during encoding ($p_{\text{cluster}} < 0.001$,
31 Figure 2a, left and middle). The significant cluster (across time space and sensors) contained robust
32 differences in the lower frequencies and showed a maximum over occipito-parietal sensors (Figure 2a,
33 middle). Consistent with our previous results², strongest differences were observed at the onset of
34 each scene. Importantly, the frequency band centered at 8 Hz was included in the cluster, which was
35 previously linked to the reinstatement of phase patterns². Testing the 8 Hz phase differences on the
36 source level revealed one broad cluster of content specificity during encoding ($p_{\text{cluster}} < 0.001$).
37 Averaging t-values across this significant cluster over time revealed highest values in occipital and
38 parietal locations (Figure 2a right). Together, these results show that every sub-scene within a video-
39 episode was associated with a content specific fingerprint in oscillatory phase, which was maximal in a
40 parieto-occipital region. In the following, we used these sub-scene specific phase patterns at the center
41 frequency of 8 Hz as handles to track replay in memory.

42 In a first step, we tested whether these phase-patterns of the video-episodes were reactivated in
43 memory. Therefore, we first contrasted phase-similarity between encoding-retrieval combinations of

1 the same video-episodes (e.g. watching video A, recalling video A) with encoding-retrieval combinations
2 of different video-episodes (e.g. watching video A, recalling video B). Similarity between encoding and
3 retrieval phase patterns was analyzed with a sliding-window approach (window size = 1 sec), providing
4 a time resolved measure of memory replay^{2,19,20} (see Figure 2c). On the source level, analysis was
5 restricted to an anatomically defined occipito-parietal region of interest (ROI) following the results from
6 the encoding phase and previous studies showing memory replay in these regions^{2,21-24} (Figure 2b).
7 Content specific phase was assessed separately at every virtual sensor and corrected for multiple
8 comparisons via random permutation considering spatial clusters. Evidence for replay was found for hit
9 trials (Hits; $p_{cluster} = 0.034$; Supplementary Figure 3a, also see Supplementary Figure 3b for unmasked
10 maps of t-values), suggesting that replay of video-episodes can be tracked in the phase of an 8Hz
11 oscillation. Notably, we found no such replay effect for Misses, i.e. trials in which subjects either
12 guessed, or did not remember the correct scene-position and/or video-episode. Furthermore, a direct
13 contrast between Hits and Misses revealed significantly stronger replay for Hits compared to Misses
14 ($p_{cluster} = 0.030$, Figure 2d), demonstrating the functional significance of this pattern-reinstatement for
15 memory.

16 The above findings confirm that content specific patterns of activity from encoding, are reinstated in a
17 purely memory driven way. This motivated us to ask in which direction and at what relative speed
18 patterns from encoding unfold during retrieval. Do patterns from the beginning of the video-episodes,
19 for instance, also reappear earlier during memory retrieval?

20 To this end, we divided the encoding interval into 6 non-overlapping windows, centered at 0.5, 1.5, 2.5,
21 3.5, 4.5, and 5.5 seconds. We then analyzed the phase-similarity to these windows across the retrieval
22 interval. The latency at which patterns reappear should be reflected in the distribution of phase-
23 similarity across time. Consequently, we compared these distributions between the distinct time
24 windows from encoding (Figure 3a).

25 Specifically, to test the direction of replay statistically across subjects, we used the following approach:
26 We cumulated the similarity distributions across the whole retrieval time. This provided the cumulated
27 similarity (CS) for every subject and every encoding-window. Similarity started at the beginning of the
28 retrieval interval with a value of zero. It ended at the end of the retrieval interval, with a value of one
29 (Figure 3c). If phase-similarity to an encoding-window "A" cumulates earlier than phase-similarity to an
30 encoding-window "B", then the cumulated similarity for "A" is higher compared to "B" and
31 consequently "A" is replayed earlier during retrieval than "B". In other words, when the CS of one
32 phase-pattern is higher than the CS of another, then the evidence for replay of that phase-pattern is
33 leading over the other at that point. If, however replay of a phase-pattern is lagging behind the replay
34 of another, the CS should be lower at that time point. We tested this relation statistically at every time
35 point by comparing the cumulated similarity across all windows for each subject. The overall tendency
36 is tested best by fitting a line across all six encoding windows. Herein, a negative slope indicates forward
37 replay, since earlier windows have higher values in the CS than later windows, a positive slope signifies
38 backward replay. We tested this slope against 0 with a two-sided t-test and corrected for multiple
39 comparisons by controlling the false discovery rate²⁵.

40 Results revealed significant forward replay in two time windows (i.e. 135ms to 1919ms, and 3458ms to
41 3473ms after cue presentation, see Figure 3d. See also Methods for some notes of caution regarding
42 the interpretation of the exact time-window). We can therefore conclude that there is a dominance of

1 early encoding-patterns in early time points at retrieval, relative to late encoding-patterns. This
2 supports the notion of forward replay (see also Supplementary Information for additional evidence
3 supporting forward replay) and corroborates the finding of forward replay from the behavioral
4 experiment.

5 In neural data, the forward direction of replay was evidenced by the tracking of content specific
6 temporal patterns. Notably, however the reactivation of temporal patterns signifies that participants
7 replayed *fragments* of the video-episodes at roughly the same speed as during encoding. Hence, these
8 data already indicate that memory replay is not the straightforward recapitulation of the original
9 experience. Instead, flexible processes must be at work to reconcile the overall compression of memory
10 with the reappearance of temporal patterns.

11 We hypothesized that the disparity between locally detailed patterns and the global compression was
12 possible through the flexible skipping between salient components in memory (e.g. sub-events); in our
13 data, the boundaries between scenes were salient elements within the video-episodes. We therefore
14 investigated whether these boundaries would serve as stepping stones enabling participants to skip
15 through their memories on a faster time-scale. Consequently, we tested statistically whether the speed
16 of replay slowed down within scenes, since the skipping between the scenes of a video-episode should
17 be easier and more likely than skipping within the individual scenes.

18 To this end, we extended the method of fitting a line across CSs to compare the compression of replay
19 within individual scenes (i.e. within sub-events) to the overall compression level. Specifically, calculating
20 the slope of the fitted line allows for an estimation of the speed of replay. This slope indicates the lag
21 between replayed patterns in the retrieval interval, such that steep slopes indicate a long lag (i.e. slow
22 replay). We fitted a separate line for each pair of encoding-windows that belonged to the same scene
23 across their respective CSs and averaged the slopes across the three lines. The time interval between
24 442ms and 2350ms displayed slopes significantly below zero, confirming forward replay within scenes.
25 This was tested with a series of one-sided t-tests, controlling the false-discovery rate. More importantly,
26 between 550ms and 2350ms at retrieval, slopes of windows within a scene were significantly steeper
27 (i.e. replay was slower) compared to the slope obtained across all encoding-windows (Figure 3e). These
28 slopes were compared again at every time point across participants with a one-sided t-test, controlling
29 the false-discovery rate. This means that when participants replayed the first and second part of a
30 scene, this replay was less compressed than we expected from the global compression level of the
31 whole video-episode. Consequently, this also means that subjects did not recapitulate every scene
32 successively in every trial. Taken together, these results show that memory replay does not occur at a
33 constant speed; instead, the speed of replay seems to change flexibly depending on the replayed
34 interval (Figure 3b, right). Finally, we repeated these tests with those trials in which subjects did not
35 remember the correct positional-scene or video-episode; however, we found no significant time-points
36 for any of the contrasts, which demonstrates the implication of these replay effects in memory (see
37 Supplementary Information). In a further control analysis, we excluded the first 800ms of the retrieval
38 interval for the similarity analysis in order to rule out that event related potentials (ERPs) drove
39 similarities. Again, we found significant negative slopes between 812ms and 1212ms and slower replay
40 within scenes in that window (see Supplementary Information). The slopes were again compared at
41 every time point across participants with a one-sided t-test, controlling the false-discovery rate. Finally,
42 we averaged the similarity for the windows that belonged to the same scene and repeated the

1 cumulated similarity analysis: Significantly negative slopes between 550ms and 1919ms ensured that
2 forwardness was not merely driven by forwardness within scenes (see Supplementary Information).

3 These results statistically support a flexible forward replay strategy. Via cross-correlations, we next
4 derived a descriptive measure of the delay between the six sub-events during flexible memory replay
5 (550ms-2350ms). The cross-correlation was computed on pairs of averaged and smoothed similarity
6 distributions (Figure 3b), which retained a time lag value for every combination of the six sub-events.
7 The adaptive replay that we found is also visible in the pattern of time lags and can be illustrated with
8 shorter lags between time windows that belong to different scenes compared to time windows that
9 belong to the same scene (Figure 3B, right). In contrast, to illustrate a strict and inflexible forward replay
10 strategy, lags between the sub-events should increase linearly according to their position at encoding
11 (illustrated in Figure 3B, right).

12 In this study, we tracked the replay of continuous episodes from memory. We used a paradigm in which
13 participants associated unique word-cues with one out of three distinct scenes in seamless video-
14 episodes. We prompted replay by asking volunteers in which exact position (1, 2, or 3) they had learned
15 each word-cue. Behavioral and neural data indicated that replay of memories takes place in a forward
16 direction and at a compressed speed, i.e. memory replay was faster relative to perception. Notably, on
17 a neural level, we found indications for different speeds of replay: Fragments of temporal patterns
18 reappeared at the same speed and the speed of replay within sub-events (i.e. scenes) of continuous
19 video-episodes was slower than the overall compression level. Notably, our method assesses the
20 reinstatement of patterns in oscillatory phase over time, i.e. discarding spatial information and
21 amplitude information in the signal. This provides direct evidence for temporal 'replay' of patterns, as
22 opposed to the reinstatement of static information.

23 Importantly, our finding of different compression levels implies that memory replay acts in a flexible
24 way. The disparity between the slower speed of replay within scenes and the overall compression is
25 an aggregated observation that cannot hold on a single trial level. Specifically, it signifies that replay is
26 not a simple concatenation of fragments because in a single trial, the sequential replay of three
27 scenes would take longer than the overall compression permits. Consequently, participants must be
28 able to skip between replayed fragments; importantly the slower speed of replay within scenes
29 denotes that on average, the skipping between sub-events must take place on a faster temporal scale
30 than the skipping within sub-events. A plausible interpretation of the observed pattern is therefore
31 that replay of relevant information is initiated from the boundaries between scenes and that
32 participants can flexibly skip between them. Event boundaries²⁶ have been previously shown to
33 trigger replay events during memory encoding²⁷. They could therefore also serve as starting points
34 during memory retrieval, to initiate the replay of information on a fine-grained temporal scale.

35 Specifically, if replay proceeds from event boundaries on a slower timescale, the moments of replay
36 for the second part of a scene will, on average, be substantially delayed relative to the start of the
37 sub-scene (i.e. event boundary). On the other hand, the replay of the beginning of a new scene can
38 start relatively early after the beginning of a previous scene, because replay can be initiated from this
39 event boundary (see Figure 4).

