Neural codes track prior events in a narrative and predict subsequent memory for details

Silvy H.P. Collin[1*], Ross P. Kempner[2], Sunita Srivatsan[2], Kenneth A. Norman[2,3*]

[1]Tilburg School of Humanities and Digital Sciences, Tilburg University Tilburg, Netherlands.
[2]Princeton Neuroscience Institute, Princeton University, Princeton, USA.
[3]Department of Psychology, Princeton University, Princeton, USA.

*Corresponding author(s). E-mail(s): s.h.p.collin@tilburguniversity.edu; knorman@princeton.edu;

Abstract

Throughout our lives, we learn schemas that specify what types of events to expect in certain contexts and the temporal order in which these events typically unfold. Here, our first goal was to investigate how such context-dependent temporal structures are represented in the brain during processing of temporally-extended events. To accomplish this, we ran a 2-day fMRI study in which we exposed participants to many unique animated videos of weddings composed of sequences of rituals; each sequence originated from one of two fictional cultures (North and South), where rituals were shared across cultures, but the transition structure between these rituals differed across cultures. The results, obtained using representational similarity analysis and subsequently confirmed with an unsupervised (K-means) algorithm, revealed that context-dependent temporal structure is represented in multiple ways in parallel, including distinct neural representations for the culture, for particular sequences, and for past and current events within the sequence. Our second goal was to test the hypothesis that neural schema representations scaffold memory for specific details. In keeping with this hypothesis, we found that the strength of the neural representation of the North/South schema for a particular wedding predicted subsequent episodic memory for the details of that wedding.

Keywords: schema, episodic memory, neuroscience

1 Introduction

Over the course of our lives, we acquire knowledge of how events typically unfold, and we leverage this knowledge to make predictions about what will happen next; this knowledge about event structure is often referred to as an event schema or script (Bower et al., 1979; Schank & Abelson, 1975). In the real world, the structure of event sequences is often history-dependent (i.e., non-Markovian), meaning that current sensory observations are not sufficient to determine what will happen next; rather, we have to combine our current sensory input with a representation of what happened in the recent past to make an accurate prediction about the future. For example, consider a situation where you observe a character in a movie standing by the door of their apartment. This observation, on its own, is compatible with a very wide range of next scenes (the character could be going into the apartment, leaving for another destination, etc). However, if you remember that – in the previous scene – the character was sitting in their apartment discussing how they want to get food, this substantially narrows down the space of possible next scenes.

How exactly does the viewer in this example use the past to predict the future? Importantly, there are many distinct ways that a person could leverage the information from the previous scene. One possibility is that the viewer could load up a representation of a *specific sequence ("path") of actions* (e.g., if, in the past, they've repeatedly seen this character walk to their favorite hot dog stand when they are hungry). Another possibility is that the viewer could load up an *abstract schema* – a situation-specific predictive model for how the character will behave, not tied to a particular sequence (e.g., "if the character is hungry, they will be more likely to go into restaurants they encounter instead of walking by them, and they will be irritable when encountering people, because they've acted this way when hungry in the past"; Zacks et al., 2007). A third possibility is that the viewer could simply carry forward a representation of the scene, without loading into memory a specific predicted sequence or abstract schema. This could be useful in several ways (e.g., if you know that the character was just in their apartment and you see them outside the apartment, you can predict that they're in the process of leaving the apartment, rather than coming back to it). A computational challenge inherent in this "carrying forward" strategy is that, if the viewer uses the exact same representation for when a scene is *presently happening* vs. *happened in the past*, this could lead to confusion about what is presently happening; as we will discuss below, one possible way to disambiguate representations of past and present events is to rotate representations of past events in the brain (Libby & Buschman, 2021; van Loon et al., 2018; Wan et al., 2022; Yu et al., 2020).

