bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Relating the past with the present:

Information integration and segregation during ongoing narrative processing

Claire H. C. Chang[1*], Christina Lazaridi[2,3], Yaara Yeshurun[4], Kenneth A. Norman[1,5], Uri Hasson[1,5]

1. Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08540-1010, USA.
2. School of the Arts, Columbia University, New York, NY 10027, USA.
3. Lewis Center for the Arts, Princeton University, Princeton, New Jersey 08540-1010, USA.
4. School of Psychological Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
5. Department of Psychology, Princeton University, Princeton, New Jersey 08540-1010, USA.

* Correspondence and requests for materials should be addressed to C.H.C. (email: hcchang73@gmail.com).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Real-life events unfold continuously over multiple minutes. In this study, in order to examine how the brain continuously integrates information while segregating the accumulated information from irrelevant inputs, a professional writer actively designed a narrative. Two unrelated storylines were interleaved across 30 one-minute segments (ABAB) and merged in a last (C) part, where narrative motifs unique to the A or B storylines recurred to induce reinstatement. Our fMRI results showed neural reinstatement of storylines and motifs in regions with long processing timescale, including the default mode network, suggesting that past information can be segregated from irrelevant inputs in an inactive state for minutes. The reactivated storyline representation was updated by integrating new segments, as reflected by the increasing neural differentiation between storylines. Furthermore, we found a positive correlation between neural reinstatement of motifs and behavioral performance in relating separated events connected by motifs, again demonstrating information integration during continuous processing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Real-life events unfold over multiple timescales. A word achieves its full meaning in the context of a sentence, a sentence in the context of a paragraph, and a paragraph in the context of the story as a whole. Recently, using real-life stimuli such as stories and movies we have revealed a cortical hierarchy of timescales that accumulates information over increasing temporal receptive windows (TRWs) [1–5]. Namely, early sensory areas accumulate information over short timescales of tens of milliseconds, coinciding with the duration of phonemes and words. Adjacent areas along the superior temporal cortex accumulate information over hundreds of milliseconds, coinciding with the duration of single sentences, while high order areas, which overlap with the default mode network (DMN) [6,7], accumulate information across paragraphs as the narrative unfolds over many minutes.

Unlike classical theories of working memory[8,9], which distinguish between areas that process the incoming information and areas that accumulate and protect information in working memory buffers, the processing timescales hierarchy suggests that memory is an integral component of each processing unit (which we refer to as a process-memory unit) [10]. In our model, each cortical area accumulates and sustains past information at its preferred timescale, while it dynamically processes and integrates new incoming information. This process-memory hierarchy framework illustrates a simple recurrent mechanism for accumulating and synthesizing ongoing continuous information as a story unfolds over time.

In this work, we asked how the process-memory hierarchy manages to segregate and protect incoming information from being combined with irrelevant recently accumulated information, while at the same time integrating relevant events over time. To probe this question, we collaborated with a professional writer to craft an original fictional story with a purposefully designed narrative structure. The first part of the narrative consisted of two seemingly unrelated storylines, A, which takes place in Los Angeles, employing a distinct set of characters, and B, which takes place in New York and involves another set of characters (Fig. 1a). The two storylines were presented in an interleaved fashion over 30 segments, 15 segments for each storyline ($A_1B_1A_2B_{2\ldots}A_{15}B_{15}$). The purpose of interleaving two unrelated storylines was to assess how areas with long-timescale integrate information across minute-long segments (e.g., across story A segments or story B segments) while at the same time protecting each storyline from being integrated with (and confused with) the unrelated parallel narrative. In the last 15 segments (Part C), the two storylines merged into a unified narrative and the connections between the characters from New York and Los Angeles were revealed. One of the main techniques for bridging part C with A and B was to embed specific images/situations/phrases, i.e., narrative motifs, within either the A or B storylines. The recurrences of these motifs in part C were designed to reinstate specific moments from storylines A and B, in an attempt to integrate the two plot lines into one coherent narrative.

This design provides some control over the story structure and enables us to generate the following predictions. First, we predicted that cortical areas with long processing timescales would slowly build a unique activation pattern for each of the two storylines (A and B) as they unfold over time. More specifically, we predicted that as the story unfolds, an incoming segment from storyline A would reinstate and update the neural representation of storyline A, while the neural activity pattern associated with storyline B would subside, and vice versa. A previous study using a movie stimulus with two interleaved storylines reported higher pattern similarity

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between segments from the same storylines in the hippocampus and a few other regions [13]; however, the interleaving of storylines in that study occurred at irregular intervals dictated by the movie (some very short, some much longer), making it difficult to study the temporal dynamics of how storyline representations are constructed and reinstated. The more regular nature of the interleaving in our custom-designed narrative allows us to build on the prior study, while also providing a clearer view of how story representations are constructed and reactivated in the face of substantial (minute-long) interruptions.

Second, we predicted that during part C, the presentation of narrative motifs from the A/B segments should trigger reinstatement of past neural patterns from segments A and B associated with these motifs, which should promote the merging of the two storylines. Previous studies have shown that verbal recall of an event in a movie or a short video clip reinstated the brain activation pattern that was present during its encoding [11,12,14,15]. In this study, by carefully placing the narrative motifs throughout the story, we could construct a priori predictions about the reinstatement of past events during continuous processing of an unfolding narrative. For example, in segment A_1, the main character, Clara, makes homemade chili for her husband in LA. In part C, Clara eats and comments on a different character's (Steven's) homemade chili recipe; we predicted that this will reactivate the memory of segment A_1, making the listeners realize that she has known Steven before moving to LA, and augmenting their understanding of a prior relationship between them, thus integrating storylines A and B during part C.

Finally, if past information can be reactivated after a minute-long suspension, this raises the question of how the latent memory is preserved without interfering with ongoing processing of another storyline.. In a previous study, we found that stimulus-driven interactions between hippocampus and long-timescale cortical regions were associated with increased alignment between neural patterns in people who viewed a story split into two segments, one-day apart[3], and who viewed the story continuously. This finding suggests that hippocampus-mediated episodic memory might help to store and reinstate relevant remote past events. Therefore, we also related participants' hippocampal-cortical connectivity to the activation of storyline-specific and motif-specific patterns to see whether hippocampus contributed to participants' ability to reactivate past relevant contexts.

In sum, the intentionally interleaved storylines and the intermittently positioned narrative motifs created connections between separate and specific parts of the story. This unique design allowed us to examine how related events can be accumulated and integrated in the presence of minute-long intervening breaks, while at the same time being isolated and protected from irrelevant inputs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

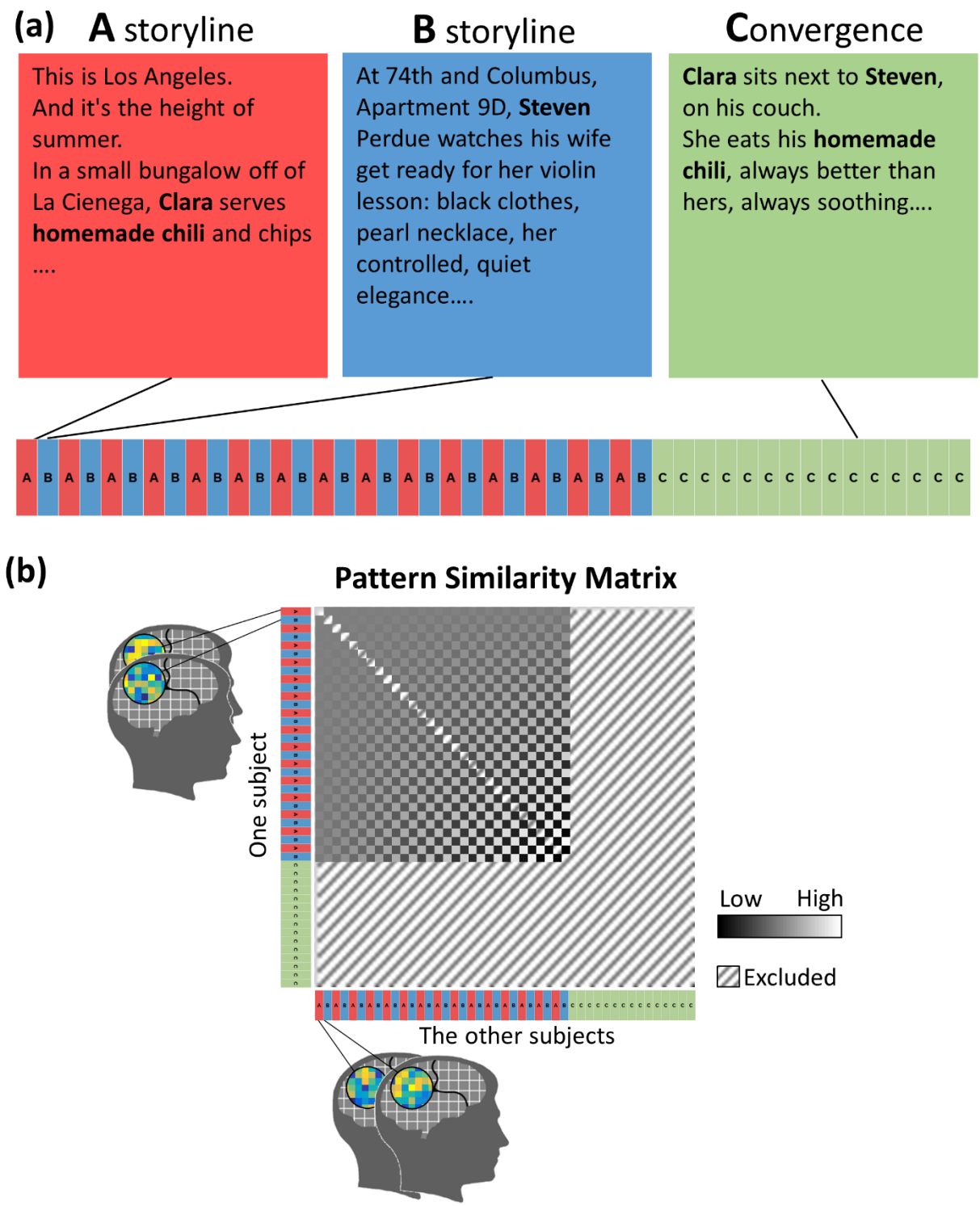

Figure 1. Stimulus and design. (a) The stimulus is a spoken narrative composed of 45 segments with two storylines, A and B. A and B interleave in the first thirty segments and converge in the last fifteen segments (C part). (b) Predicted similarity matrix between brain responses to the 45 segments.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