40 Mechanistically, the hippocampus has been suggested to preserve the temporal order of experiences
41²⁸ by storing a sparse index²⁹. Accordingly, interactions between the hippocampus and visual cortex
42 have been observed during memory replay in sleeping rodents²¹. We therefore speculate that the

1 here observed replay, which was located in posterior cortical areas, may have been triggered by the
2 hippocampus in order to execute the vivid reinstatement of sensory information. Future studies with
3 access to hippocampal and cortical signals should investigate this hypothesized interaction during
4 memory replay. Notably, our task requires subjects to rely on sensory representations and likely
5 promotes such accurate sensory pattern reinstatement. At first glance, the reinstatement of temporal
6 patterns is also at odds with the observation of compression in general. An important implication
7 from the finding of temporal pattern reinstatement under global compression is therefore that the
8 accurate reinstatement of patterns must be limited to fragments of the original perception. In other
9 words, subjects possibly omit non-informative (perhaps redundant) parts of the video-episodes and
10 therefore replay a shorter episode in memory, which contains less information. Previous work on
11 mental simulation of paths supports this interpretation. The duration that participants take to
12 mentally simulate a path increases, when this path includes more turns⁵. In the same way, the
13 duration of replay might depend on the overall number of relevant elements within a video-episode.

14 Another crucial result from our experiments is the forward direction of replay. This finding is in line
15 with recent studies showing anticipatory activation of familiar paths in the visual cortex²³ and
16 evidence of forward replay of long narratives¹². Notably, in the rodent literature, the task of spatial
17 navigation appears to determine whether replay is backward or forward. At the end of a path, awake
18 rodents replay in a backward fashion⁶, whereas animals that plan the path towards a goal display an
19 anticipatory activation of place-cells in the forward direction³⁰. Task requirements in our design
20 could indeed have prompted participants to step mentally through the video-episodes in a forward
21 manner. Speculatively, other designs (e.g. tasks requiring recency judgments) might therefore cause a
22 backwards replay. This would be well in line with the flexibility in memory replay that we observed in
23 the neural data, since a flexible mechanism could arguably guide replay in a forward and backward
24 direction when skipping through events. An interesting additional question arising from this is
25 whether replay of fine-grained temporal patterns in the cortex can also be backwards.

26 Importantly our study also demonstrates how one can investigate these open questions. The design
27 that we used to trigger the replay of distinct sub-events in a continuous episode can easily be adapted
28 to a working memory context and our method to track oscillatory patterns allows for the investigation
29 of replay in working memory, during rest and during sleep. We have repeatedly shown how to use the
30 similarity in oscillatory phase to track content-specific reactivation, even when the exact onset of
31 memory-reactivation is unknown. We here extended our previously developed method² to track
32 distinct sub-events from continuous representations: In a statistically robust way we aggregated
33 evidence across several repetitions and compared their distribution across time.

34 This investigation of temporal dynamics during human episodic memory replay has only recently
35 become an option, when the tracking of multivariate patterns was extended to human
36 electrophysiology^{2,9,11,15,16,27,31}. Leveraging a paradigm in combination with a method that can detect
37 the individual fingerprints in oscillatory patterns, we were now able to observe the fine-grained
38 dynamics of memory replay on a behavioral and on a neural basis. Our data render memory replay as
39 a flexible process, namely the compression level varies within replayed episodes: Some fragments
40 reappear on a timescale that resembles the original perception and replay is less compressed within
41 sub-events of continuous episodes, which suggests that participants were able to flexibly skip
42 between sub-events during memory replay.

1 Methods

2 Participants

3 No statistical methods were used to pre-determine sample sizes but our final sample sizes are similar
4 to those reported in previous publications ^{2,3}.

5 **Behavioral pilot and experiment**

6 For the behavioral pilot only 12 subjects (8 female, 4 male) participated that were on average 22.58
7 years old (youngest: 19, oldest: 29). 2 of the female participants were left handed, the rest were right
8 handed. Data from 24 right handed volunteers (18 female, 6 male) was acquired for the behavioral
9 experiment. The average age of this sample amounted to 22.79 years (youngest: 20, oldest 34).

10 **MEG experiment**

11 For the MEG experiment 24 volunteers (13 male, 11 female) participants were tested. Subjects were
12 between 18 and 34 years old (mean: 23.92 years). 6 participants were left handed, 18 participants were
13 right handed. 1 of the 24 subjects was excluded after pre-processing because of a persistent electrical
14 artifact in the data that could not be removed with filtering.

15 6 additional subjects (4 female, 2 male) aged 19 to 28 years (mean: 22) were recorded but not analyzed.
16 2 subjects moved excessively throughout the recording session (maximal movement: 1.8 cm and 2.7
17 cm), 1 subject moved excessively throughout the session (maximal movement 1.4 cm) and fell asleep
18 during the experiment. 1 subject felt unwell and aborted the experiment after approx. 10 % of the
19 recording session, 1 subject only completed approx. 70 % of the recording session and moved more
20 than 2 cm throughout the experiment. Finally 1 subject was lost due to technical failure during the
21 recording. After preprocessing, the maximal movement of included participants across all trials (i.e. the
22 range of all positions) was on average 5.89mm (s.d. = 2.62, min = 1.69, max = 9.09).

23 All participants in the pilot studies, behavioral experiments and the MEG experiment, were native
24 English speakers. Before participation they were screened for any neurological or psychiatric disorders.
25 Their informed consent was obtained according to the ethical approval that was granted by the
26 University of Birmingham Research Ethics Committee (ERN_15–0335A).

27 Material and experimental set up

28 Stimulus material and allocation of experimental conditions were pseudo-randomized, as described in
29 the corresponding sections below.

30 **Videos**

31 For each of the balancing pilots (Supplementary Information), a total of 12 short video-clips were used.
32 Videos stemmed from a pool that was provided by Landesfilmdienst Baden-Württemberg, Germany,
33 some of them were additionally edited. Each video-clip was a 2-second-long colored, dynamic scene
34 that featured a single action (i.e. a ship sailing or a diver jumping into the water). During the task, video-
35 clips were always superimposed with a transparent text box (white box with alpha value 0.9) in which
36 the word-cue could appear. According to the behavioral results from balancing pilot 1, we edited or

1 changed some of the scenes before the second balancing pilot. The final video-clips were 12 different
2 scenes that belonged to four general topics. For the behavioral experiments and the MEG experiment,
3 the video-clips were then grouped into four seamless sequences of frames that formed a video-episode
4 (i.e. a sequence of three scenes that belong to a general topic and form a short story). The 3 scenes of
5 each video-episode were clearly distinguishable.

6 According to the second balancing pilot, scenes that were assigned to be in 1st, 2nd or 3rd position of
7 video-episodes, did not differ significantly in difficulty (percent correct responses), when associated
8 with a word-cue. Pairwise comparisons with t-test of positions 1 and 2 ($t_{17} = 0.86, p = 0.4$), 2 and 3 (t_{17}
9 $= 0.15, p = 0.88$) and 1 and 3 ($t_{17} = 1.4693, p = 0.16$) and Bayes-Factor analysis supported the null
10 Hypothesis of no difference between positions. This was supported either by substantial ($BF_{01} > 1.6$) or
11 strong ($BF_{01} > 3.3$) evidence for the null hypothesis in the comparison of positions 1 and 2 ($BF_{01} = 2.97$)
12 of positions 2 and 3 ($BF_{01} = 4.07$) and of positions 1 and 3 ($BF_{01} = 1.65$). Importantly, reaction times in
13 the second balancing pilot did not differ significantly between the video-clips that we finally assigned
14 to be in position 1, 2 or 3. Pairwise comparisons with t-test of assigned positions 1 and 2 ($t_{17} = -0.59, p$
15 $= 0.56$), 2 and 3 ($t_{17} = -0.31, p = 0.76$), and 1 and 3 ($t_{17} = -1, p = 0.33$) and Bayes-Factor analysis supported
16 the null Hypothesis for the comparison of positions 1 and 2 ($BF_{01} = 3.53$) of positions 2 and 3 ($BF_{01} =$
17 3.95), and positions 1 and 3 ($BF_{01} = 2.67$).

18 **Word-cues**

19 Word-cues were downloaded from the MRC Psycholinguistic Database ³². For the balancing pilots
20 (Supplementary Information), we divided 540 word-cues into 18 lists. Those lists did not differ in
21 Kucera-Francis written frequency (mean = 20.80, s.d. = 8.55), concreteness (mean = 506.50, s.d. =
22 90.07), imageability (mean = 521.04, s.d. = 69.51), number of syllables (mean = 1.63, s.d. = 0.68),
23 number of letters (mean = 5.61, s.d. = 1.42) or word-frequencies taken from SUBTLEXus (mean = 15.22,
24 s.d. = 14.07); specifically, “Subtlwf” was used ³³. In the balancing pilots, 12 of the lists were associated
25 with a video-clip and 6 of the lists were assigned to become a distractor word. Across subjects each list
26 was associated with every movie once and served as a distractor word six times. This was done to
27 additionally control for list specific effects across subjects. An additional 9 words were randomly
28 selected for practice.

29 For the behavioral pilot, the behavioral experiment and the MEG experiment, we divided 360 word-
30 cues into 12 lists. Those lists were likewise balanced for Kucera-Francis written frequency (mean =
31 20.41, s.d. = 7.47), concreteness (mean = 518.72, s.d. = 78.39), imageability (mean = 530.78, s.d. =
32 60.17), number of syllables (mean = 1.56, s.d. = 0.62), number of letters (mean = 5.44, s.d. = 1.30) and
33 word-frequencies taken from SUBTLEXus (mean = 15.07, s.d. = 13.04); again, “Subtlwf” was used ³³.
34 Across participants, each of the lists was associated with every video-clip twice. An additional 6 words
35 were randomly selected for practice.

36 **Response options**

37 To create the response options (see Figure 1c-d), we took Screenshots from the video-clips. In the
38 balancing pilots (Supplementary Information), we adjusted brightness and contrast, so that no
39 screenshot appeared more salient. For the behavioral pilot, the behavioral experiment and the MEG
40 experiment the numbers 1, 2 and 3 were framed by a square which resembled a frame from an old
41 film. Those represented the first response options, i.e. the choice between scene 1, scene 2 or scene 3.

1 For the second response, i.e. the response about the correct video-episode, the 3 screenshots from the
2 concatenated video-clips were presented next to each other for each of the 4 choices. In the control
3 task of the behavioral experiment, the response option intact/rearranged was realized with a
4 screenshot which was of the same size as the videos during presentations. This screenshot was
5 superimposed by a transparent textbox containing a word-cue. The words intact and rearranged were
6 displayed at the left and right of the textbox as response options. The left/right position of these options
7 was balanced across participants.

8 **Behavioral setup**

9 Visual content was presented on an LED monitor (Samsung syncmaster 940n at a distance of
10 approximately 60 cm from the subject's eyes. The monitor was set to a refresh rate of 60 Hz. On a
11 screen size of 1280 x 1024 pixels, the video-clips had the dimension of 360 pixels in width and 288 pixels
12 in height on the screen. "Helvetica" was chosen as the general text font, font size was set to 22 for
13 instructions and to 28 for word-cues. Black text (rgb: 0, 0, 0) and movies were presented against a white
14 background (rgb: 255, 255, 255).

15 **MEG setup**

16 MEG was recorded at the Sir Peter Mansfield Imaging Center (SPMIC) in Nottingham, UK. Subjects
17 performed the experiment in a seated position at a distance of approximately 60 cm from a white
18 screen. The image was projected onto the screen using a PROPixx projector (VPixx Technologies, Saint-
19 Bruno, Canada) that operated at a refresh rate of 60Hz and a resolution of 1920 x 1080 px. The
20 projected image appeared at a size of approx. 40 x 22.5 cm on the screen. Accordingly, the video-clip
21 appeared in a dimension of approx. 15 x 12 cm. An eye tracker (EyeLink 1000 plus, SR Research, Ontario,
22 Canada) was placed in front of the screen. The tracker was mounted in an upwards facing orientation,
23 slightly below the visible display, on a small wooden board. In this setup it tracked the subject's left eye
24 from below and from a distance of approximately 55 cm.