Previous fMRI studies have identified history-dependent representations in the brain during context-dependent sequence learning tasks (Clarke et al., 2022; Hsieh et al., 2014) but these studies have not sought to discriminate between the three possibilities listed above. The *first goal* of our study is to test for all three types of history-dependent representation during context-dependent sequence learning. The *second goal* of our study is to look at how these different types of history-dependent

representation affect subsequent memory for episodic details. A large body of prior work has demonstrated that schemas at encoding can provide a "scaffold" onto which specific event details can be attached (e.g., Abbott et al., 1985; Alba and Hasher, 1983; Bransford and Johnson, 1972; McClelland et al., 2020; Tompary and Thompson-Schill, 2021), and other work has argued that the inferred schema serves to contextualize episodic memories (DuBrow et al., 2017; Shin & DuBrow, 2020) – e.g., memories formed while your "looking for food" schema was active will be easier to access when that same schema is active later. In keeping with this claim, a recent study found that the strength of neural schema representations during perception of audio and movie narratives was related to later memory for details from these narratives (Masis-Obando et al., 2022). However, what remains to be investigated is whether it is indeed specifically the schema representations that relate to memory for details, rather than the other types of history-dependent memory representations described above. As described below, the design of our study makes it possible to investigate the possible relation between later memory for episodic details and several types of history-dependent memory representations (schema representation, path representation, rotated representation of preceding rituals) during encoding.

To achieve these goals, we ran a 2-day functional magnetic resonance imaging (fMRI) study (Figure 1) in which we exposed participants to many unique animated videos of wedding ceremonies with sequences of rituals originating from a novel (fictional) culture, where each video was accompanied by a unique narrative. The rituals unfolded according to the graph structure shown in Figure 1B. Each wedding was from one of two fictional cultures (indicated by the couple being from the North or the South of a fictional Island). At the start of each wedding video, the culture (North or South) was mentioned (but this information did not persist on screen, so participants had to remember it). After a short introduction to the wedding, one of two "stage 2" rituals (campfire or flower) was shown, with equal probability – there was no way to predict which stage 2 ritual would be shown. Following this unpredictable stage 2 ritual, the subsequent rituals were fully predictable given knowledge of whether the wedding was for a couple from the North or South (e.g., in North weddings, campfire was always followed by coin, and flower was always followed by torch). Importantly, while the weddings from the two cultures incorporated the same rituals (with the exact same frequency), the *transition structures* of the weddings were distinct and opposite from each other (e.g., in South weddings, campfire was followed by torch, not coin; and flower was followed by coin, not torch). This sequence design leads to the non-Markovian property that (starting with the transition from stage 2 to stage 3) participants need to represent what happened earlier in the wedding in order to predict what rituals they will observe going forward. In a previous study that measured behavior only (i.e., no fMRI), we demonstrated using these materials that participants could successfully learn the ritual transition structure of the weddings through experience (i.e., without seeing the actual graph shown in Figure 1B) if they were given blocked training (i.e., where they viewed multiple weddings from the North, followed by multiple weddings from the South), but they performed much worse when given interleaved training (i.e., strictly alternating weddings from