fMRI data were collected from 25 subjects while they listened to a structured narrative that lasted for approximately 1 hour. The narrative has two interleaved, seemingly unrelated storylines, A & B, that converge in the later C part. In the first set of analyses, we tested how ongoing information from each of the two unrelated storylines was accumulated across minute-long segments while being segregated from the parallel unrelated interleaved storyline. In the second set of analyses, we tested how information in part A & B was reactivated in part C, so as to create a unified narrative. The two storylines are connected to part C using 28 specifically designed, recurrent narrative motifs. These motifs are planted at specific, strategic moments of the narrative by the author (58 occurrences in parts A & B, and 36 occurrences in part C). The participants' overall comprehension and their understanding of the relations created by these motifs were assessed based on post-scan questionnaires. Using representational similarity analysis (RSA) [16,17] on brain activation patterns within ROIs independently defined by a whole-brain parcellation of resting-state fMRI[18], we tested whether the structure of the story induced the reinstatement of storylines and narrative motifs, as it was designed to do.

Overall comprehension of the story

The comprehension scores were evaluated based on 28 multiple choice questions in a post-scan questionnaire. The group overall comprehension score was 91% (with a range of 64%-100% across listeners), indicating that most subjects were engaged with the story and were able to follow the plot well.

Relation score of narrative motifs

Most subjects could explicitly report the relation created by the motifs between part C and parts A/B. 15 open questions were included in the post-scan questionnaire to evaluate the relation score. The participants' answers were rated by the author (Lazaridi) using a 0-2 scale (2= correct relation between C and A/B; 1 = either only C or A/B event was mentioned; 0 = wrong relation or do not remember any related events). Below is a sample question and real answers from participants with different rated scores (see Fig. 1a for the related events in the story and Supplementary Table 1 for all the answers and scores):

The following prompts are words, sentences, or phrases that recurred in the story. Please explain their significance to the story:
Q. Home made chili=?
Ans. (score 2): Steven's tradition that Clara adopts.
Ans. (score 1): Steven makes it.
Ans. (score 0): Clara eats Gary's homemade chili.

18 out of the 25 subjects scored above 1 on average across all motifs. The group mean relation score was 1.1 (range 0.33-1.67 across subjects).

Neural reinstatement of storyline

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

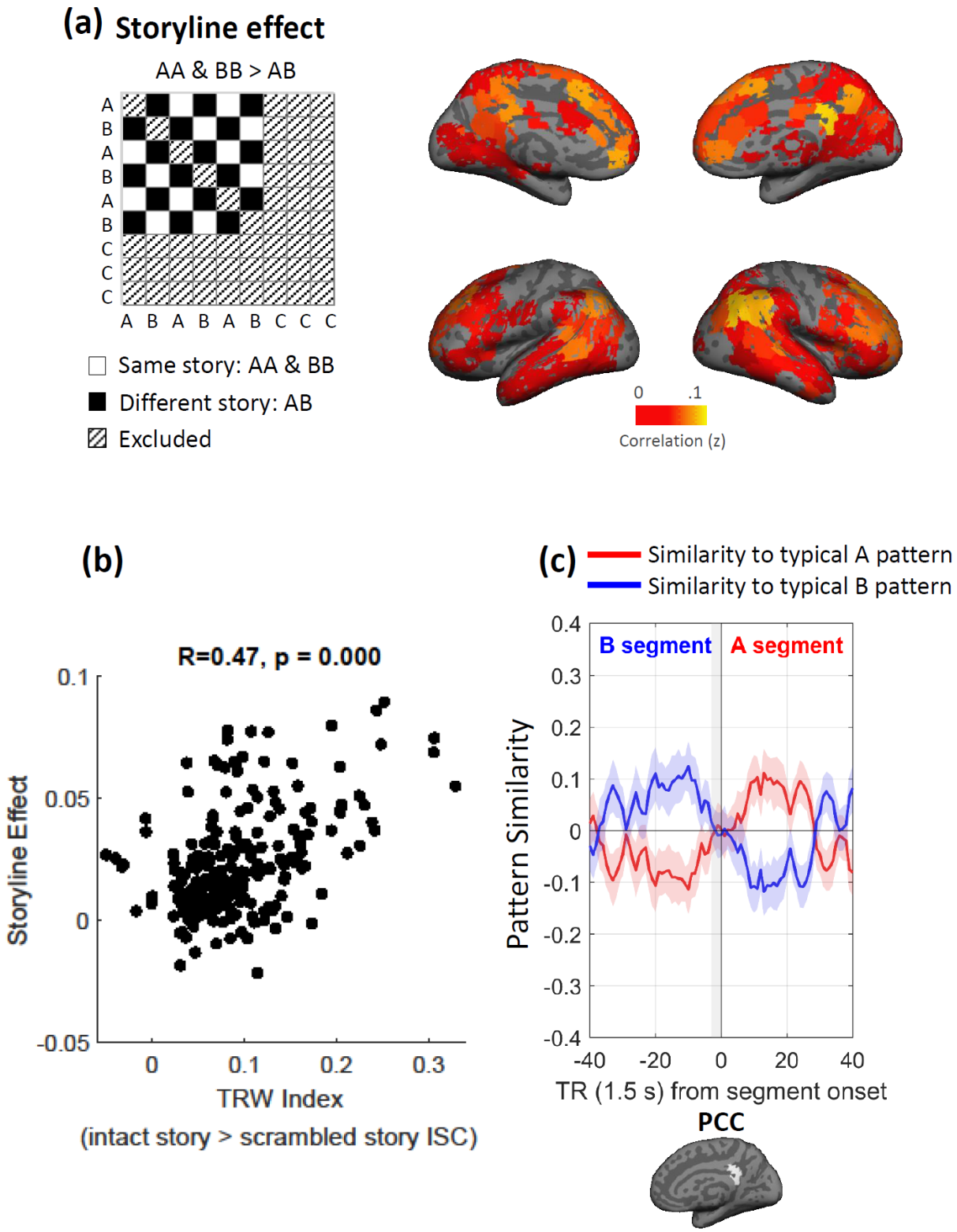

Figure 2. storyline effect. (a) Regions showing larger pattern similarities between segments of the same vs. different storylines (N=25, p < .05, FWE). (b) Correlation between temporal receptive window index and storyline effect across regions. (c) Pattern similarity between typical

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A or B storyline patterns and -40~40 TRs surrounding the boundary between B and A segments in right PCC. Shaded areas indicate 95% confidence interval (CI) across subjects. The vertical gray-shaded area shows the silent pause at boundary.

We first examined whether, and if so, where in the brain the two seemingly unrelated storylines (A&B) had distinc cortical representations. Using RSA we compared the neural patterns within each storyline (AA & BB) to the neural patterns between the two storylines (AB). Within each ROI, we averaged over time within each segment (lasting approximately one minute) to extract a spatial pattern of activity for that segment . We then compared pattern similarity between segments from the same storyline to pattern similarity between segments from different storylines (Fig. 2a). Note that all of these comparisons were done between participants (see methods).

Higher within-storyline pattern similarity was revealed in a large set of regions, including language areas (superior/middle temporal gyrus, inferior frontal gyrus, and supplementary motor cortex), areas in the default mode network (including PCC, precuneus, mPFC, SFG, posterior parietal cortex, angular gyrus, posterior hippocampus, and parahippocampal cortex), areas in the executive network, (including anterior insula, MFG, MCC, and supramarginal gyrus), high order visual areas (including cuneus and fusiform gyrus), and subcortical areas (including putamen, thalamus, and caudate)(please see Supplementary Table 3 for all the brain area abbreviations we used). We also observed the separation between storylines in anatomically defined hippocampus ROIs (Supplementary Fig. 5)[13]. The largest separation across the two storylines was found in areas within the PCC/precuneus.