25 Procedure

26 **Behavioral pilot and behavioral experiment**

27 In the behavioral pilot (Figure 1a, b, d top-left), subjects saw video-episodes that consisted of 3 distinct
28 scenes. Those scenes comprised of the video-clips from balancing pilot 2 (Supplementary Information),
29 which ensured that no material specific differences were to be expected between position 1, 2 and 3
30 of the video-episodes; not in memory performance and most importantly not in reaction time.
31 Participants first completed the screening questionnaire and gave informed consent. After instruction
32 with the task, they saw the video-episodes twice for familiarization and were instructed to pay attention
33 to their 3-scene-structure, such that they could confidently identify the first, second and third scene of
34 each video-episode.

35 After a short practice version of the task, the experiment started. It was again a sequence of encoding,
36 distractor and retrieval blocks. In each encoding block subjects learned a series of associations (Figure
37 1a). They first saw a fixation cross on the screen for 2 seconds. Thereafter one of the four video-
38 episodes played for 6 seconds. During this video-episode a transparent textbox was overlaid on the
39 video. In one of the three scenes, a word-cue appeared in the textbox and disappeared again with the
40 end of the scene. Subjects were instructed to form a vivid association between the word and the precise

1 scene of the video-episode, such that they could later recall that exact scene and video-episode upon
2 presentation with the word-cue. We randomized the presentation of the associations in a balanced
3 way, such that no video-episode was presented more than twice in a row and a word-cue did not appear
4 in the same position more than twice in a row. Additionally, every position within every video-episode
5 was associated with a word cue once within 12 subsequent associations.

6 After each video-episode a fixation-cross showed for 1 second then subjects rated the plausibility of
7 the association between word-cue and scene. Three response options were labelled with “not
8 plausible”, “plausible” and “very plausible” and could be selected with the buttons 4, 5 and 6 on the
9 numerical pad of the keyboard. The plausibility rating served to keep participants engaged in the task
10 and support memory formation. In the distractor block, subjects were presented again for 45 seconds
11 with simple math problems and had to decide which one of two single digit sums was either bigger or
12 smaller (Figure 1b). For the retrieval block the word-cues were now randomized again in a balanced
13 way, such that word-cues corresponding to the same video-episode regardless of position, or to the
14 same position regardless of video-episode, did not appear more than twice in a row.

15 Retrieval blocks (Figure 1d, top-left) started with a fixation cross, displayed for 2 seconds. Then a word-
16 cue appeared in the center of the screen and the three framed numbers appeared on a triangle around
17 the word-cue. Participants were instructed to select, as quickly as possible, in which of the three scenes
18 they learned the word. For this choice they only saw the numbers 1, 2 and 3; after they made this
19 choice, screenshots forming the four video episodes appeared in the four corners of the screen.
20 Participants were asked to indicate now, to which of the four episodes the selected scene belonged.
21 The position of the numbers 1, 2 and 3 as well as the mappings of the four screenshot-sequences to
22 the screen positions were randomized in a balanced way, namely all possible permutations of 1, 2 and
23 3 were randomly mapped onto the three positions within 6 subsequent trials and all possible
24 permutations of the four positions of the video-episode screenshots were used within 24 trials. This
25 was done to control for any potential effects from specific screen positions on reaction times or position
26 specific response preferences. In order to respond, volunteers were asked to place the index finger of
27 their dominant hand on the number 5 of the numerical pad on the keyboard. The surrounding numbers
28 4, 6 and 8, which form a triangle around the number 5 were highlighted with red stickers and served as
29 the response options for the scene-response (first response: 1, 2 or 3). Those buttons corresponded
30 spatially to the position of the permuted numbers 1, 2 and 3 on the screen. Accordingly, the buttons 1,
31 7, 9 and 3 which form a square on the numerical pad, were available for the second response which
32 informed about the correct video-episode. Importantly subjects were instructed to make all responses
33 with the index finger of their dominant hand and go back to the starting position after every response,
34 i.e. leave the finger resting on the button 5. At the end of every retrieval trial, a scale appeared on which
35 subjects rated the confidence in their response. Three options were labelled with “guess”, “sure” and
36 “very sure” and corresponded to the buttons 4, 5 and 6 on the numerical pad.

37 Participants performed a variable amount of runs of encoding, distractor and retrieval blocks that varied
38 in length according to their individual memory performance. The first block comprised of 24 items,
39 subsequently its length was adjusted. If more than 70% of items were recalled correctly in the last block
40 (i.e. correct scene and movie were selected), 12 items were added to the next block, if less than 50%
41 were recalled correctly, 12 items were removed from the following block. All blocks comprised at least
42 of 12 associations that had to be learned. All participants completed 360 trials in total.

1 In the final behavioral experiment (Figure 1a, b, d top row) subjects performed exactly the same task
2 as in the behavioral pilot experiment, however, every other block was performed with a different
3 retrieval task. Specifically subjects performed the same learning paradigm, yet they did alternating
4 retrieval blocks of cued-recall (CR, see above) and associative recognition (AR). In the AR blocks (Figure
5 1d, top-right) subjects were presented with a screenshot of a single video-clip, representing one of
6 three scenes within a video-episode. The center of the screenshot was again superimposed with a
7 transparent textbox containing one of the previously learned word-cues. The association between
8 word-cue and video-clip could either be intact, i.e. the word was learned in this exact position within
9 the video-episode, or it could be rearranged. In the latter scenario, a different video-clip from the same
10 video-episode was superimposed by the word-cue. This means that word-cues were either presented
11 in the correct position or in the wrong position within the video-clip. Participants were again instructed
12 to decide as quickly as they could, whether the association was intact or rearranged. Block-size was
13 adjusted in the same way with percent of correct responses measured as $200 * (\text{Hits} - \text{False Alarms}) / N$,
14 with Hits being the number of correctly identified intact associations and False Alarms referring to the
15 number of rearranged associations that were declared intact and N referring to the number of trials in
16 the last block. Response buttons for the intact/rearranged choice were 4 and 6 on the numerical pad,
17 which are in equal distance from the number 5, where the index finger of participants' dominant hand
18 rested comfortably at the beginning of each trial. After the experiment participants answered a few
19 interview questions regarding eventual strategies and their subjective experience of the task.

20 **MEG experiment**

21 In the MEG experiment (Figure 1a, b, c) volunteers learned associations between video-episodes and
22 word-cues in the same way as in the behavioral experiment. Memory retrieval was similar to the
23 behavioral pilot experiment (i.e. a cued-recall task); however, a fast response was not required (see
24 below). Upon informed consent and screening questionnaires, participants received the instructions
25 for the task on a laptop outside the scanner. They familiarized themselves with the video-episodes
26 twice, paying close attention to their structure. It was ensured that every participant was able to
27 identify the three different scenes of a video-episode. In a short practice, they performed a block of
28 encoding, distractor and retrieval with the six example words. The head-localization coils of the MEG
29 system were attached to the participants' head and their positions were logged along with the shape
30 of the participant's head (see Data Collection). Subsequently, volunteers were seated in a comfortable
31 position under the MEG helmet. Subjects used a single button on each of two response pads with their
32 left and right index finger. After the eye tracker was mounted and calibrated, the experiment started.

33 The MEG experiment was again a sequence of encoding, distractor and retrieval blocks. In each
34 encoding block (Figure 1a), subjects learned a series of associations between scenes in video-episodes
35 and unique word-cues. Participants first saw a fixation-cross on the screen for 1 second. After that one
36 of the four video-episodes played for 6 seconds overlaid with a transparent textbox. In one of the three
37 scenes of the video-episode, the unique word-cue appeared in the textbox and disappeared again with
38 the end of the scene. The task was again to form a vivid association between the word and the precise
39 scene of the video-episode in order to later recall the exact scene and video-episode, when only
40 presented with the word-cue.

41 After the video-episode, the fixation-cross appeared again for 500ms. Finally, the two response options
42 'plausible' and 'not plausible' appeared on the left and right of the screen. Subjects used the left or

1 right button to indicate whether the association between video-scene and word-cue was plausible to
2 them. This task kept participants engaged and supported their memory performance.

3 The order of presentation was randomized in a balanced way: no video-episode was presented more
4 than twice in a row and a word-cue did not appear in the same position more than twice. Additionally,
5 every position within every video-episode was associated with a word cue once within 12 subsequent
6 associations. In the distractor block (Figure 1b) subjects solved simple math problems for 45 seconds:
7 They had to decide which one of two single digit sums was either bigger or smaller, using a left or right
8 button press. Participants received feedback on the distractor task in form of the words “correct” or
9 “wrong” displayed in green or red respectively. For the retrieval block the word-cues were now
10 randomized again in a balanced way, such that word-cues corresponding to the same video-episode
11 regardless of position, or to the same position regardless of video-episode, did not appear more than
12 twice in a row.

13 Trials of the retrieval block (Figure 1c) started with a fixation cross that was displayed for 1 second.
14 Then a word-cue appeared in the center of the screen for 3.5 seconds. In this time interval subjects
15 remembered in which exact scene they had seen this word. For a random time interval between 250ms
16 and 750ms, a fixation cross was shown again, then the response options appeared. The time interval
17 for retrieval was chosen based on reaction-time data from the behavioral experiments, such that
18 participants could comfortably remember the correct association. The first response option required
19 the selection of the correct scene. To this end pictograms featuring the numbers 1, 2 and 3 were
20 displayed on the top of the screen. The mapping of the numbers 1, 2 and 3 to the three screen-positions
21 was randomized in a balanced way such that all possible permutations appeared within 6 subsequent
22 trials. Participants could now move a red square, which framed the current selection. By pressing the
23 left button they changed their selection by moving the frame clockwise. This selection was confirmed
24 by pressing the right button. Note that this button assignment ensured that subjects would always
25 prepare the same response during the retrieval trial, regardless of the memory content. This is
26 important to control for trivial but systematic differences that correlate with memory content in the
27 retrieval interval.

28 After the position was selected, the two other position pictograms were overlaid with transparency
29 ($\alpha = 0.9$), such that the selected option remained highlighted on the screen. The concatenated
30 screenshots from the video-episodes appeared below the position-pictograms and the red selection
31 frame could be moved clockwise with the left button. Again, the selection was confirmed with the right
32 button. To ensure that subjects followed instructions and tried to recall the position as soon as they
33 were presented with the word-cue (and did not wait until the response options were presented), a time
34 limit of 4 seconds was imposed to select the correct position and again to select the movie. To allow
35 for flexibility due to hasty or imprecise selections, 200ms were added to this time limit, whenever the
36 selection-frame was moved. Participants did not know about this increment. If the time limit was
37 exceeded, the message ‘too slow’ appeared at the center of the screen for 5 seconds. Altogether the
38 time limits were designed, such that subjects could comfortably remember the correct association
39 during the presentation of the word-cue, and were eager to select the two responses straight away.
40 After the associated video-episode was selected, unselected response options were overlaid with
41 transparency for 300ms, then the two options ‘guess’ and ‘know’ were presented on a new screen to
42 give the participant the opportunity to communicate, whether the selected answers were based on a
43 guess.