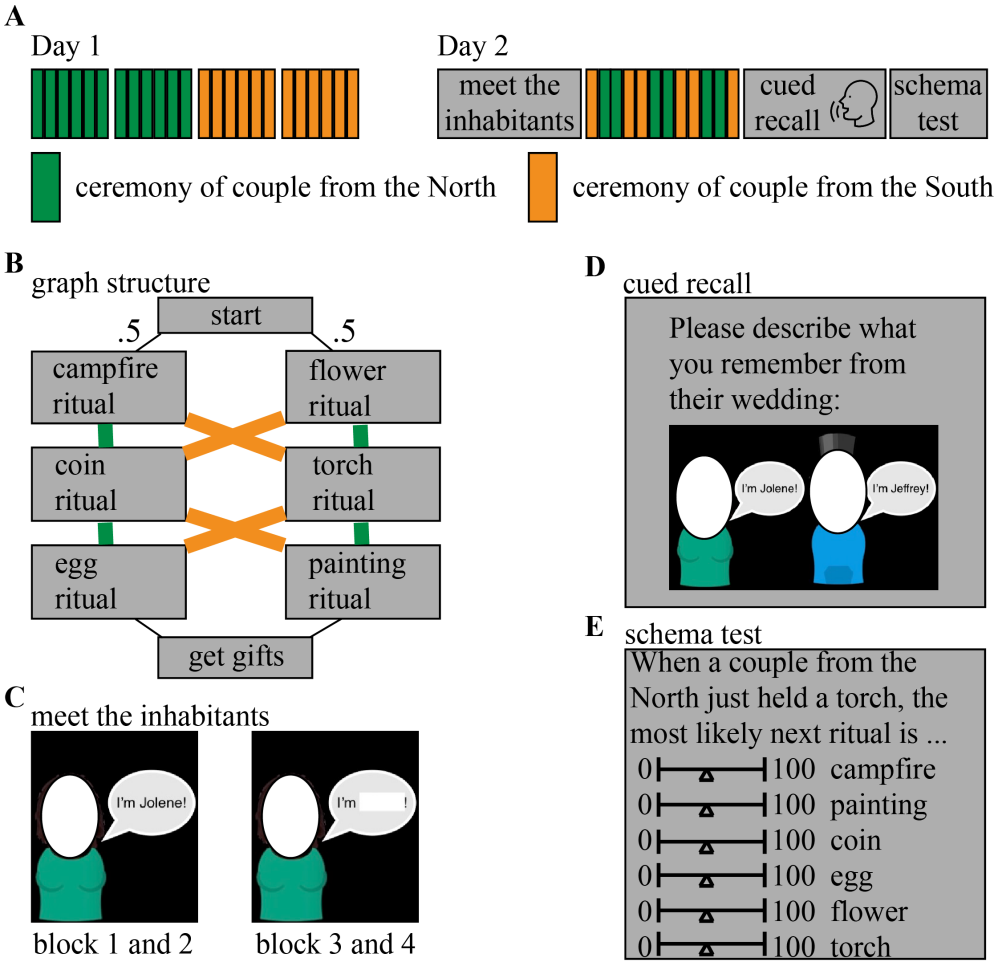

Fig. 1 Task design. (A) Overview of the experiment, consisting of two days. (B) Graph structure that indicates how the ritual sequences unfold in the North graph vs the South graph. The critical sections of this graph structure for analyses are stages 2 (campfire, flower), 3 (coin, torch) and 4 (egg, painting). (C) Screenshots of the pretraining task in which participant were introduced to the characters before viewing the actual wedding videos. Pictures of faces omitted in the figure for privacy reasons. (D) Screenshot of the (self-paced) cued recall task that participants performed on day 2 of the experiment. Pictures of faces omitted in the figure for privacy reasons. (E) Screenshot of the schema test.

the North and South; Beukers et al., 2023). Here, we leveraged this blocked training procedure on day 1 of our study to facilitate successful learning of the ritual transition structure. Next, participants came back on day 2 to view ceremonies that also followed this structure. An important feature of the present study is that each wedding video contained details unique to that wedding; on day 2, we tested participants on their memory for these unique details after they viewed the weddings – this arrangement allowed us to assess the relationship between history-dependent neural

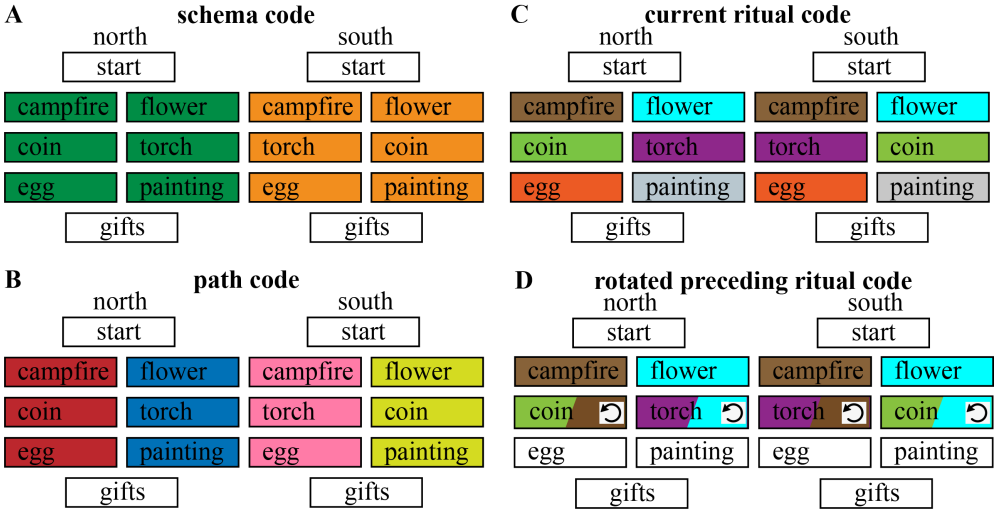

Fig. 2 The four different hypothesized codes. (A) Schema code, representing North schema vs South schema. (B) Path code, representing each of the 4 sequences separately. (C) Current ritual code, representing each of the six rituals separately. (D) Rotated preceding ritual code, representing a rotated version of the preceding ritual along with the (non-rotated) current ritual. The rotation symbol indicates when we predict a rotated version of the preceding ritual to be present.

representations in the brain and subsequent memory for wedding-specific details.