Neural reinstatement of storyline is stronger in areas with long processing timescales

To define areas with short, medium and long processing timescale, we used an independent dataset [1] to generate a temporal receptive window (TRW) index for each ROI, i.e. the difference in inter-subject correlation between an intact story and its scrambled version. Higher TRW indices were found in prior studies to be associated with increased capacity to accumulate information over long-timescale [1,5]. If the storyline effect only reflected a difference in low-level properties such as wording or acoustic features (note that the same narrator read all segments), regions with low TRW, i.e. regions insensitive to word scrambling, should also show a storyline effect as strong as in high TRW regions. On the contrary, we found a significant positive correlation between TRW index and storyline effect (Fig. 2b). In other words, areas that are capable of accumulating information over long-timescale had a larger difference between storylines.

We selected a long TRW ROI, i.e. PCC, to demonstrate the time course of storyline effect at the segment boundary. We computed the pattern similarity between each of the -40~40 TRs around segment boundaries and the typical A or B storyline patterns. As shown in Fig. 2c, at the boundary between B and A segments, the similarity to typical B pattern rapidly dropped, while the similarity to typical A pattern increased. The two waveforms crossed around the boundary. Similar results were obtained for the complementary transition from A to B segments (Supplementary Fig. 4).

Early vs. late storyline effect

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

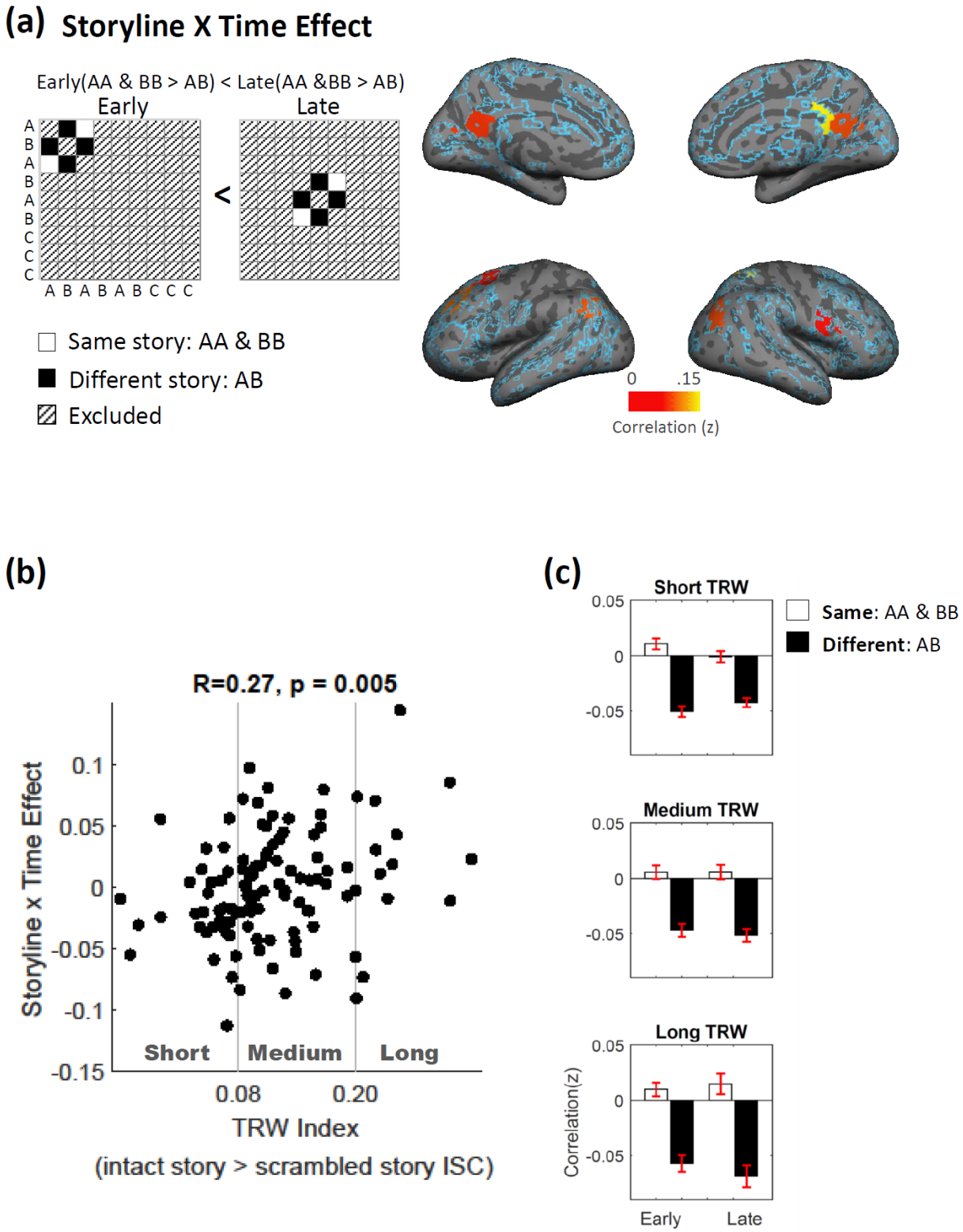

Figure 3. Storyline x Time effect within regions showing significant storyline effect. (a) Regions showing larger storyline effect in the late than early half of the AB part (N=25, p < .05, FWE). The blue outline marks regions showing a significant storyline effect. (b) Correlation

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between TRW index and storyline x time effect across regions. (c) Averaged pattern similarity between segments of the same and different storylines in the early and late halves of the AB part within regions of three TRW bins. Error bars indicate 95% CI across subjects.

We predicted that the segregation of the two storylines (A & B) should increase as the story unfolds and subjects accumulate further information about the unique context of each storyline. To test this hypothesis, we examined whether the storyline effect increased over time by dividing the AB part into the early and later halves (Fig. 3a). Within areas showing the separation between storylines, an increase in the separation of patterns at the later phase (leading to a significant interaction between time and storyline) was found in PCC/precuneus, left AG/IPL, left SFG, right IFG, right MTG/MOG, and right SPL.

The separation of response patterns across the two unrelated storylines increased over time mainly in areas with long processing timescales. As shown in Fig. 3b, we found a significant positive correlation between TRW index and the storyline x time effect. Furthermore, by dividing all the ROIs into three TRW bins (short, medium, and long), Fig. 3c shows that the two storylines became more distinctive in the late AB part and that such effect was larger in regions with longer TRW. A 3-way repeated measure ANOVA revealed an interaction between TRW, time (early/late), and storyline (same/different) ($F(2, 48) = 16.70$, $p < .001$). Post-hoc paired t-tests showed significant storyline effect in all TRW bins, either in early (short TRW: $t(1,24) = 12.70$, $p < .001$, Cohen's $d = 2.54$, CI = 0.05~0.07; medium: $t(1,24) = 9.27$, $p <.001$, Cohen's $d = 0.66$, CI = 0.04~0.06; long: $t(1,24) = 10.46$, $p <.001$, Cohen's $d = 2.09$, CI = 0.05~0.08) or late AB part (short TRW: $t(1,24) = 9.55$, $p <.001$, Cohen's $d = 1.91$, CI = 0.03~0.05; medium: $t(1,24) = 9.91$, $p = <.001$, Cohen's $d = 1.98$, CI = 0.05~0.07; long: $t(1,24) = 9.14$, $p = <.001$, Cohen's $d = 1.83$, CI = 0.06~0.10).

It is worth noticing that the anti-correlations in Fig. 3c do not mean that the two storylines had opposing activation patterns. The two patterns are forced to average to approximately zero by the need to subtract the global mean response before computing the correlation across segments [19,20]. Therefore, the correlation values only reflect the relative, but not the absolute, difference between conditions.