1 Participants performed a variable amount of runs of encoding, distractor and retrieval blocks. The
2 blocks varied in length according to individual memory performance. The first block comprised of 24
3 items, subsequently its length was adjusted. If more than 90% of items were recalled correctly in the
4 last block (i.e. correct scene and movie were selected), 24 items were added to the next block. If more
5 than 70% of items were recalled correctly, 12 items were added to the next block, if less than 50% were
6 recalled correctly, 12 items were removed from the following block, if less than 40% were recalled
7 correctly, 24 items were removed. All blocks comprised at least 12 associations that had to be learned,
8 i.e. block-size was never reduced below 12 items; all participants learned and recalled a total of 360
9 associations.

10 Data Collection

11 Stimulus presentation and the collection of behavioral data was realized on a standard desktop
12 computer running MATLAB 2014b (MathWorks) under Windows 7, 64 Bit version. Stimuli were
13 presented through the Psychophysics Toolbox Version 3³⁴. In the behavioral experiments, responses
14 were collected from button presses on the numerical pad of a wired keyboard (Model 1576, Microsoft
15 Corporation, Redmond, US). In the MEG experiment, fiber optic response pads were used.

16 Neurophysiological data were collected with 275-channel CTF MEG (CTF, Coquitlam, BC, Canada) at the
17 Sir Peter Mansfield Imaging Center (SPMIC) in Nottingham, UK. The system was used in third-order
18 gradiometer configuration, recording at a sampling frequency of 600 Hz over the whole duration of the
19 experiment. Three localization coils that were attached to the participants' left preauricular point (LPA),
20 right preauricular point (RPA) and to a point slightly above the nasion (NAS) were energized during the
21 recording session. This was done to localize the head position relative to the sensors.

22 Head digitization was collected with a Polhemus ISOTRAK device (Colchester, Vermont, USA). A
23 minimum of 500 points on the scalp were logged relative to the positions of the three fiducial points
24 (LPA, RPA, NAS). Individual anatomical data was acquired via magnetic resonance imaging (MRI) (3T
25 Achieva scanner; Philips, Eindhoven, the Netherlands) with an MPRAGE sequence covering the whole
26 head at 1mm³ resolution. MRIs were either measured at the SPMIC or at the Centre for Human Brain
27 Health at the University of Birmingham (CHBH).

28 For 17 of the included subjects (23), eye tracking (EyeLink 1000 Plus, SR Research, Ontario, Canada) was
29 recorded on a separate Computer provided by the manufacturer at a sampling rate of 2000 Hz. The
30 data was additionally written into 3 analog input channels of the MEG system via the EyeLink Analog
31 Card. The eye tracker was used in remote mode tracking the pupil and corneal reflection with a 16mm
32 lens. It was calibrated and validated using 13 points on 80% of the screen, which contained all of the
33 task relevant information.

34 Analysis of Reaction Times

35 We defined reaction time (RT) as the time to the first response after onset of the word-cue (Figure 1d,
36 bottom-row). All RTs faster than 200ms were considered implausible and discarded from further
37 analysis. Additionally, RTs that were 2.5 standard deviations above the mean RT were discarded. The
38 means of remaining RTs were then tested statistically. Data distribution of reaction times was assumed
39 to be normal but not formally tested. Individual data points are shown in Figure 1d. To account for

1 potentially non-normal distribution of RTs ³⁵, all statistical tests are also reported for log-transformed
2 RTs.

3 Preprocessing of Neural Data

4 The data was preprocessed in MATLAB 2015a (MathWorks) with a combination of functions from the
5 Fieldtrip toolbox for EEG/MEG analysis ³⁶ and custom written scripts.

6 For the sensor level analysis, the 3rd order gradiometer correction was first applied, then the continuous
7 recording was filtered with a Butterworth IIR filter of 4th order with a stopband of 49.5 to 50.5 and its
8 harmonics (99.5 - 100.5, 149.5 - 150.5, 199.5 - 200.5, and 249.5 - 250.5) to reduce the line noise artifact.
9 Additionally, the data was filtered with a stopband of 59 – 60 to attenuate noise with a center frequency
10 of 59.5 Hz.

11 Subsequently, the data was segmented into trials that started 1.5 seconds prior to video-onset and
12 ended 7.5 seconds after video-onset at encoding. Trials at retrieval started 1.5 seconds prior to the
13 onset of the word-cue and ended 5 seconds after onset of the word-cue. If available, the dataset was
14 combined with the downsampled and segmented trials from the eye tracking.

15 To remove activity from eye blinks and noise, and to detect heartbeats, Independent Component
16 Analysis (ICA) was used ³⁷. For the computation of the ICA unmixing matrix, trials containing coarse
17 artifacts or showing strong muscle activity were heuristically excluded. Additionally, the data was
18 downsampled to 250 Hz and cut to 1-second-long segments; the obtained unmixing matrix was then
19 applied to all original trials.

20 When possible, we compared independent components with the eye tracking data; we removed those
21 components that picked up eye-blinks or eye-movement related activity. Additional components that
22 picked up electrical noise were removed from the data. A copy of components which contained a clear
23 R-wave of the QRS complex in a heartbeat was stored for later peak-detection and regression. All
24 remaining components were projected back to a channel representation.

25 Finally, all data was inspected visually and trials containing artifacts were removed from later analysis.
26 After visual inspection, 84.26 % (S.D. = 8.29 %) of trials remained.

27 Heartbeats were removed with a regression based approach: An iterative peak detection algorithm was
28 applied to the ICA-component showing the clearest R-wave; it served as a proxy for ECG. This was done
29 only for the remaining trials after visual inspection. Before peak-detection the heartbeat-component
30 was highpass-filtered (4Hz, 4th order Butterworth). The peak detection algorithm first calculated a
31 plausible maximum of heartbeats that were not to be exceeded. The signal was z-scored and
32 thresholded. Local peaks were detected by finding local maxima in clusters of z-scores that were above
33 threshold. Subsequently the threshold was lowered stepwise, down to a z-score of 2. With lowering
34 threshold, increasingly bigger areas around the peaks were excluded from further peak detection. If
35 the maximum number of plausible peaks was exceeded, the threshold was no longer lowered. A
36 heartbeat template was now created by averaging 500ms long segments around the peaks. Gaps in the
37 continuous recording were subsequently zero-padded in order to convolve the component with the
38 template. Peak detection was then repeated on the convolved time course and a new template was
39 built from these peaks for subsequent convolution ³⁸. After a few repetitions, the template converged
40 and the resulting peaks were controlled manually, even though errors rarely needed to be corrected.

1 Instead of simply subtracting the averaged template from the data, the trials were now split into four
2 big segments and a general linear model (GLM) was built around the peaks in each segment. A high
3 pass filter (1Hz, 4th order Butterworth) was applied to the data, only for the purpose of fitting the model.
4 The GLM consisted of a separate repeated measure factor for each time point in the heartbeat,
5 beginning 280ms before the peak and ending 720ms after the peak ³⁸. Additionally, a separate factor
6 was included for every heartbeat, which modelled the offset between 280ms pre-peak and 720ms post-
7 peak. Furthermore an offset factor for the overall segment was included. The solved model was then
8 applied to every channel. The data model \hat{y} was built by using only the repeated measure factors, which
9 modelled each time point within the heartbeat (i.e. the beta weights for offsets were set to 0). After
10 visual inspection, this resulting model of the heartbeat was subtracted from each original channel.

11 For the source level analysis, the anatomical data was first aligned to the digitized head positions. This
12 was done by extracting the surface of the head from the anatomical MRI; in a first step a rough
13 alignment was done manually, then the Iterative Closest Point (ICP) algorithm implemented in fieldtrip
14 ³⁶ was used to match the surface to the point-cloud of the head digitization, finally this solution was
15 controlled and eventually corrected again manually. The transformation to the aligned space was
16 subsequently applied to the segmentation of the brain, which was likewise extracted from the
17 anatomical images. To correct for head movements, the average head positions within the trials were
18 first clustered, such that one positional-cluster was built for every 10 trials. Subsequently a separate
19 lead field was computed for every cluster and then averaged. Hereby, an average lead field across all
20 trials was obtained for each participant ³⁹. Importantly 'all trials' refers to the trials that were included
21 in a given contrast (e.g. for the contrast of Hits and Misses at retrieval, encoding trials were not included
22 in the computation of the lead field). Before the source level analysis, the 3rd order gradiometer
23 correction was applied to the cut raw-data, lead fields were adjusted accordingly. Finally, the data was
24 demeaned and bandpass filtered between 4 and 15 Hz. The position of virtual sensors in individual
25 brains was derived from a 1 cm spaced grid, which was placed 6mm below the surface of the cortex
26 into the MNI brain and then spatially warped into individual brains. This was done via the inverse of the
27 transformation describing their normalization and resulted in 1407 individual virtual sensor positions
28 which were anatomically equivalent. Finally, to reconstruct activity on virtual sensors a regularized
29 linearly constrained minimum variance (lcmv) beamforming approach, implemented in the Fieldtrip
30 toolbox ³⁶, was used. Filter coefficients were again computed on all data in a given contrast.

31 Analysis of oscillatory power

32 To estimate oscillatory power at retrieval (Supplementary Figure 1), the Fourier-transformed data was
33 multiplied with a complex Morlet wavelet of six cycles. This was done in steps of 10ms for every full
34 frequency between 2 and 40Hz. The raw power was then obtained from the squared amplitude of the
35 Fourier spectrum. Across all trials within the contrast (i.e. Hits and Misses), a baseline was computed
36 as the average power between 1 second pre-stimulus and 4 second after stimulus onset ⁴⁰. Trials were
37 then normalized by subtracting the baseline and dividing by it ($\text{activity}_{tf} - \text{baseline}_f$)/ baseline_f , with t
38 indexing time and f indexing frequency.

39 Region of Interest (ROI)

40 An occipito-parietal region of interest (ROI) was derived from the AAL atlas ⁴¹ (Figure 2b). To obtain the
41 ROI in form of a group of virtual sensors, the sensor-positions in MNI-space were assigned to the

1 nearest described AAL-region, based on their Euclidean distance. The occipito-parietal ROI comprised
2 of bilateral AAL-regions: angular gyrus, calcarine sulcus, cuneus, inferior occipital cortex, inferior
3 parietal lobule, lingual gyrus, middle occipital gyrus, precuneus, superior occipital gyrus, superior
4 parietal lobule, supramarginal gyrus.

5 Content specific oscillatory phase at encoding

6 During encoding, participants repeatedly watched the same video-episodes. Hence, it was possible to
7 assess content specific properties if they were more similar between trials of same content than
8 between trials of different content (Figure 2a). In order to determine whether the ongoing oscillatory
9 phase was specific to individual perceptual content, trials were grouped into 4 sets according to the
10 video-episode that was perceived. The complex Fourier spectrum was again derived by multiplying the
11 Fourier-transformed data with a complex Morlet wavelet of six cycles. Then, inter-trial phase coherence
12 ⁴² (ITPC) was computed across the trials of same content (i.e. for each of the four trial-groups). This was
13 done at every full frequency between 2 and 40 Hz in steps of 10ms starting 1 second before the onset
14 of the video-episodes and ending 7 seconds after the offset of the video-episodes. Following that, the
15 trials were shuffled and grouped randomly into 4 sets of mixed-content-trials. Sets were of equal size
16 to the 4 sets of same-content-trials. Again, ITPC was computed separately for each of the 4 sets. To
17 balance the contribution of the 4 sets, a Rayleigh Z-correction was applied with $N \cdot \text{ITPC}^2$, where N refers
18 to the number of trials in a set. Finally, the corrected ITPC was averaged across the 4 sets in the ordered
19 and in the shuffled condition. Their difference indicated content specificity of phase which could be
20 statistically tested ^{17,43}. The analysis in source-space was done in the same way using the virtual sensors;
21 however, the frequency was restricted to 8 Hz.