An important feature of our task design is that it gave us the ability to tease apart the three kinds of history-dependent neural representations described above:

- Abstract schema code (North / South): If participants express one neural pattern throughout North weddings that is distinct from the neural pattern expressed throughout South weddings, this is "abstract" in the sense that it can not be explained in terms of representing specific rituals (on average, each ritual occurs equally often in North and South weddings) or in terms of representing a single sequence (since North and South each encompass multiple sequences of rituals; e.g., North weddings can follow the campfire, coin, egg sequence or the flower, torch, painting sequence). As noted above, this representation (if present) would be sufficient to fully predict all of the rituals that are predictable (i.e., in stages 3 and 4).

- Path (sequence) code: In addition to (or instead of) representing the North / South schema, participants could represent the specific "path" (sequence of rituals) that a particular wedding follows. Each of the North and South schemas encompass two distinct paths, making for four possible paths in total. Knowing the schema plus the identity of the current ritual makes it possible to deduce the full path (e.g., North + campfire uniquely specifies that the full path will be campfire, torch, egg). Path codes can be viewed as instances of a schema or script (in the sense that they capture a regularity that is present across multiple observed weddings), but they are

less abstract than a representation of North / South, which (as noted above) would encompass multiple paths. Path codes are also somewhat less useful for prediction than North / South codes (insofar as they can not be used to predict the stage 2 to 3 transition without also knowing whether the couple was from the North or South).

- Previous ritual code: In addition to the above codes, participants could also represent the immediately preceding ritual. On its own, this would not be useful for predicting the stage 2 to 3 transition (i.e., simply knowing that the preceding ritual was campfire, without knowledge of whether the couple was from the North or South, does not specify whether the next ritual will be torch or coin). However, accurately representing the current stage 3 ritual and the previous (stage 2) ritual would be sufficient to predict the stage 4 ritual (i.e., knowing that torch was preceded by campfire is sufficient to deduce that the next ritual will be egg). As noted above, representing the preceding ritual and the current ritual at the same time could cause confusion if both are represented in the same form. A recent study looking at context-dependent sequence representation in mice (Libby & Buschman, 2021) suggests that one can avoid confusion between current and preceding items by rotating the neural code for the preceding item (note that, here, we are using "rotation" in the linear algebraic sense). fMRI and EEG studies of working memory tasks in humans have obtained evidence consistent with this view, suggesting that rotations are used to differentiate between currently-relevant and currently-irrelevant items held in working memory (van Loon et al., 2018; Wan et al., 2022; Yu et al., 2020); these human studies specifically found that representations of currently-irrelevant items were represented in a format that is anticorrelated with the representations of currently-relevant items, analogous to a "photo negative" (i.e., such that neural features that are especially active when an item is currently-relevant are especially inactive when that same item is currently-irrelevant). Based on these findings, we expected that representations of previous rituals, if present, would be anticorrelated with the patterns that are evoked by these same rituals when they are currently happening.

In summary, there were (at least) three distinct ways that participants in our study could represent the context-dependent transition structure of the wedding videos: an abstract North / South schema code (henceforth, *schema code*); a representation of specific sequences (henceforth, *path code*); and a representation of the previous ritual, which we hypothesized would be rotated (henceforth, *rotated previous ritual code*). We also expected that some regions would simply represent the current ritual (henceforth, *current ritual code*).

As we will describe below, we devised methods for quantifying all four of these neural codes on a per-wedding basis using representational similarity analysis, and we correlated these per-wedding neural measures with behavioral measures of subsequent memory for unique details from these weddings. The results of our study show evidence for all four types of codes, but only one of them (the schema code) predicted subsequent memory.

2 Results

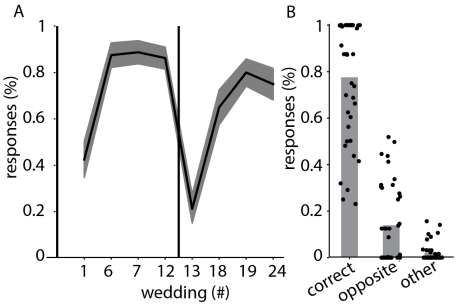

Fig. 3 Participants learned to distinguish North from South schema. (A) Amount of correct answers on 2 alternative forced choice questions throughout the learning task on Day 1. The vertical line between wedding 12 and 13 represents the switch from one to the next schema. (B) percentage assigned to the correct answer option ("correct"), the opposite answer option ("opposite") and the 4 other answer options ("other") during the test for schema learning at the end of Day 2. The percentage for "other" is divided by 4 since it contains 4 out of 6 answer possibilities.