Neural reinstatement of narrative motifs

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

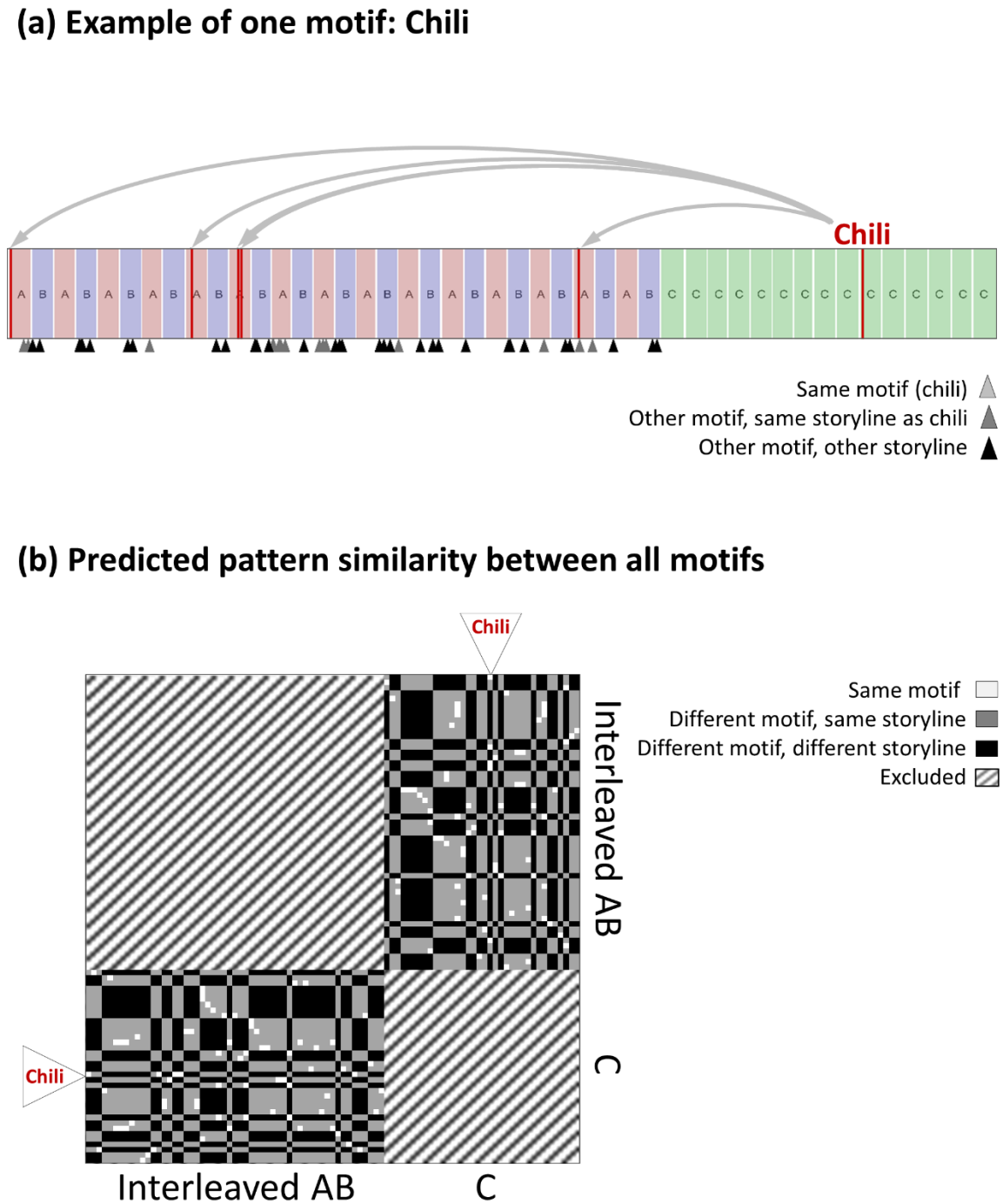

Figure 4. Predicted motif effect. (a) Fragments of the convergence part (C part) were bridged to specific fragments in the AB part by motifs, e.g. chili. (b) predictions about pattern similarity based on motifs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In contrast to the first thirty segments of the story, where the subjects need to segregate information across the two unrelated storylines, in the last C part of the story, subjects are required to integrate information across all segments into a single unified storyline. To test how related but separated events are integrated in areas with long processing timescale, we examined how past information from part AB was reinstated during part C upon the recurrence of the narrative motifs. For each occurrence of motifs in the story, we averaged the 5 TRs after its onset. Then, we correlated each reoccurrence of a narrative motif in part C with all its occurrences in part A or B (Fig. 4). The correlation between matching motifs was compared to the correlation between non-matching motifs from the same storyline (shared storyline) and to the correlation between non-matching motifs from the competing storyline (unrelated segments).

The reappearance of the narrative motifs in part C reinstated specific neural patterns seen when the motifs were encountered during the A/B segments in areas with long integration timescales ($p < .05$, FWE)(Fig. 5, middle). Furthermore, by applying the above analysis TR by TR around the onsets of narrative motifs in part C, we found that the correlation rapidly increased after motif onset and lasted for 5-7 TRs, approximately for 3-6 sentences (Fig. 5, upper and lower panels). The reinstatement effect was specific to matching motifs and was not seen between non-matching motifs, either within or across storylines. Furthermore, this result was mainly found within areas with long processing timescale but not in areas with short processing timescale (see Heschl's gyrus in Fig. 5).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

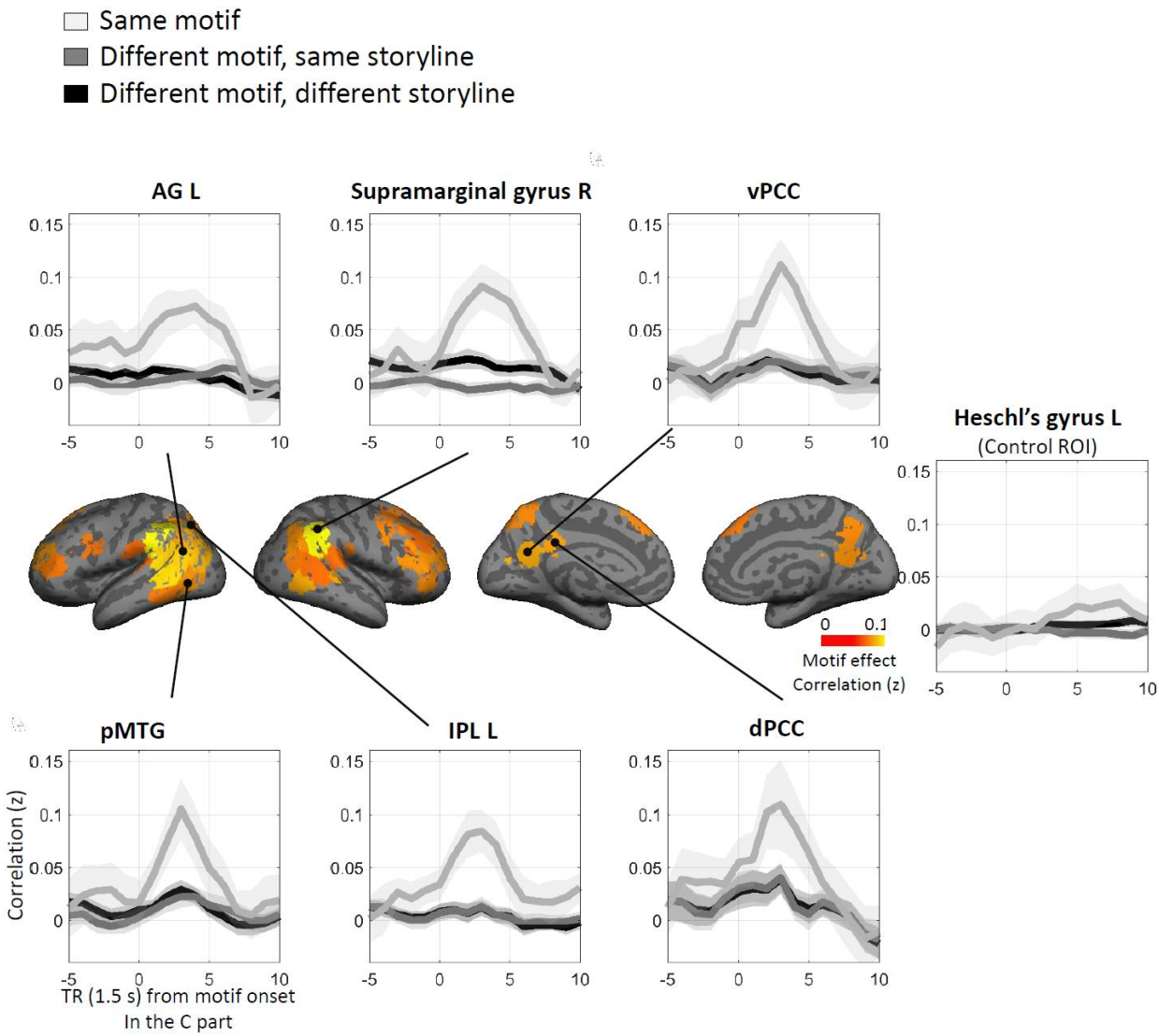

Figure 5. Motif effect. Middle: Regions showing a significant motif effect ($p < .05$, FWE). Upper & Lower: Pattern similarities between motifs in the AB part and TRs around motif onsets in the C part. Shaded areas indicate 95% CI across participants.

Narrative motifs vs. high-frequency word effects

To make sure that the reinstatement of patterns after motif onsets reflects retrieval of narrative information (as opposed to simple reactivation of word representations shared between the A/B and C segments, e.g., the representation of the word "chili"), we performed the same analysis on a set of high-frequency words that occurred in the C part and in either the A or B storyline (e.g., "watch"). We analyzed 28 high-frequency words to match the number of narrative motifs. If the neural reinstatement effect that we observed for motifs simply reflected the reactivation of word representations, the same effect should be observed when we look at the repetition of high-frequency words like "watch" that have no particular narrative significance. In all ROIs with long processing timescale, beside the dorsal PCC, the correlation between matching items was significantly higher for the narrative motifs compared to the high-frequency

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

words, which hovers around zero (p < .05, FWE corrected, Fig. 6). This indicates that word repetition alone was not sufficient to drive neural reinstatement in long-timescale regions; rather, the words had to refer to significant narrative events (as is true for "chili" but not for "watch").