22 Content specific phase similarity between encoding and retrieval

23 The reactivation of temporal patterns (Figure 2c-d, Supplementary Figure 3) was estimated on virtual
24 sensors for the frequency of 8 Hz. To this end, the oscillatory phase coherence between encoding and
25 retrieval was contrasted between trial-combinations of same content (e.g. watching video-episode A,
26 recalling video-episode A) and random trial-combinations of different content (e.g. watching video-
27 episode A, recalling video-episode B). The combinations were balanced, such that in both conditions
28 (same vs. different combinations) exactly the same trials were used in the same amount of
29 combinations. We only changed the pairing between encoding and retrieval trials. For each trial-
30 combination, 1-second long windows from the encoding trial were now compared to every time point
31 at retrieval starting at the onset of the word-cue and ending at its offset after 3.5 seconds. This
32 comparison was done with a sliding window approach. As a metric of phase-similarity, the phase
33 coherence across time ^{2,19,20} (i.e. across the 1 second window) was computed. All possible windows
34 from encoding were used in this sliding window approach, with the first window ranging from 0 to 1
35 seconds and the last window ranging from 5 to 6 seconds during the video-episode (compare Figure 2
36 c). Note that the response options set on between 250ms and 750ms after the word-offset, additionally
37 the first response-screen did not contain content-information (only the numbers 1, 2, and 3) and all
38 responses required a button-press on the left button. Therefore, no confounds from the response
39 interval were expected to bleed into the tested interval. Oscillatory phase was estimated by multiplying
40 the Fourier-transformed data with a complex Morlet wavelet of six cycles in steps of 15.6ms for
41 consistency with our previous analyses ². The average similarity between all time-windows and
42 combinations was subsequently averaged to derive a single value of similarity for combinations of same

1 content and a single value for combinations of different content at each virtual sensor. Note that this
2 method enables the investigation of highly dynamic patterns in a robust way, because a measure that
3 captures dynamic changes in ongoing oscillations is accumulated across encoding time, retrieval time
4 and ten thousands of trial-combinations.

5 Time courses of Replay

6 To observe the temporal scale of reactivation (Figure 3), the distribution of similarity to the
7 remembered stimulus content (i.e. phase coherence) across retrieval was compared between different
8 sliding windows from encoding. By definition, a distribution is normalized to an area under curve of 1
9 and therefore accounts for differences in total similarity between windows. To robustly compare the
10 distribution of similarity between 6 non-overlapping windows, phase-coherence was accumulated
11 across time, such that at the beginning of the retrieval time, zero similarity to all windows was present
12 and at the end of retrieval (i.e. at 3.5 seconds after word onset) 100 % of similarity was reached (Figure
13 3c). This made it possible to compare at each time point, whether the similarity to a window had come
14 up earlier than to another window. In other words: If patterns from window “A” tend to appear earlier
15 than patterns from window “B” across subjects, then the cumulated similarity to window A should be
16 statistically higher than the cumulated similarity to window “B”, at several time points.

17 In order to test for a general tendency for forward replay, a line was fitted across all 6 windows and
18 tested against a slope of 0 (Figure 3d). Hence a negative slope of this line means that earlier windows
19 from encoding appear earlier during retrieval. In order to test the hypothesis that the replay of
20 individual scenes takes place on a slower timescale (Figure 3e), 3 lines were fitted across the 2 non-
21 overlapping windows within each scene, and their slope was averaged. A more negative average slope
22 of these 3 lines compared the slope of the line across all windows supports the hypothesis that
23 replaying individual scenes takes place on a slower temporal scale.

24 Importantly this way of cumulating the similarity distributions allows for robust testing across subjects
25 at the expense of introducing temporal dependencies between time points. Specifically, if more
26 similarity to a window is present at an early point this can propagate to later points, if similarity
27 thereafter increases at the same speed for all windows. The extent of significant time intervals should
28 therefore be interpreted with caution. Another disadvantage of this method is that the slope is interval
29 scaled and its absolute value is not interpretable.

30 In order to compensate for this disadvantage and quantify the actual lag between time windows from
31 encoding descriptively (Figure 3a-b), the distributions of similarity were averaged across subjects and
32 smoothed with a moving average kernel of 250ms, to attenuate noise (Figure 3a, right). The cross-
33 correlation between distributions was then computed to estimate the lag between them: The shape of
34 one similarity distribution is matched to another (Figure 3b). This was done within the time interval in
35 which the slowing down of replay was observed; specifically, in which the slope for lines fitted within a
36 scene was significantly more negative than the slope across all windows (i.e. between 550ms and
37 2350ms at retrieval).

38 Statistical analyses

39 **Behavioral performance and Reaction times**

1 Behavioral performance was tested with a repeated-measures ANOVA, on the percent of correct
2 responses. Partial eta-square (η^2) was calculated as a measure of effect-size. Greenhouse-Geisser
3 correction was used with all repeated-measures ANOVAs. Post-hoc tests were then performed with 2
4 separate repeated-measures ANOVAs for the final behavioral experiment and with a series of one-
5 sample two-tailed t-test (see Supplementary Information). T-tests were one-tailed if specific
6 hypotheses were tested, t-tests were two-tailed, if no assumption about the direction of effects was
7 made. All confidence intervals were derived via bootstrapping with 10,000 iterations ⁴⁴. Data
8 distribution of percent correct responses was assumed to be normal but not formally tested. Individual
9 data points of behavioral performance are shown in Supplementary Figure 8.

10 RTs in the balancing pilots (Supplementary Information) were first contrasted with two-tailed one-
11 sample t-tests. In order to statistically test the null hypothesis the Scaled JZS Bayes Factor ⁴⁵ to the one-
12 sample t-tests was contrasted against a prior effect size of 0.707. RTs in the behavioral pilot experiment
13 were compared with a repeated-measures ANOVA with the factor position (1, 2 and 3). In the final
14 behavioral experiment, a 2x3-repeated-measures ANOVA was computed with the factors retrieval task
15 (cued-recall vs. associative recognition) and position (1, 2 and 3). Post-hoc tests were then performed
16 with 2 separate repeated-measures ANOVAs. Reaction times for the 3 different positions were
17 subsequently compared with a series of post-hoc one-sample t-tests. All confidence intervals were
18 derived via bootstrapping with 10,000 iterations ⁴⁴. Greenhouse-Geisser correction was used with all
19 repeated-measures ANOVAs, null-effects of interest were tested with Bayesian t-tests against a prior
20 effect size of 0.707 ⁴⁵.

21 **Content specific oscillatory phase at encoding**

22 Content specific phase at encoding was statistically tested by contrasting average ITPC across arranged
23 groups with the average ITPC across shuffled groups. This was done with a series of one-tailed t-test at
24 every time point between 0 and 6 seconds after onset of the video-episode, at every frequency
25 between 2 and 40 Hz and at every sensor. Multiple comparison correction was done via Monte-Carlo
26 permutation of contrast labels as implemented in the fieldtrip toolbox ^{36,46}, treating each subject as a
27 unit of observation. 3-dimensional clusters and cluster-sums were formed across time, frequency and
28 sensors. Cluster-sums in the original contrast were compared to the distribution of cluster-sums under
29 random label assignment in order to derive p-values. The cluster-forming threshold corresponded to
30 the critical t-value ($\alpha < 0.05$) of a single-sided one-sample t-test, 1000 random permutations were
31 drawn. On the source level, content specific phase was assessed for the frequency of 8Hz. Again, the
32 ITPC of arranged groups and the ITPC of shuffled groups were contrasted with a one sample t-test that
33 was computed at every time point and every virtual sensor. Clusters were summed across neighboring
34 sensors and time points in 1000 random permutations. To obtain time courses within the parieto-
35 occipital ROI, t-values were averaged across all virtual sensors within the ROI.

36 **Content specific phase similarity between encoding and retrieval**

37 Based on previous results ², statistical testing for content specific reactivation was done for the
38 frequency of 8 Hz, restricted to an occipito-parietal region of interest (ROI) derived from the AAL atlas
39 ⁴¹. Averaged similarity values of encoding-retrieval combinations were contrasted between
40 combinations of same content and combinations of different content. This was done with a one-sample
41 t-test on every virtual sensor within the ROI. Subsequently t-values were thresholded with a t-value
42 corresponding to a one-sided alpha value of 0.05; clusters were built across neighboring virtual sensors.

1 Statistical testing was done again via 1000 random permutations using the Monte-Carlo method
2 implemented in the fieldtrip toolbox ³⁶ and treating each subject as the unit of observation. Cluster-
3 sums in the original contrast were compared to the distribution of cluster-sums under random label
4 assignment in order to derive p-values. A series of post-hoc t-tests was done on every time-point at
5 retrieval in order to estimate the contribution to the effect from encoding windows (see Supplementary
6 Information, specifically Supplementary Figure 3a).

7 **Time courses of replay**

8 Time courses were obtained by averaging across the ROI, which allows for an unbiased investigation of
9 the time-courses of reactivation (see Supplementary Information, specifically Supplementary Figure
10 2a). Specifically, the cluster correction approach results in a biased noise-distribution within the cluster
11 of significant reactivation. This renders the interpretation of its shape and any post-hoc analysis on
12 sensors within the cluster problematic ⁴⁶, see also ⁴⁷. Since 86.46% of the t-values in the ROI were
13 positive, we therefore decided to average across all virtual sensors within the anatomical ROI for the
14 analyses of all time courses that were statistically tested.

15 Likewise, similarity densities were computed on the averaged similarity values across all virtual sensors
16 within the ROI. The cumulated similarity density distributions for 6 non-overlapping encoding-windows
17 were obtained for every subject. Consequently, at every retrieval time-point a line could be fitted across
18 6 values for every subject. The slope of that line was subsequently subjected to a two-sample t-test
19 against 0 across all subjects. Data distribution of slope was assumed to be normal but not formally
20 tested. Individual data points of slope are shown in Supplementary Figure 8. The resulting time-course
21 of t-values across the whole retrieval time was finally subjected to a multiple comparisons correction
22 by controlling the false discovery rate ²⁵. To compare the speed of replay within scenes, to the overall
23 speed, the average slope fitted across two windows each (windows within scenes) was statistically
24 tested against the slope across all encoding windows with a series of one-tailed one-sample t-tests. T-
25 values were obtained again at every time point during retrieval and the false discovery rate was
26 controlled in order to correct for multiple comparisons. To estimate at which time-points reinstatement
27 could be detected best (Supplementary Information, Figure 2a), a series of one-tailed one-sample t-
28 tests was computed at every retrieval time point, between encoding-retrieval similarity of same content
29 combinations and encoding-retrieval combinations of different content combinations (see
30 Supplementary Information).

31 Finally, the average similarity to all encoding time points was compared within the ROI, between trials
32 in which an association from the first, second or third scene was recalled (Supplementary Information,
33 Figure 2b). This was done with a repeated-measures ANOVA with the factor position and pairwise post-
34 hoc one-tailed one-sample t-tests.