Our experiment relies on participants' ability to learn the schemas that we presented. We used blocked schema training, as Beukers et al. (2023) showed that this is effective in the paradigm that we used. To verify that participants learned the schemas, we tested whether participants learned to distinguish the ritual order in the North schema as opposed to the South schema. To assess their learning, participants received stop-and-ask two-alternative forced choice questions asking them to predict what will happen next throughout the task. Additionally, they were tested on their knowledge of the schemas at the end of day 2 using questions like, "If a couple from the North just planted a flower, what is the most likely ritual to happen next?". They were given all possible rituals as answer options and were told to allocate a total of 100 percent across these answer options.

Results from these tasks indicated that participants indeed learned the two schemas (Figure 3), assigning a significantly higher percentage to the correct answer option as opposed to the answer option of the opposite schema ($W = 807$, $p < .001$).

2.1 Various neural codes to represent context-dependent sequences

We used a representational similarity analysis (RSA; Kriegeskorte et al., 2008) searchlight approach to discover which brain regions would represent a certain hypothesized code. We calculated an average neural pattern for each of the combinations of ritual + schema (e.g., "campfire + North") across all weddings but one (referred to as "template patterns"), and then compared these templates to the neural patterns evoked by rituals in the left-out-weddings. Then we looped through all weddings (see Methods for details).

2.1.1 Schema code across the brain

For the schema code, we expected that different neural patterns would be active for rituals that belong to North or South weddings, regardless of which specific ritual and