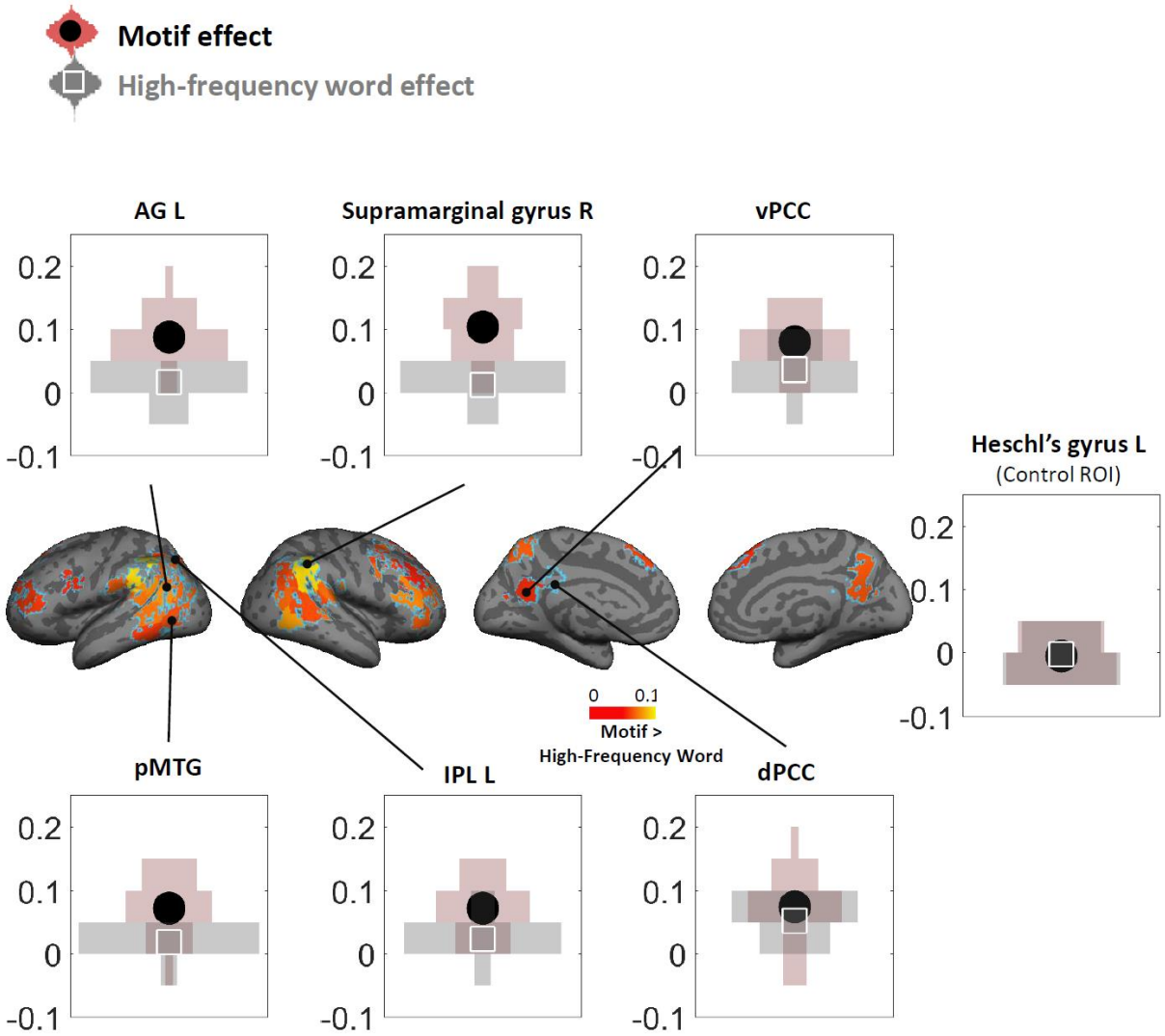

Figure 6. Motif vs. high-frequency word effect. Among regions showing a significant motif effect (marked by the blue outline), the effect of storyline-specific high-frequency words was computed using the same RSA method. The two effects were compared using one-tailed one-sample test (N=25; p < .05, FWE). Shaded areas show the distribution across subjects.

Correlation between motif reinstatement and the behavioral relation score.

We next asked whether the reinstatement of motif-specific activation patterns in segment C was necessary for the understanding of the story plotline as a whole. The neural reinstatement triggered by motifs in part C was correlated with the subjects' ability to relate part C with storyline A and B, as revealed in the relation test (see behavioral results above). In short, each subject was asked to freely recall events related to a given motif from different parts of the story.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Within ROIs showing a significant motif effect, including the dPCC, left IPL, left supramarginal gyrus, left pMTG, dmPFC, right MFG, and left IFG/MFG (one-tailed, FDR corrected q < .05, Fig. 7), we observed a correlation between the neural reinstatement of motifs and the individual relation scores. In other words, participants who showed a stronger neural reinstatement of motifs were also better at explaining the narrative related connections among separate events sharing the same motifs. This result was strongest in areas with long processing timescale and was not seen in areas with short processing timescale (see Heschl's gyrus in Fig. 7).

Narrative motifs, emotional engagement, and memorability

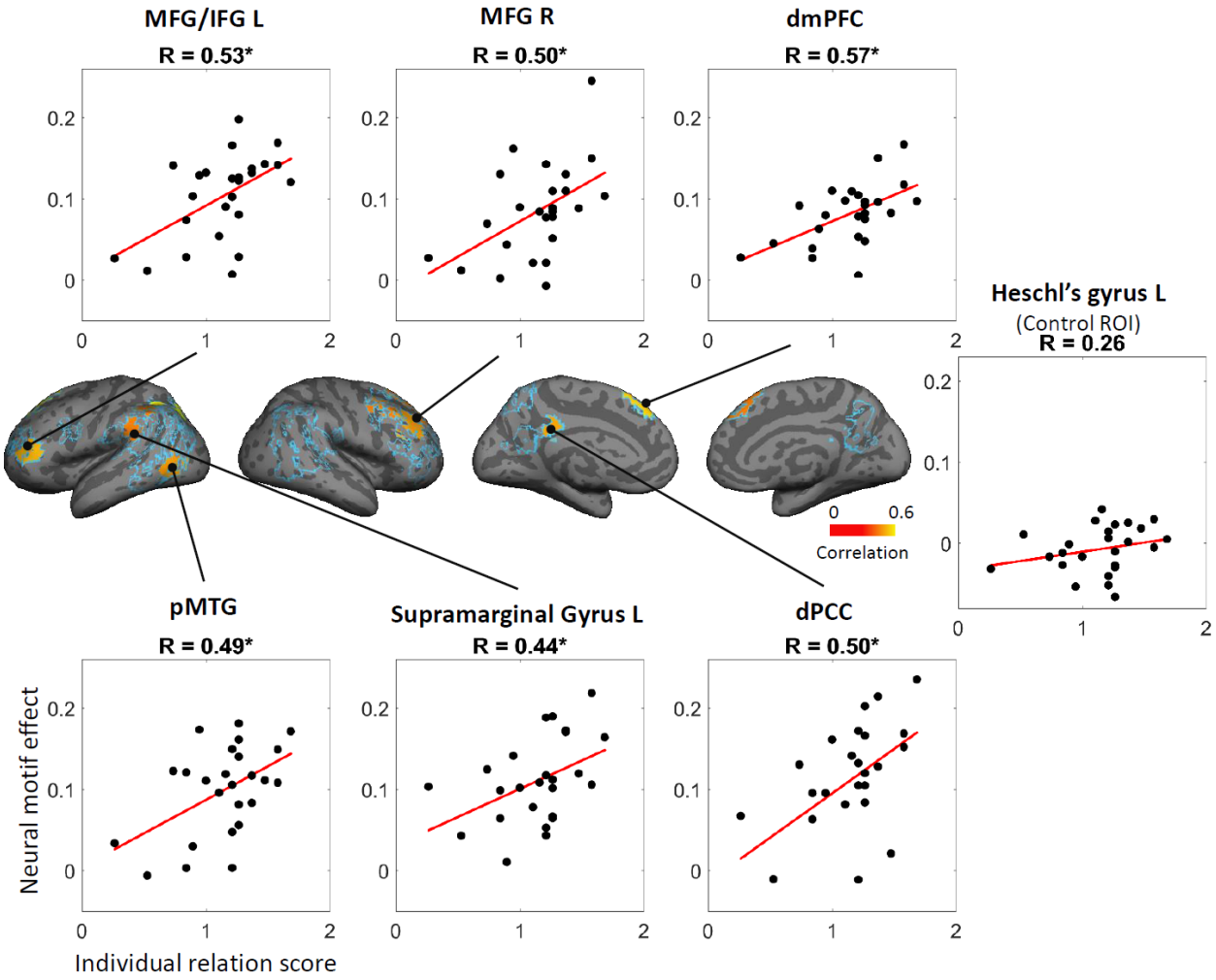

Figure 7. Correlation between the neural reinstatement of motifs and the individual relation scores. Only ROIs showing significant motif effect were tested (marked by the blue outline). Asterisks indicate significant correlation (N = 25, one-tailed FDR corrected q < .05).

Lazaridi, the author of the story, embedded the motifs in highly emotional scenes and provided emotional weightings for the occurrences of motifs in the AB part by a 5 points scale, predicting that heightened emotional context would contribute to the memorability of a motif and the motif relation effect. We computed the Pearson correlation between those emotional

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

weightings provided by the author and the behavioral relation scores of the fifteen narrative motifs, which have corresponding questions in the behavioral questionnaire. A non-significant tendency toward a positive correlation was found (R = .42, p = .06) (Supplementary Fig. 7).