35 **Oscillatory power**

36 Baseline corrected oscillatory power was contrasted on the sensor level with a series of one-sample t-
37 tests (see Supplementary Information). Multiple-comparison correction was realized with a cluster-
38 based Monte-Carlo permutation as implemented in the fieldtrip toolbox ³⁶. 1000 permutations of
39 contrast-labels were used; the clusters were formed from neighboring values below a threshold (see
40 below). Neighboring values were derived across time from 0 to 4 seconds after the onset of the word-
41 cue, across frequency from 2 to 40 Hz and spatially across sensors. The threshold was the t-value which

1 corresponds to a threshold of $\alpha = 0.05$ for a single sided test. The maximal cluster-sum of real data
2 was then compared to the distribution of maximal cluster-sums under random permutations in order
3 to derive a p-value. In order to find the most robust frequencies that showed oscillatory power
4 decreases, a one-tailed t-test was computed for the average power difference across time (0 – 4s),
5 sensors and frequencies. On the source level, baseline-corrected power at 8 Hz was averaged over time
6 between 0 and 4 seconds and subjected to a one-sample t-test. Multiple comparison correction was
7 addressed with the same cluster-based permutation approach; however, clusters were formed across
8 neighboring virtual sensors.

1 **Acknowledgements**

2 The authors would like to thank the Sir Peter Mansfield Imaging Centre (SPMIC), specifically Dr. George
3 O'Neill, Dr. Benjamin A.E Hunt and Dr. Lauren Gascoyne for their help with data collection. This work
4 was supported by the ERC Grant Code4Memory (647954) awarded to S.H., who is further supported by
5 the Wolfson Society and the Royal Society. B.P.S. is supported by a Sir Henry Dale Fellowship jointly
6 funded by the Wellcome Trust and the Royal Society (107672/Z/15/Z). The funders had no role in study
7 design, data collection and analysis, decision to publish or preparation of the manuscript.

8 **Competing interests**

9 The authors declare no competing interest.

10 **Author Contributions**

11 Conceived and designed the experiments: SM BPS SH.

12 Performed the experiments: SM.

13 Analyzed the data: SM under supervision of SH.

14 Contributed reagents/materials/analysis tools: SM SH HB.

15 Wrote the paper: SM SH commented and edited by BPS HB.

16 **Code availability**

17 Analysis scripts of this project are deposited in a public repository
18 (<https://doi.org/10.25500/eData.bham.00000254>).
19

20 **Data availability**

21 Group statistical data of this project is deposited in the Dryad Digital Repository
22 (<https://doi.org/10.25500/eData.bham.00000254>). The data that support the findings of this study
23 will be available from the corresponding author upon request.

1 References

- 2 1. Tulving, E. What is episodic memory? *Curr. Dir. Psychol. Sci.* **2**, 67–70 (1993).
- 3 2. Michelmann, S., Bowman, H. & Hanslmayr, S. The Temporal Signature of Memories:
4 Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLoS Biol.* **14**,
5 (2016).
- 6 3. Michelmann, S., Bowman, H. & Hanslmayr, S. Replay of Stimulus-specific Temporal Patterns
7 during Associative Memory Formation. *J. Cogn. Neurosci.* **30**, 1577–1589 (2018).
- 8 4. Arnold, A. E. G. F., Iaria, G. & Ekstrom, A. D. Mental simulation of routes during navigation
9 involves adaptive temporal compression. *Cognition* **157**, 14–23 (2016).
- 10 5. Bonasia, K., Blommestejn, J. & Moscovitch, M. Memory and navigation: Compression of space
11 varies with route length and turns. *Hippocampus* **26**, 9–12 (2016).
- 12 6. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place
13 cells during the awake state. *Nature* **440**, 680–683 (2006).
- 14 7. Carr, M. F., Jadhav, S. P. & Frank, L. M. Hippocampal replay in the awake state: A potential
15 substrate for memory consolidation and retrieval. in *Nature Neuroscience* **14**, 147–153 (2011).
- 16 8. Yaffe, R. B., Shaikhouni, A., Arai, J., Inati, S. K. & Zaghoul, K. A. Cued Memory Retrieval
17 Exhibits Reinstatement of High Gamma Power on a Faster Timescale in the Left Temporal Lobe
18 and Prefrontal Cortex. *J. Neurosci.* **37**, 4472–4480 (2017).
- 19 9. Staudigl, T., Vollmar, C., Noachtar, S. & Hanslmayr, S. Temporal-Pattern Similarity Analysis
20 Reveals the Beneficial and Detrimental Effects of Context Reinstatement on Human Memory. *J.*
21 *Neurosci.* **35**, 5373–5384 (2015).
- 22 10. Zhang, H. et al. Gamma Power Reductions Accompany Stimulus-Specific Representations of
23 Dynamic Events. *Curr. Biol.* **25**, 635–640 (2015).
- 24 11. Wimber, M., Maaß, A., Staudigl, T., Richardson-Klavehn, A. & Hanslmayr, S. Rapid memory
25 reactivation revealed by oscillatory entrainment. *Curr. Biol. CB* **22**, 1482–6 (2012).
- 26 12. Chen, J. et al. Shared memories reveal shared structure in neural activity across individuals. *Nat.*
27 *Neurosci.* **20**, 115–125 (2017).

- 1 13. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems
2 neuroscience. *Front. Syst. Neurosci.* **2**, 1–28 (2008).
- 3 14. Staresina, B. P. et al. Hippocampal pattern completion is linked to gamma power increases and
4 alpha power decreases during recollection. *eLife* **5**, (2016).
- 5 15. Yaffe, R. B. et al. Reinstatement of distributed cortical oscillations occurs with precise
6 spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci.* **111**, 18727–
7 18732 (2014).
- 8 16. Kurth-Nelson, Z., Barnes, G., Sejdinovic, D., Dolan, R. & Dayan, P. Temporal structure in
9 associative retrieval. *eLife* **4**, e04919 (2015).
- 10 17. Ng, B. S. W., Logothetis, N. K. & Kayser, C. EEG Phase Patterns Reflect the Selectivity of
11 Neural Firing. *Cereb. Cortex* **23**, 389–398 (2013).
- 12 18. Schyns, P. G., Thut, G. & Gross, J. Cracking the Code of Oscillatory Activity. *PLoS Biol.* **9**,
13 e1001064 (2011).
- 14 19. Lachaux, J.-P. et al. Studying single-trials of phase-synchronous activity in the brain. *Int. J.*
15 *Bifurc. Chaos* **10**, 2429–2439 (2000).
- 16 20. Mormann, F., Lehnertz, K., David, P. & E. Elger, C. Mean phase coherence as a measure for
17 phase synchronization and its application to the EEG of epilepsy patients. *Phys. Nonlinear*
18 *Phenom.* **144**, 358–369 (2000).
- 19 21. Ji, D. & Wilson, M. A. Coordinated memory replay in the visual cortex and hippocampus during
20 sleep. *Nat. Neurosci.* **10**, 100–107 (2007).
- 21 22. Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C. & De Lange, F. P. Shared representations for
22 working memory and mental imagery in early visual cortex. *Curr. Biol.* **23**, 1427–1431 (2013).
- 23 23. Ekman, M., Kok, P. & de Lange, F. P. Time-compressed preplay of anticipated events in human
24 primary visual cortex. *Nat. Commun.* **8**, 15276 (2017).
- 25 24. Bosch, S. E., Jehee, J. F. M., Fernandez, G. & Doeller, C. F. Reinstatement of Associative
26 Memories in Early Visual Cortex Is Signaled by the Hippocampus. *J. Neurosci.* **34**, 7493–7500
27 (2014).

- 1 25. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
2 approach to multiple testing. *J. R. Statistic Soc. Ser. B* **57**, 289–300 (1995).
- 3 26. Radvansky, G. A. & Zacks, J. M. Event boundaries in memory and cognition. *Curr. Opin. Behav.*
4 *Sci.* **17**, 133–140 (2017).
- 5 27. Sols, I., DuBrow, S., Davachi, L. & Fuentemilla, L. Event Boundaries Trigger Rapid Memory
6 Reinstatement of the Prior Events to Promote Their Representation in Long-Term Memory. *Curr.*
7 *Biol.* **27**, 3499-3504.e4 (2017).
- 8 28. Davachi, L. & DuBrow, S. How the hippocampus preserves order: the role of prediction and
9 context. *Trends Cogn. Sci.* **19**, 92–99 (2015).
- 10 29. Buzsáki, G. & Tingley, D. Space and Time: The Hippocampus as a Sequence Generator. *Trends*
11 *Cogn. Sci.* **22**, 853–869 (2018).
- 12 30. Johnson, A. & Redish, A. D. Neural Ensembles in CA3 Transiently Encode Paths Forward of the
13 Animal at a Decision Point. *J. Neurosci.* **27**, 12176–12189 (2007).
- 14 31. Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W. & Duzel, E. Replay of very early
15 encoding representations during recollection. *J. Neurosci. Off. J. Soc. Neurosci.* **34**, 242–8
16 (2014).
- 17 32. Coltheart, M. The MRC psycholinguistic database. *Q. J. Exp. Psychol. Sect. A* **33**, 497–505
18 (1981).
- 19 33. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: A critical evaluation of current
20 word frequency norms and the introduction of a new and improved word frequency measure for
21 American English. *Behav. Res. Methods* **41**, 977–990 (2009).
- 22 34. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
- 23 35. Ratcliff, R. Group reaction time distributions and an analysis of distribution statistics. *Psychol.*
24 *Bull.* **86**, 446–461 (1979).
- 25 36. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for
26 Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell.*
27 *Neurosci.* **2011**, 1–9 (2011).