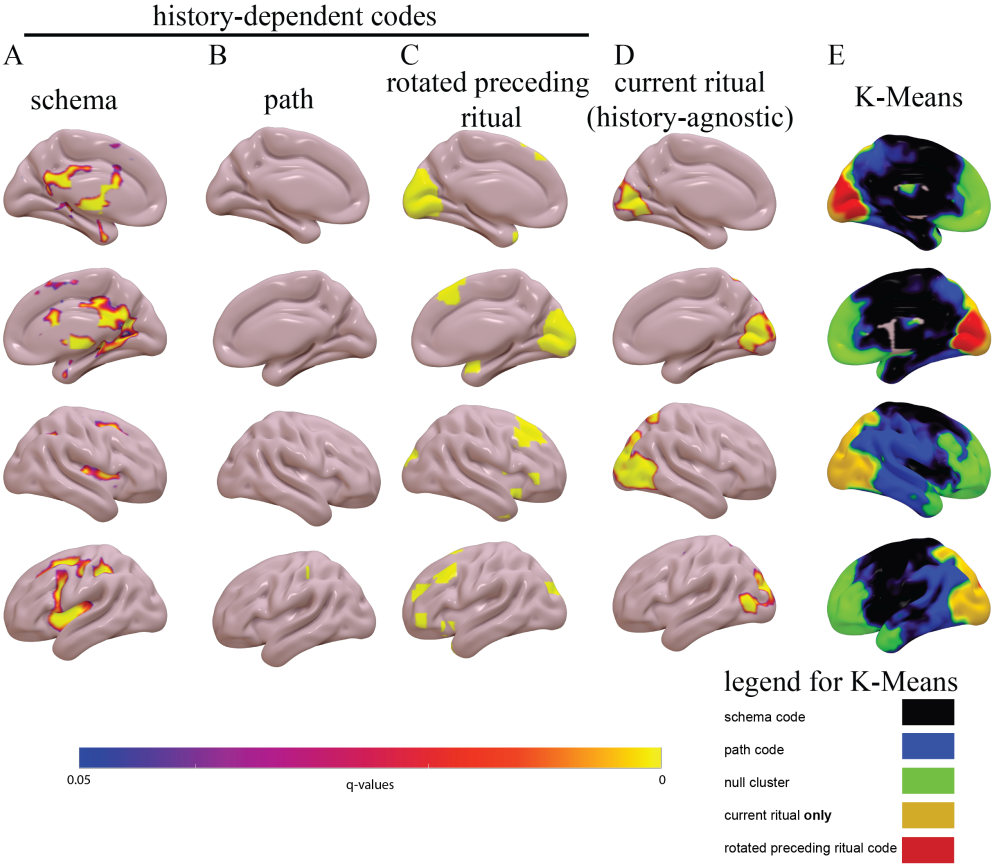

Fig. 4 Results of the searchlight RSA showing each of the 4 different neural codes (FDR corrected), and of the K-means clustering analysis. (A) FDR corrected brain map for schema code. (B) FDR corrected brain map for path code. (C) FDR corrected brain map for rotated preceding ritual code. (D) FDR corrected brain map for current ritual code. The current ritual code and rotated preceding ritual code regions (C and D) are defined in a way that is not mutually exclusive (see Methods), which is why the regions are partly overlapping. (E) K-means clustering results (see section on "Exploratory analysis: unsupervised clustering of the whole brain").

path was being shown at that time. This was calculated by contrasting the similarity of neural patterns corresponding to the same schema and neural patterns corresponding to different schemas (see Figure 10 and Methods). The whole-brain RSA searchlight approach showed a significant schema code (FDR corrected) in several regions, most prominently in the thalamus, pallidum, caudate, posterior medial cortex, hippocampus, parahippocampus, fusiform gyrus, inferior and superior temporal regions, temporal pole, superior frontal regions (Figure 4A). Figure 5 and Appendix Figure A1 show a posthoc confirmation of the pattern.

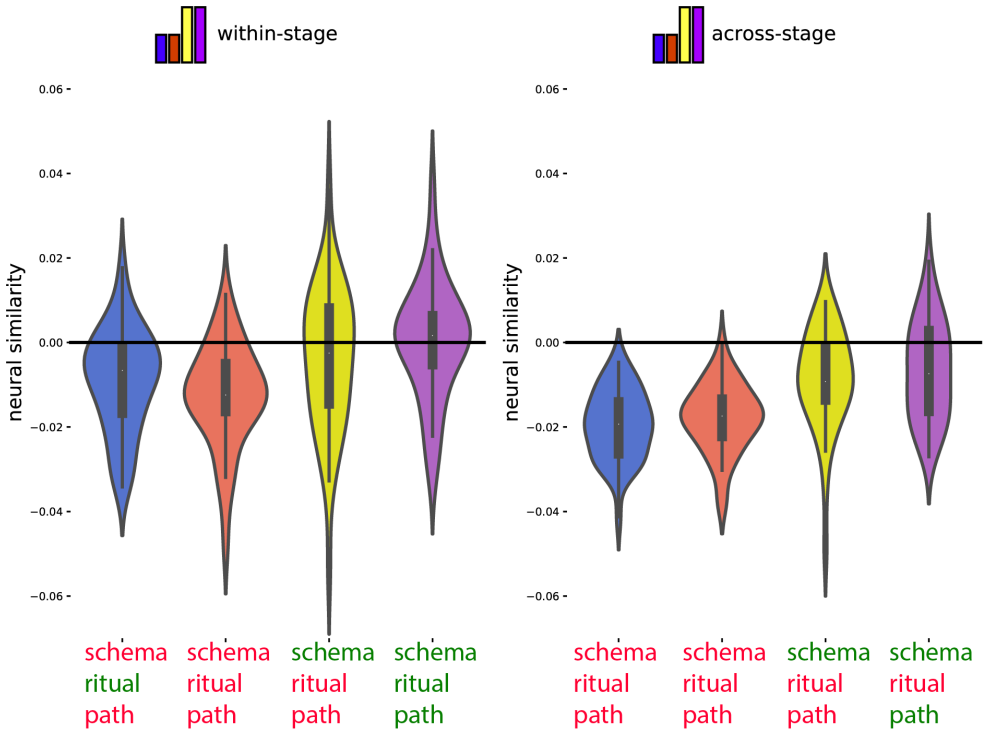

Fig. 5 Posthoc confirmation of the pattern for the schema neural code. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. The small bars at the top of each little plot indicate the predicted pattern. The schema code pattern was present for both within-stage and across-stage comparisons, as evidenced by higher similarity for same-schema conditions (i.e., yellow and purple) vs. different-schema conditions (i.e., blue and orange); within-stage: $t(39) = 4.214$, $p < 0.001$; across-stage: $t(39) = 6.310$, $p < 0.001$. While these regions showed higher similarity for same vs. different schemas, they did not show sensitivity to same vs. different path, as evidenced by non-significant within-stage and across-stage comparisons for same-path, same-schema (purple) vs different-path, same-schema (yellow); within-stage: $t(39) = 1.230$, $p = 0.226$, across-stage: $t(39) = 0.690$, $p = 0.494$.