Hippocampal-cortical ISFC and cortical reinstatement of storyline and motif

To examine whether storyline reinstatement is dependent on connectivity with the hippocampus, we examined the correlation between storyline effect and hippocampal-cortical inter-subject functional correlation (ISFC)[21] in ROIs showing a significant storyline effect. For each A/B segment (except for the first segment of each storyline), we correlated the storyline effect with 3hippocampal-cortical ISFC during that segment and ran the correlation across segments (within participants) and across participants (within segments). Because we did not have strong predictions about the relevant time windows for computing the storyline effect and ISFC, we ran an exploratory grid search across a range of analysis parameters. The ISFC between hippocampus and mPFC showed a strong correlation with the storyline effect (FDR corrected q < .05 across ROIs) for multiple settings of analysis parameters (Supplementary Fig. 6), although the result did not survive multiple-comparisons correction when factoring in the full set of analysis parameters, so it should be interpreted with caution. We also examined the correlation between hippocampal ISFC and motif reinstatement in ROIs showing a significant motif effect but did not find a significant correlation.

Discussion

For this study, we actively designed a structured narrative in collaboration with a professional author to test how related events are dynamically and flexibly integrated by the brain while being protected and segregated from intervening irrelevant events. Our results indicate that the memory traces of recent events can be reactivated as a function of current input in areas with long processing timescales within the DMN. This is seen in Fig. 2c, where the neural patterns associated with the current storyline were reactivated at segment boundaries, while the activation patterns of the irrelevant storylines subsided. The reinstatement of relevant past events was also tested in Section C (where the two separate A & B storylines converged into one coherent storyline) by using motifs to reactivate and update particular moments from both storylines (Fig. 5). As predicted, the presentation of specific motifs in part C triggered reinstatement of associated neural patterns from part A and B. Taken together, these results revealed a dynamic shift between currently active context and latent inactive contexts, which helps to integrate information over minute-long interruptions while protecting the accumulated information from irrelevant input.

In support of the idea that the reactivated information is integrated with new input, we found that participants showing stronger neural motif reinstatement performed better in relating separate events sharing the same motif (Fig. 7); this idea is also supported by the finding of increased neural differentiation between storylines over time in regions with a long temporal receptive window (Fig. 3). Using a similar design, in which subjects watched a movie with two interleaved storylines in two parallel worlds (the movie Sliding Doors), a recent study observed a similar increase in neural differentiation between storyline representations within the hippocampus[13]. Our study replicated the neural differentiation between storylines in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

hippocampus (Supplementary Fig. 5) and extended it by revealing similar patterns of results in the default mode network regions. The regular structure of interleaving storylines in our study also allowed us to show the time course of swapping storyline representations in cortex at segment boundaries (Fig. 2c) and explore how hippocampal connectivity related to this storyline effect in the cortex (see below).

After demonstrating the reinstatement of relevant past information, we next addressed how the remote memory was reinstated. Previous work by Chen et al. (2016) showed that hippocampal-cortical interaction helped participants to integrate information across movie segments separated by a 1-day break. In our exploratory analyses, we found that storyline reinstatement increased with functional connectivity to hippocampus in mPFC both across subjects and across segments (Supplementary Fig. 6) for several (but not all) settings of analysis parameters. This finding provides preliminary support for the idea that hippocampally-mediated episodic memory may help to store and reinstate storyline representations. Notably, the degree of inter-subject functional connectivity between hippocampus and cortex did not reliably predict neural reinstatement triggered by motifs in part C. One possible explanation is that the inter-subject functional connectivity method involves averaging over multiple time points; this may make it less useful for detecting brief reinstatement events triggered by motifs.

In addition to episodic memory, there are other mechanisms that might also contribute to reinstating storyline representations. In particular, recent studies of working memory have shown that past information could be held in the cortex during delay period without persistent activity [24–26] and such inactive neural patterns can be reactivated on task demand [27], by probe stimuli [25], and even by transcranial magnetic stimulation [28]; several computational models have been built to account for the latent memory [29] and short-term synaptic plasticity in cortex has been proposed to be the underlying mechanism[30–32]. More work is needed to delineate the roles of hippocampal episodic memory vs. short-term plasticity within cortex in supporting memory for currently-irrelevant storylines.

Our design successfully induced reactivation of neural patterns associated with specific storylines and motifs, which, we believe, reflects the reinstatement of narrative information relating to past events rather than simple reactivation of word representations. First, we found a significant motif effect even when taking storyline-specific high-frequency words as the baseline (Fig. 6). Second, as noted above, we found a correlation between the behavioral relation scores and the neural motif effects (Fig. 7). In other words, the same set of narrative motifs yielded greater neural reinstatement in subjects who demonstrated a better understanding of the narrative. Third, the same motif was not always expressed in the same words ("throwing up" vs. "Clara feels sick, as the coffeecake rises to her throat"). As for the storyline effect, the strongest difference between storylines was found in regions with long TRW, i.e. regions where randomly ordered words failed to elicit reliable responses[1] (Fig. 2b). In addition, 27% of the word tokens in the early AB part are storyline specific, while only 24% of the word tokens in the late AB part are storyline specific. Therefore, the difference in wording could hardly explain the stronger storyline effect in the later AB part (Fig. 3).

In conclusion, real-life events require dynamic integration of past and present information. This study demonstrated that cross-disciplinary collaboration is a fruitful approach to a better understanding of the continuous processing of real-life stimuli. We showed how events that unfold over many minutes are integrated into a coherent plotline, while at the same time they are protected from being mixed with irrelevant inputs. Taken together, these results

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

suggest that process-memory may have two states, a state in which prior events are active and influence ongoing information processing, and an inactive state, in which the latent memory does not interfere with the ongoing neural dynamics [10,24]. These findings not only illuminate the neural mechanisms underlying process-memory but also explain (in neuroscientific terms) why the practiced narrative design techniques applied by Lazaridi are so effective. The results of these kinds of collaborative works might also be applied to guide the future design of narratives, in order to induce specific reactions and event interpretations, affecting audience response.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Participants

Twenty-eight participants were recruited. They were all right-handed native English speakers. All participants provided written consent forms before the experiment. Twenty-five participants were included for further analyses (14 females, age 18-40). Three were excluded, one due to anatomical anomalies, one due to excessive motion artifacts in T1 image, and one slept during the story. The experimental protocol was approved by the Institutional Review Board of Princeton University.

Stimulus

The stimulus was created by Lazaridi ("The 21st Year" -- Excerpt, copyright 2019), who has been in collaboration with our lab for a number of years [33]. She has years of experience in practicing and evolving the technique of organizing the audience's understanding, memory and interpretation of a narrative through screenplay writing and professional screenplay development around the world [34]. Compared to other types of writing, the creation of a screenplay is highly audience-driven due to the large investment (in time, collaboration, and financing) inherent in film-making. Furthermore, watching a film is a more continuous experience than reading a book, requiring the screenwriter to guide and unite the audience's understanding and overall response to the narrative without loss of focus or inner thought digressions.

Lazaridi designed the narrative stimulus as a stand-alone fiction text that incorporated her experience-guided narrative techniques of traditional screenplay writing. The narrative consisted of 45 segments, and two seemingly unrelated storylines, A and B. A and B segments were presented in an interleaved manner for the first 30 segments. In the last 15 segments (Part C), the two storylines merged into a unified narrative. Each segment lasted for 41-57 TRs (mean: 46 TRs = 70 sec). They were separated by silent pauses of 3-4 TRs. The narrative was recorded by a professional actress (June Stein), who is a native English speaker, and directed by Lazaridi to ensure that the actor's interpretation matched the author's intent. The recording is 56 minutes long.

In the A and B segments, the author incorporated unique narrative motifs (Fig. 1 and Fig. 4), i.e., specific images/situations/phrases, that recurred in part C. The recurrence of motifs in part C is designed to trigger the reinstatement of specific moments from part AB, in order to evolve their meanings and to integrate the two storylines. In total, there were 28 different narrative motifs, occurring 58 times in the AB part, and 36 times in part C. The same narrative motif was not always realized with the same words. The main technique Lazaridi used to make motifs memorable was to embed them in emotionally heightened narrative moments. For example, at the beginning of the narrative, (Part A) Clara serves chili during a party in LA and her interactions with the dish (serving, eating, throwing up after eating it) map a series of seminal emotional moments in her personal narrative.

Procedure

The recording of the narrative was presented using MATLAB (MathWorks) and Psychtoolbox [35] through MRI-compatible insert earphones (Sensimetrics, Model S14). MRI-safe passive noise-canceling headphones were placed over the earbuds for noise reduction and safety. To remove the initial signal drift and the common response to stimulus onset, the narrative was

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

preceded by a 14 TR long musical stimulus, which was unrelated to the narrative and excluded from fMRI analysis. Participants filled a questionnaire after the scanning, to evaluate their overall comprehension of the narrative and their ability to relate events in different parts of the story that shared the same motifs.