- 1 37. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG
2 dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
- 3 38. Tal, I. & Abeles, M. Cleaning MEG artifacts using external cues. *J. Neurosci. Methods* **217**, 31–
4 38 (2013).
- 5 39. Stolk, A., Todorovic, A., Schoffelen, J. M. & Oostenveld, R. Online and offline tools for head
6 movement compensation in MEG. *NeuroImage* **68**, 39–48 (2013).
- 7 40. Long, N. M., Burke, J. F. & Kahana, M. J. Subsequent memory effect in intracranial and scalp
8 EEG. *NeuroImage* **84**, 488–494 (2014).
- 9 41. Tzourio-Mazoyer, N. et al. Automated anatomical labeling of activations in SPM using a
10 macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–
11 289 (2002).
- 12 42. Tallon-Baudry, C., Bertrand, O., Delpuech, C. & Pernier, J. Stimulus specificity of phase-locked
13 and non-phase-locked 40 Hz visual responses in human. *J. Neurosci. Off. J. Soc. Neurosci.* **16**,
14 4240–9 (1996).
- 15 43. Busch, N. A., Dubois, J. & VanRullen, R. The Phase of Ongoing EEG Oscillations Predicts
16 Visual Perception. *J. Neurosci.* **29**, 7869–7876 (2009).
- 17 44. Hentschke, H. & Stüttgen, M. C. Computation of measures of effect size for neuroscience data
18 sets. *Eur. J. Neurosci.* **34**, 1887–1894 (2011).
- 19 45. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for
20 accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).
- 21 46. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci.*
22 *Methods* **164**, 177–190 (2007).
- 23 47. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. & Baker, C. I. Circular analysis in systems
24 neuroscience: The dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
- 25
26

1 Figure 1 Experimental design and behavioral results

2 (a) During encoding subjects repeatedly saw one out of four video-episodes. In one of three scenes that
3 comprise a video-episode, a word-cue appeared in the center of the screen. (b) In the distractor block
4 participants identified either the bigger or the smaller one of 2 simple sums. (c) In the MEG experiment,
5 participants saw the static word-cue during retrieval for 3.5 seconds, followed by a fixation cross for
6 250ms - 750ms. Subsequently they first picked the scene-position in which they learned the association
7 and then confirmed the correct video-episode. (d) In the cued-recall (CR) task of the behavioral
8 experiment (left) 24 participants selected the correct scene position as quickly as possible during
9 retrieval. In an associative-recognition (AR) control task (right) they decided whether the presented
10 association (word superimposed on a screenshot) was intact or rearranged. In CR blocks, subjects were
11 faster to recall an association that was learned in earlier scene-positions during encoding (bottom left).
12 Importantly, in the control task, they performed the same encoding task and needed source memory
13 for AR retrieval, however no modulation of reaction times was found. The y-axis denotes the difference
14 to each participant's average reaction time in the respective task. Spaghetti-plots show individual
15 subjects. Boxplots are 25th and 75th percentile and the median; whiskers are maxima and minima,
16 excluding outliers. Red dots within the boxplots depict the arithmetic mean. Significant differences are
17 marked with a star and denote a significant one-tailed $t_{23} = -1.870$, $p = 0.037$, $CI = [-\infty -10ms]$, Cohen's
18 $d = 0.382$ (left) and $t_{23} = -2.767$, $p = 0.006$, $CI = [-\infty -67ms]$, Cohen's $d = 0.565$ (right), n.s. denotes non-
19 significant in a post-hoc paired t-test comparison ($ps > 0.199$).

20

1 Figure 2: Reinstatement of oscillatory patterns from encoding

2 (a) During encoding, the different video-episodes elicited content specific phase patterns. The left panel
3 shows the averaged t-values (N = 23) across sensors in the cluster of significant content-specificity (i.e.
4 t-values of PLV within groups of same content vs. groups of mixed content). Topographies in the middle
5 are t-values within the same cluster (across time, frequency and space), averaged across time and
6 across all frequencies (top) or only for 8 Hz (bottom). Both topographies show maximal values over
7 occipital and parietal sensors. The right panel shows the average t-values across time on virtual sensors,
8 within the temporo-spatial cluster of significant differences at 8Hz. Occipital and parietal sensors
9 expressed the maximal t-values. (b) Occipito-parietal region of interest (ROI) that we used for statistical
10 testing of content-specific reactivation. (c) Time course of content specific phase at 8 Hz during
11 encoding, averaged across the ROI. Below, the sliding window approach is illustrated, in which all
12 possible time windows from encoding were compared to each retrieval time window via phase
13 coherence. Subsequently, combinations of same and different content combinations were contrasted.
14 (d) Cluster of significant differences between content-specific reactivation for successfully remembered
15 and forgotten associations ($p_{cluster} = 0.030$).

16

1 Figure 3: Chronometry of memory replay

2 (a) The 6 non-overlapping time windows from encoding illustrated next to a video-episode (left). The
3 average (across 23 subjects) similarity densities to these windows are on the right. The blue bar denotes
4 where replay was significantly slower within scenes (see e). (b) Cross correlations of similarity densities
5 within this window show the adaptive pattern. The matrix shows the combination of windows that are
6 correlated in each cell. The times in ms at which cross correlation is biggest are displayed in the color-
7 coded cells. In this, lags between windows within scenes are bigger than lags between windows across
8 scenes (right, top); with strict forward replay, all scenes would be replayed in order (right, bottom). (c)
9 Illustration of the cumulative similarity (CS) approach used to test replay-dynamics. If evidence for a
10 window statistically precedes evidence for another during retrieval, its cumulated similarity is higher.
11 Fitting a line through those subsequent encoding windows will therefore result in a negative slope. (d)
12 Average slope of lines fit across all windows' CS, for each subject and time point. Negative slope
13 indicates that earlier encoding-windows have higher CS values and signify forward replay. The slope
14 was tested against 0 with a series of two-sided t-tests. (e) Contrast of average slopes from the average
15 fit across windows within scenes and a fit across all windows, supporting an adaptive replay framework.
16 A series of one-sided t-tests was used to contrast the slope across participants at every time point. The
17 horizontal blue bars in d and e indicate significance controlling the false discovery rate at a level of 5%.

18

1 Figure 4: Illustration of several replay speeds and their aggregation

2 Temporal patterns from different time-windows during the video-episodes are reinstated during
3 retrieval. The temporal patterns (colored according to the corresponding time window) signify replay
4 at the same speed (no compression), yet overall the speed of replay is compressed. If replay can start
5 from the boundaries in the video, the moments of replay for the second part of a scene will be
6 substantially delayed relative to the start of the sub-scene (local compression). On the other hand, the
7 replay of the beginning of a new scene does not have to start too long after the beginning of a previous
8 scene, because replay can be initiated from this event boundary, because of skipping between
9 boundaries. Replay patterns from single trials (t_1, t_2, t_3, t_4) will then aggregate such that patterns from
10 the same scene are statistically further apart than patterns from different scenes (bottom row, color
11 coded according to the time-window). The global compression level will be higher than expected from
12 the local compression level within scenes and substantially higher than expected from temporal pattern
13 reinstatement. Only such a dynamic replay framework that allows for skipping between patterns, can
14 explain the observed result of various speeds within the replay interval (Figure 3).

15

16

17

18

19

20

21

22

23

24

25

26

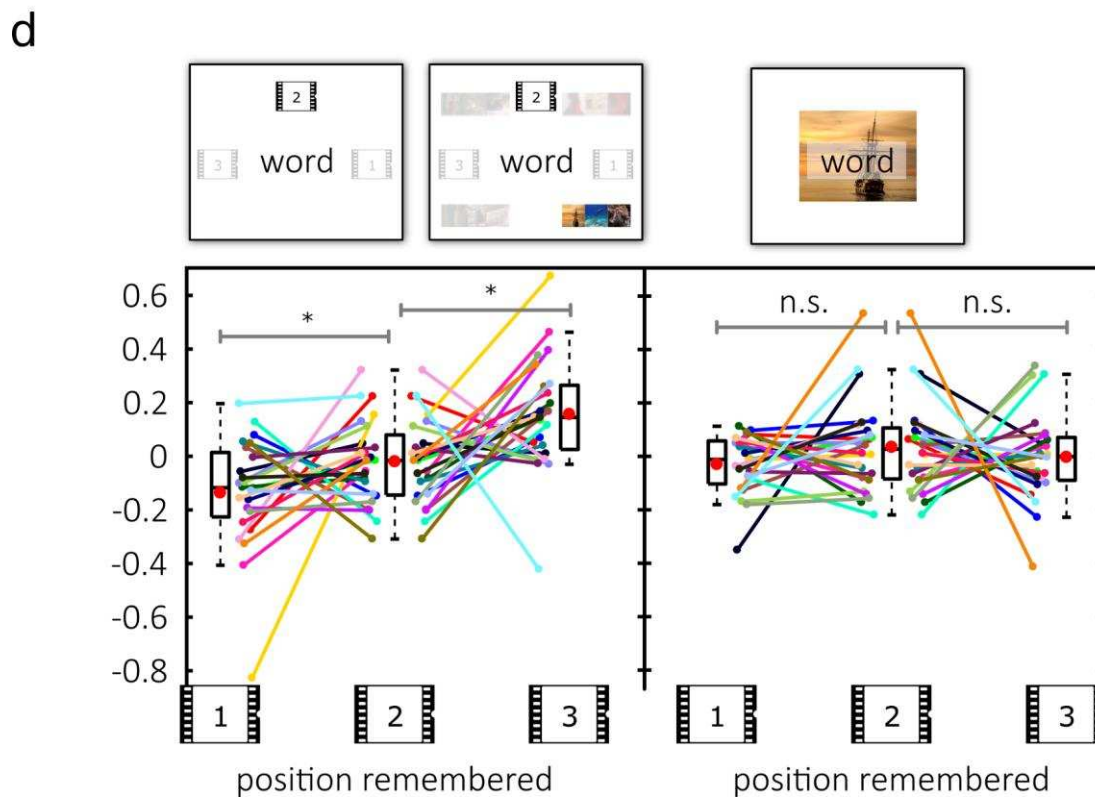
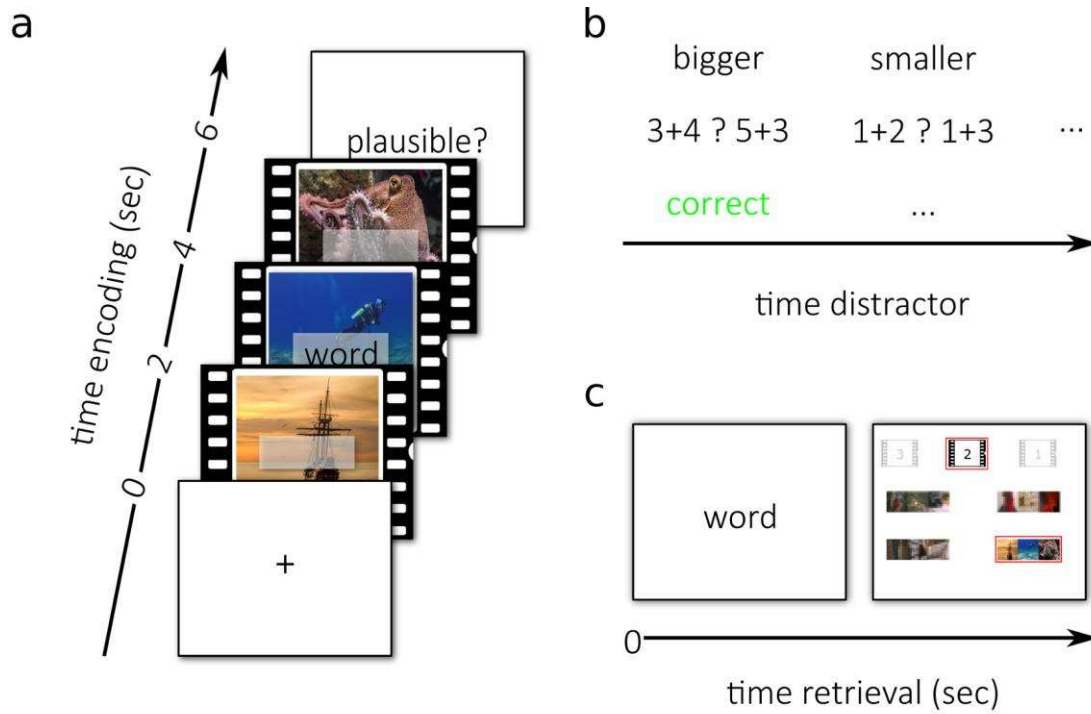
27