2.1.2 Path code in postcentral gyrus

For the path code, we expected that different neural patterns would be active for each of the four different sequences ("paths"). This was calculated by contrasting the similarity of neural patterns belonging to the same path and neural patterns belonging to the other path from the same schema (see Figure 10 and Methods). The whole-brain RSA searchlight approach showed a significant path code (FDR corrected) in postcentral gyrus (Figure 4B). Figure 6 and Appendix Figure A2 show a posthoc confirmation of the pattern.

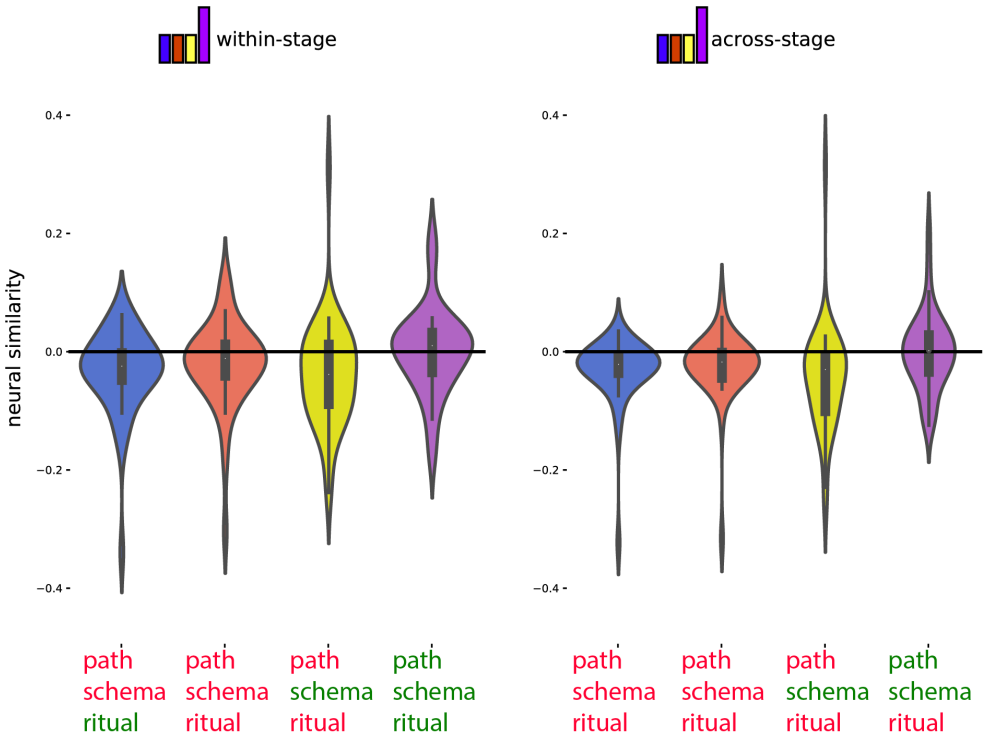

Fig. 6 Posthoc confirmation of the pattern for the path neural code. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. The small bars at the top of each little plot indicate the predicted pattern. These regions show a path code pattern for both within-stage and across-stage comparisons, as evidenced by higher similarity for the same-path, same-schema condition (i.e., purple) compared to the different-path, same-schema condition (i.e., yellow); within-stage: $t(39) = 2.514$, $p = 0.016$; across-stage: $t(39) = 3.109$, $p = 0.003$. The presence of a significant across-stage path code effect is especially important – the within-stage path code effect is confounded with same vs. different ritual so it does not provide strong evidence for path coding *per se*. All of the across-stage comparisons looking at same vs. different path involve different rituals, so they are not subject to this confound.