MRI acquisition

Subjects were scanned in a 3T full-body MRI scanner (Skyra, Siemens) with a 20-channel head coil. For functional scans, images were acquired using a T2*-weighted echo planar imaging (EPI) pulse sequence (repetition time (TR), 1500 ms; echo time (TE), 28 ms; flip angle, 64°), each volume comprising 27 slices of 4 mm thickness with 0 mm gap; slice acquisition order was interleaved. In-plane resolution was 3×3 mm^2 (field of view (FOV), 192×192 mm^2). Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (TR, 2300 ms; TE, 3.08 ms; flip angle 9°; 0.86 x 0.86 x 0.9 mm^3 resolution; FOV, 220 x 220 mm^2). To minimize head movement, subjects' heads were stabilized with foam padding.

MRI analysis

Preprocessing

MRI data were preprocessed using FSL 5.0 (http://fsl.fmrib.ox.ac.uk/), including BET brain extraction, slice time correction, motion correction, high-pass filtering (140 s cutoff), and spatial smoothing (FWHM 6 mm). All data were aligned to standard 3 mm MNI space. Only voxels that were active in all subjects were included for further analysis.

Following preprocessing, the first 19 TRs were cropped to remove the music preceding the narrative (14 TRs), the time gap between scanning and narrative onset (2 TRs), and to correct for the hemodynamic delay (3 TRs). To verify the temporal alignment between the fMRI data and the stimulus, we computed the temporal correlation between the audio envelop of the stimulus (volume) and the subjects' mean brain activation in left Heschl's gyrus following Honey et al. [36]. The left Heschl's gyrus mask was from Harvard-Oxford cortical structural probabilistic atlases (thresholded at 25%). The audio envelope was calculated using a Hilbert transform and down-sampled to the 1.5 s TR. The correlations were computed with -100-100 TRs lag to find the time lag that showed the highest correlation. The averaged peak time was 0.12 TR across subjects, indicating that the narrative and fMRI data were temporally well-aligned.

To account for the low-level properties of the stimulus, a multiple-regression model was built for each voxel. The regressors included an intercept, the audio envelope, and the boxcar function of the between-segment pauses, convolved by the canonical hemodynamic response function and its derivatives with respect to time and dispersion as given in SPM8 (https://www.fil.ion.ucl.ac.uk/spm/). For the effect of audio amplitude and between-segment pause, please see Supplementary Fig. 1. The residuals of the regression model were used for the following analyses.

ROI masks

We used 238 functional regions of interest (ROIs) defined independently by Shen et al.[18] based on whole-brain parcellation of resting-state fMRI data. A control anatomical ROI was also included: left Heschl's gyrus defined using the Harvard-Oxford cortical structural probabilistic atlas, thresholded at 25%. A bilateral anatomical hippocampal mask was obtained using the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

threshold. This mask was divided into anterior (MNI coordinate y>-19), middle (-30 < y < =-19), and posterior parts (y <= -30) ROIs following Collin et al. (2015)[37]. All ROIs had more than 50 active voxels in our data.

Shared response model

When comparing activation patterns across subjects, the mismatch of functional topographies could decrease analysis sensitivity even after anatomical alignment [38,39]. Therefore, we functionally aligned data within each ROI across subjects using the shared response model (SRM) [40]. SRM projects all subjects' data into a common low-dimensional feature space by capturing the components of the response shared across subjects. The input to SRM was a TR x voxel x subject matrix, and the output was a TR x feature x subject matrix. We used fMRI data from the whole story (z-scored over time first) to estimate an SRM with 50 features. Note that no information about storyline or motif was submitted to SRM. Therefore, while this projection inflated the overall inter-subject pattern similarity, it could not artifactually give rise to the storyline or motif effect shown here. The output of SRM was z-scored over time. Unless otherwise stated, all the pattern analyses described below were run based on the resulting 50 features.

We also performed the same analyses without the application of SRM. Generally speaking, a subset of the areas that were significant in the analysis with SRM were also significant in the analysis without SRM. Please see Supplementary Fig. 2 for the results.

RSA of storyline effect

To examine the storyline effect, we performed representational similarity analysis (RSA) [16,17] on brain activation patterns and tested whether the representational similarity between segments from the same storyline was higher than that of segments from different storylines. For each ROI, we first averaged activation patterns within each segment across TRs. The resulting data were z-scored over segments. Pairwise pattern similarities between the 45 activation maps were computed with the leave-one-subject-out method (Fig. 1). Namely, the averaged activation pattern was extracted for each segment. Then the Pearson correlation coefficients between one subject's activation patterns and the averaged patterns of the remaining subjects were computed. The output correlation coefficients (45 x 45 segments) were normalized with Fisher's z-transformation. This procedure was repeated for each of the 25 subjects and each ROI.

We then contrasted the averaged within- and between-storyline similarities in the AB part, excluding the within-segment similarities (the diagonal of the 45 x 45 similarity matrix), to obtain 25 contrast values (Fig. 2a) for each ROI. These contrast values were compared to zero by a one-tailed one-sample t-test and thresholded at $p < .05$ (FWE correction for multiple comparisons).

To examine whether the storyline effect increased over time, for regions showing a significant storyline effect, we computed the storyline effect in the early (segment 1-14) and later (segment 15-30) halves of the AB part separately. 25 contrast values were generated by comparing the late and early storyline effects (late (same > different storyline) > early (same > different storyline)). These contrast values were again submitted to a one-tailed one-sample t-test ($p < .05$, FWE). The results were projected back onto the whole-brain surface and visualized using Freesurfer v6 (http://surfer.nmr.mgh.harvard.edu/).

Temporal receptive window index

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Following Yeshurun et al. [5], the TRW index was generated based on an independent dataset from Lerner et al. [1], which includes an intact story ("Pieman", ~7 min long) and the same story with scrambled word order. Inter-subject correlation between averaged time series of each ROI was computed, using the leave-one-subject-out method, and normalized using Fisher's z transformation. TRW index was then calculated by subtracting the ISC of the scrambled story from that of the intact story.

We examined the correlation between TRW and storyline effect (Fig. 2b) across regions, as well as the correlation between TRW and storyline x time effect within regions showing significant storyline effect (Fig.3b).

In addition, we also divided the TRW index into three bins (bin width: 0.126 r(z)) and grouped all the ROIs into the three bins (Supplementary Fig. 3). For regions with short, medium, and long TRW, averaged pattern similarities between segments of the same and different storylines in the early and late halves of the AB part were computed (Fig. 3c).

Time course of the storyline effect at segment boundary

To further illustrate the time course of the storyline effect, a long TRW ROI, i.e. posterior cingulate cortex, was selected. We computed the pattern similarities between each of the -40-40 TRs around segment onsets and the typical A and B storyline patterns using a leave-one-subject-out method. For example, for the boundary between segment 1 and segment 2, -40~40 TRs around the onset of segment 2 were extracted from one subject. The typical A storyline pattern was obtained by averaging all the A storyline TRs, except for the segments analyzed here, i.e. segment 1 and 2, from the rest of the subjects. The typical B storyline pattern was obtained in the same manner. Pearson correlation between the 81 TRs around segment 2 onset and the typical A and B patterns were calculated and normalized with Fisher's z-transformation. The same procedure was repeated for each subject and each boundary. Fig. 2c shows the transition from B to A segments. Please see Supplementary Fig. 4 for the transition from A to B segments.

RSA of narrative motif effect

For each narrative motif occurrence, the corresponding activation pattern was obtained by averaging 5 TRs immediately after its onset. Pearson correlation coefficients between activation patterns of motifs in the AB part and motifs in the C part were computed with the leave-one-subject-out method and normalized with Fisher's z transformation. As shown in Fig. 4, pattern similarities between narrative motifs were grouped into three types: (1) same motif, (2) different motifs from the same storyline, and (3) different motifs from different storylines (unrelated). For example, pattern similarities between different occurrences of "chili" belong to (1). Similarities between "chili" and other A storyline motifs belong to (2). Similarities between chili and B storyline motifs belong to (3). Motif effect of each "chili" token in part C was defined as the averaged type (1) similarity minus the averaged type (2) similarity, in order to eliminate the confound of storyline effect.

The group motif effect was thresholded with a permutation test. For each ROI, the above procedure was repeated after shuffling the labels of motifs within storylines for 10000 times, creating a null distribution. To correct for multiple comparisons across ROIs, the largest motif effect across ROIs in each of the 10000 iterations was extracted, resulting in a null distribution of the maximum motif effects. Only ROIs with a group motif effect exceeding 95% of the null distribution were considered significant ($p < .05$, FWE)(Fig. 5, middle).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Time course of the narrative motif effect

To further illustrate the time course of the motif effect, for each motif in C, the Pearson correlation coefficients between activation patterns of -5~10 TRs around its onset and the activation patterns of motifs in the AB part were computed. Motifs with a -5~10 TRs time window that overlapped with the between-segment silent pauses were excluded from this analysis. The resulting coefficients were normalized with Fisher's z-transformation and averaged by categories (same motif and same storyline, different motif but same storyline, and unrelated). Fig. 5 shows the resulting pattern similarity around narrative motif onset.

Narrative motif vs. High-frequency word effect

To verify that the motif effect did not result from repeated wordings or word-level semantics, we replaced the narrative motifs with storyline-specific high-frequency words and performed the same RSA. More specifically, among words that only occurred in A and C parts and words that occurred only in B and C parts, we chose the 28 words with the highest lemma/word stem frequencies (Supplementary Table 2). Two out of the twenty-eight narrative motifs were included in this list. Together, these words occurred 111 times in the AB part and 110 times in the C part. Among regions showing a significant motif effect, we calculated the difference between the real motif effect and the effect elicited by high-frequency words for each subject. The 25 difference values were entered into a one-sample one-tailed t-test. The results were thresholded at $p < .05$ (FWE, Fig. 6).