28

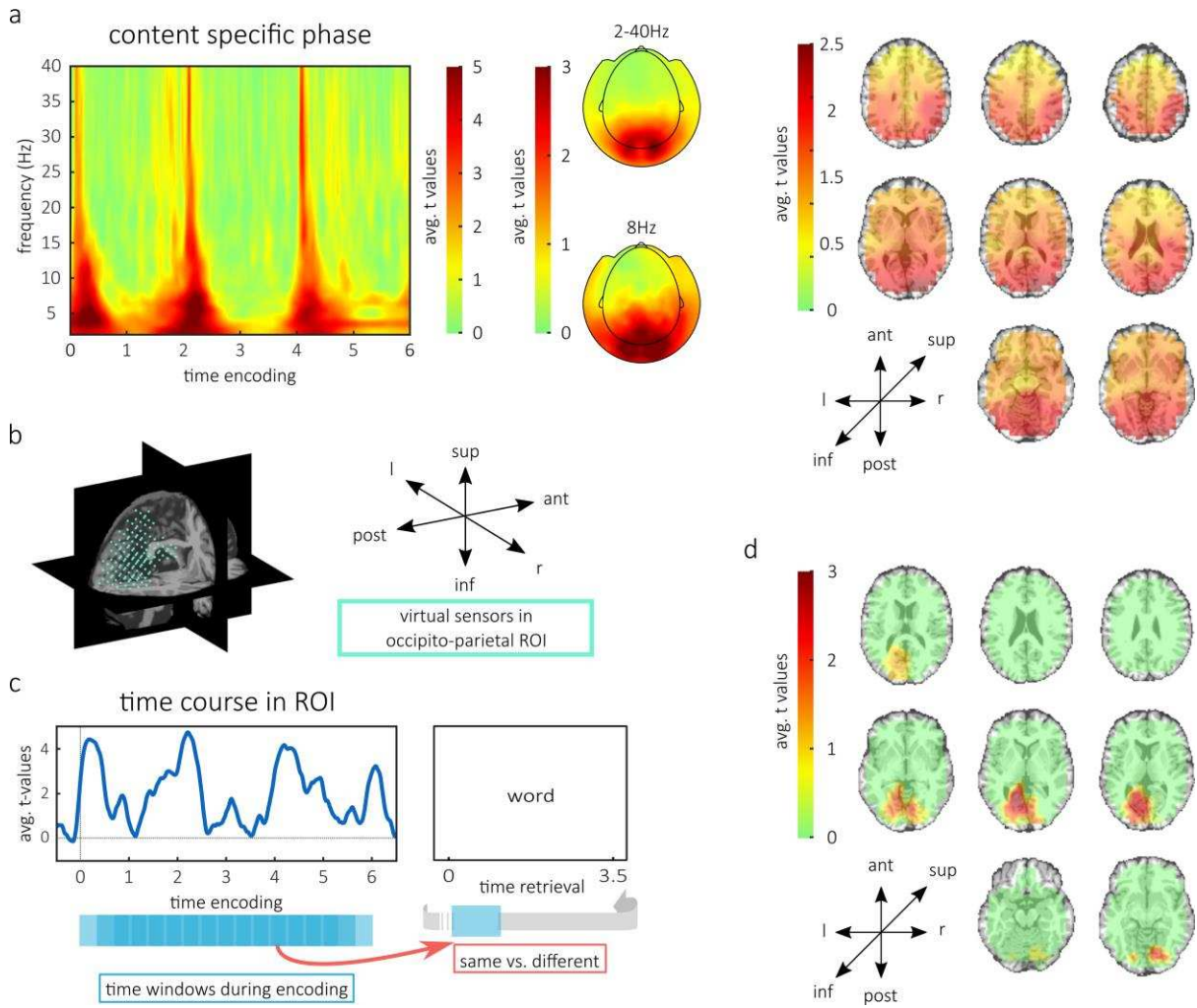
29

30

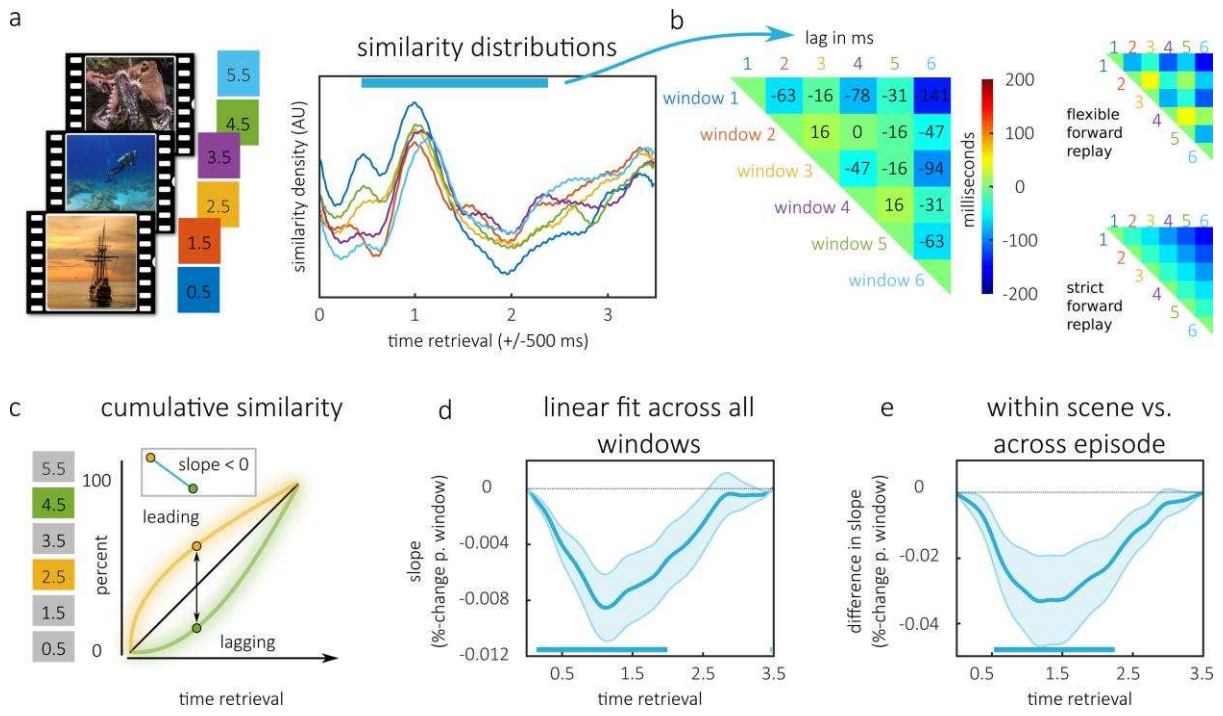
31



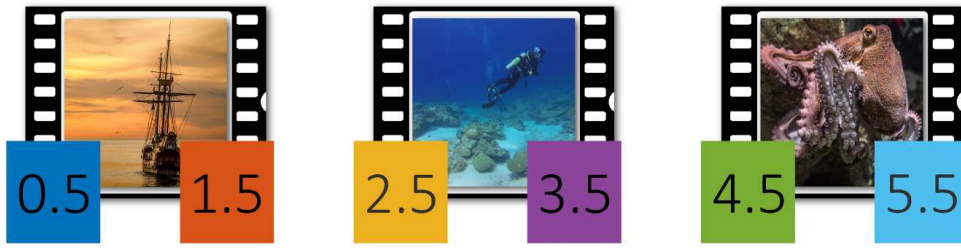
1



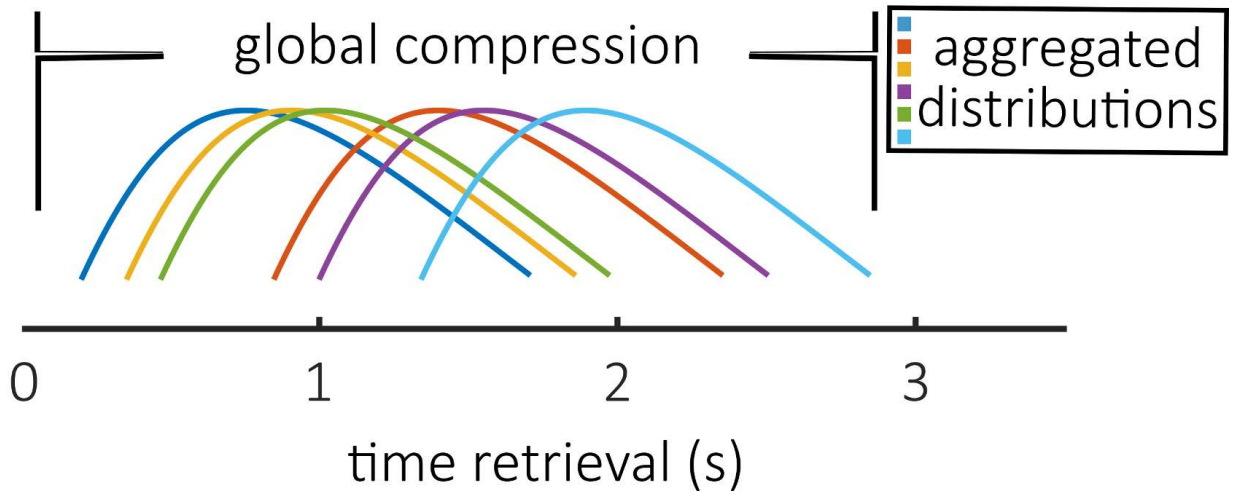
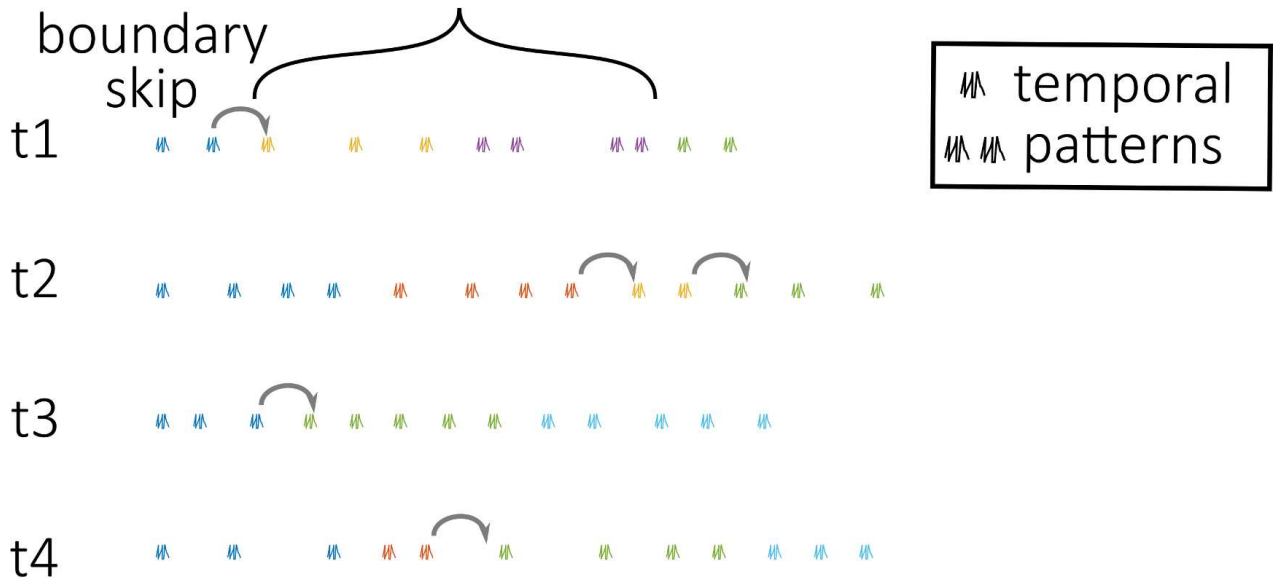
- 1
- 2
- 3



1



local compression



1