Correlation between hippocampal-cortical ISFC and cortical reinstatement of storyline and motif

To examine whether the cortical reinstatement of storyline was dependent on connectivity with hippocampus, we examined the correlation between hippocampal-cortical inter-subject connectivity (ISFC)[21] and the storyline effect across segments for each subject, within ROIs showing a significant storyline effect.

The ISFC was computed within the 0-40 TR time window after the onset of each segment using the leave-one-subject-out method, i.e. the correlation between one subject's hippocampal activity and the averaged cortical activity of the other subjects. We used the preprocessed data without regressing out the effects of between-segment pause and the audio envelope because it is possible that the activation pulse between segments (Supplementary Fig. 1) does not only reflect the silence but also memory encoding or retrieval[41]. SRM was not applied because topographical alignment is not a concern when comparing the averaged time series between ROIs. The hippocampus seed was defined using Harvard-Oxford cortical structural probabilistic atlas thresholded at 25%.

The storyline effect for each segment was also computed using the leave-one-subject-out method. The typical activation patterns for A and B storylines were first estimated by averaging data from all but one subject, excluding the current segment. Pattern similarity between the resulting typical A & B patterns and the left-out subject's activation pattern for the current segment was then computed. The storyline effect was defined as the difference between pattern similarity to the relevant storyline and the similarity to the irrelevant storyline, taking the previous segment as the baseline, for example, for a B segment: (current segment's similarity to B - similarity to A) - (previous segment's similarity to B - similarity to A).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The correlation between ISFC and the storyline effect across segments was computed for each subject, excluding the first segment of each storyline. The R-values were entered into a one-tailed t-test after Fisher's z-transformation. This initial analysis did not yield a significant result after correction for multiple ROIs ($N = 25$, $p < .05$, FDR correction). In exploratory follow-up analyses, we then systematically examined the influence of the time window of ISFC, the time window of the storyline effect, the hippocampus seed (whole vs. posterior: MNI $y <= -30$), and the baseline of the storyline effect. We also examined the correlation across subjects in each segment. For each combination of analysis parameters, we corrected for multiple-comparisons across ROIs using the FDR method. Please see Supplementary Fig. 6 for the results. No significant correlation was found after simultaneous FDR-correction for multiple-comparisons across ROIs and the twelve sets of analysis parameters.

We examined the correlation between hippocampal ISFC and motif reinstatement in a similar manner. The motif effect was defined and computed using the RSA method described above (based on 5 TRs after motif onsets, using similarity between different motifs of the same storylines as baseline)(Fig. 4). Across motifs in the C part, correlation between the motif effect and ISFC after motif onset was then computed for each subject. We also examined the correlation across subjects for each motif and the influence of ISFC time windows and hippocampus seeds.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data availability. The data used in this study have been publicly released as part of the "Narratives" collection. Raw MRI data are formatted according to the Brain Imaging Data Structure (BIDS) with exhaustive metadata and are publicly available on OpenNeuro: https://openneuro.org/datasets/ds002245. The data corresponding to this study are indicated using the "21styear" task label. These data can be cited using the following reference:

Nastase, S. A., Liu, Y.-F., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N., Chen, J., Honey, C. J., Yeshurun, Y., Regev, M., Nguyen, M., Chang, C. H. C., Baldassano, C. B., Lositsky, O., Simony, E., Chow, M. A., Leong, Y. C., Brooks, P. P., Micciche, E., Choe, G., Goldstein, A., Halchenko, Y. O., Norman, K. A., & Hasson, U. Narratives: fMRI data for evaluating models of naturalistic language comprehension. https://doi.org/10.18112/openneuro.ds002245.v1.0.3

References

1. Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *J. Neurosci.* 31, 2906–2915 (2011).
2. Honey, C. J. *et al.* Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron* 76, 423–434 (2012).
3. Chen, J. *et al.* Accessing Real-Life Episodic Information from Minutes versus Hours Earlier Modulates Hippocampal and High-Order Cortical Dynamics. *Cereb. Cortex* 26, 3428–3441 (2016).
4. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* 95, 709-721.e5 (2017).
5. Yeshurun, Y., Nguyen, M. & Hasson, U. Amplification of local changes along the timescale processing hierarchy. *Proc. Natl. Acad. Sci.* 201701652 (2017). doi:10.1073/pnas.1701652114
6. Raichle, M. E. *et al.* A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–82 (2001).
7. Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. The brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38 (2008).
8. Cowan, N. Chapter 20 What are the differences between long-term, short-term, and working memory? *Prog. Brain Res.* 169, 323–338 (2008).
9. Baddeley, A. Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839 (2003).
10. Hasson, U., Chen, J. & Honey, C. J. Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn. Sci.* 19, 304–313 (2015).
11. Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A. & Hasson, U. How we transmit memories to other brains: constructing shared neural representations via communication.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv 27, 1–25 (2016).

12. Chen, J., Leong, Y. C., Norman, K. A. & Hasson, U. *Shared experience, shared memory: a common structure for brain activity during naturalistic recall*. *bioRxiv* (2016). doi:10.1101/035931

13. Milivojevic, B. *et al.* Coding of Event Nodes and Narrative Context in the Hippocampus. *J. Neurosci.* 36, 12412–12424 (2016).

14. Bird, C. M., Keidel, J. L., Ing, L. P., Horner, A. J. & Burgess, N. Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *J. Neurosci.* 35, 14426–14434 (2015).

15. Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R. & Fried, I. Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322, 96–101 (2008).

16. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103, 3863–8 (2006).

17. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4 (2008).

18. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* 82, 403–415 (2013).

19. Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B. & Bandettini, P. A. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *Neuroimage* 44, 893–905 (2009).

20. Garrido, L., Vaziri-Pashkam, M., Nakayama, K. & Wilmer, J. The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Front. Neurosci.* 7, 174 (2013).

21. Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7, 12141 (2016).

22. Yassa, M. A. & Stark, C. E. L. Pattern separation in the hippocampus. *Trends Neurosci.* 34, 515–525 (2011).

23. Milivojevic, B., Vicente-Grabovetsky, A. & Doeller, C. F. Insight reconfigures hippocampal-prefrontal memories. *Curr. Biol.* 25, 821–830 (2015).

24. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405 (2015).

25. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871 (2017).

26. Sprague, T. C., Ester, E. F. & Serences, J. T. Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron* 91, 694–707 (2016).

27. Watanabe, K. & Funahashi, S. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nat. Neurosci.* 17, 601–611 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

28. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science (80-.).* 354, 1136–1139 (2016).

29. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125 (2009).

30. Miller, E. K., Lundqvist, M. & Bastos, A. M. Working Memory 2.0. *Neuron* 100, 463–475 (2018).

31. Zucker, R. S. & Regehr, W. G. Short-Term Synaptic Plasticity. *Annu. Rev. Physiol.* 64, 355–405 (2002).

32. Mongillo, G., Barak, O. & Tsodyks, M. SynaptiC Theory of Working Memory. *Science (80-.).* 319, 1543–1546 (2008).

33. Yeshurun, Y. *et al.* Same Story, Different Story: The Neural Representation of Interpretive Frameworks. *Psychol. Sci.* 28, 307–319 (2017).

34. Lazaridi, C. *Stories that Change: A Diagnostic Manual for Troubleshooting Your Screenplay*. (Mediterranean Film Institute, 2012).

35. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436 (1997).

36. Honey, C. J., Thompson, C. R., Lerner, Y. & Hasson, U. Not Lost in Translation: Neural Responses Shared Across Languages. *J. Neurosci.* 32, 15277–15283 (2012).

37. Collin, S. H. P., Milivojevic, B. & Doeller, C. F. Memory hierarchies map onto the hippocampal long axis in humans. *Nat. Neurosci.* 18, 1562–1564 (2015).

38. Brett, M., Johnsrude, I. S. & Owen, A. M. The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3, 243–249 (2002).

39. Sabuncu, M. R. *et al.* Function-based Intersubject Alignment of Human Cortical Anatomy. *Cereb. Cortex* 20, 130–140 (2010).

40. Chen, P.-H. *et al.* A Reduced-Dimension fMRI Shared Response Model. *Neural Inf. Process. Syst. Conf.* 460–468 (2015).

41. Ben-Yakov, A. & Dudai, Y. Constructing realistic engrams: poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J. Neurosci.* 31, 9032–42 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.16.908731; this version posted January 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgment

This study is supported by a Magic Grant from Princeton's Humanities Council, Taiwan Ministry of Science and Technology (105-2917-I-564 -007 -), and by the National Institute of Mental Health (R01-MH112357).

Author contributions

C.L., Y.Y., and U.H. designed the experiment; C.L authored the stimulus and C.L and Y.Y created the audio rendition of the narrative. C.H.C. and Y.Y. acquired the data; C.H.C. analyzed and interpreted the data with input from U.H. C.L. and K.A.N.; C.H.C. drafted the manuscript; U.H., K.A.N., Y.Y., and C.L. substantively revised the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.