2.1.3 Rotated preceding ritual code in medial occipital regions

For this code, we expected that the preceding ritual would be represented as a rotated (photo-negative) version of how that ritual is represented when it is currently unfolding. To identify regions with this property, we compared templates acquired at the campfire/flower stage to patterns acquired at the coin/torch stage (and vice-versa). Specifically, we looked for regions where patterns during the coin/torch stage were *especially dissimilar* to the template corresponding to the preceding ritual (e.g., if coin were preceded by campfire, we would expect it to be *less similar* to the campfire

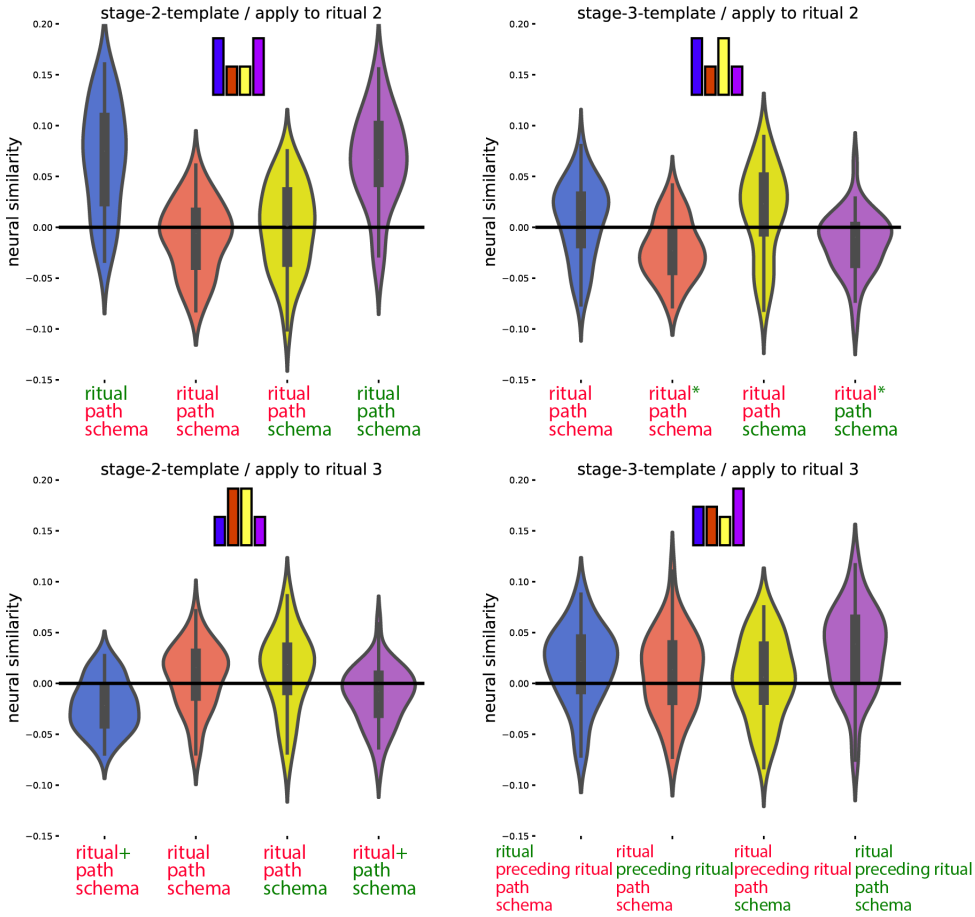

Fig. 7 Posthoc confirmation of the pattern for the rotated preceding ritual neural code. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/preceding-ritual/path are the same (green) or different (red) for the template and held-out ritual. * in green indicates that the held-out pattern corresponds to the ritual that *immediately preceded* the template ritual. + in green indicates that the held-out pattern corresponds to the ritual that *immediately followed* the template ritual. The 4 sub-figures correspond to different combinations of using stage 2/3 as the template and applying the template to stage 2/3. The small bars in the top left corner of each little plot indicate the predicted pattern. This indeed shows a rotated code pattern, as evidenced by the following t-tests across the four subplots: First, in the upper-right plot, there was lower similarity for orange and purple (held-out ritual immediately preceded the template ritual) than blue and yellow, $t(39) = 4.670$; $p < .001$. In the lower-left plot, there was lower similarity for blue and purple (held-out ritual immediately followed the template ritual) than orange and yellow, $t(39) = 4.628$; $p < .001$. The upper-left plot provides evidence that these regions also coded for the current ritual: Similarity was higher for same-ritual (i.e., blue and purple) than across-ritual (i.e., orange and yellow) comparisons, $t(39) = 8.384$, $p < .001$. The lower-right plot provides evidence that these regions coded for both the current ritual and the preceding ritual: Similarity was higher when both the current and preceding ritual were the same (purple) than when only the current ritual was the same (blue), $t(39) = 2.071$; $p = 0.045$, or only the preceding ritual was the same (orange), $t(39) = 2.763$; $p = 0.009$.

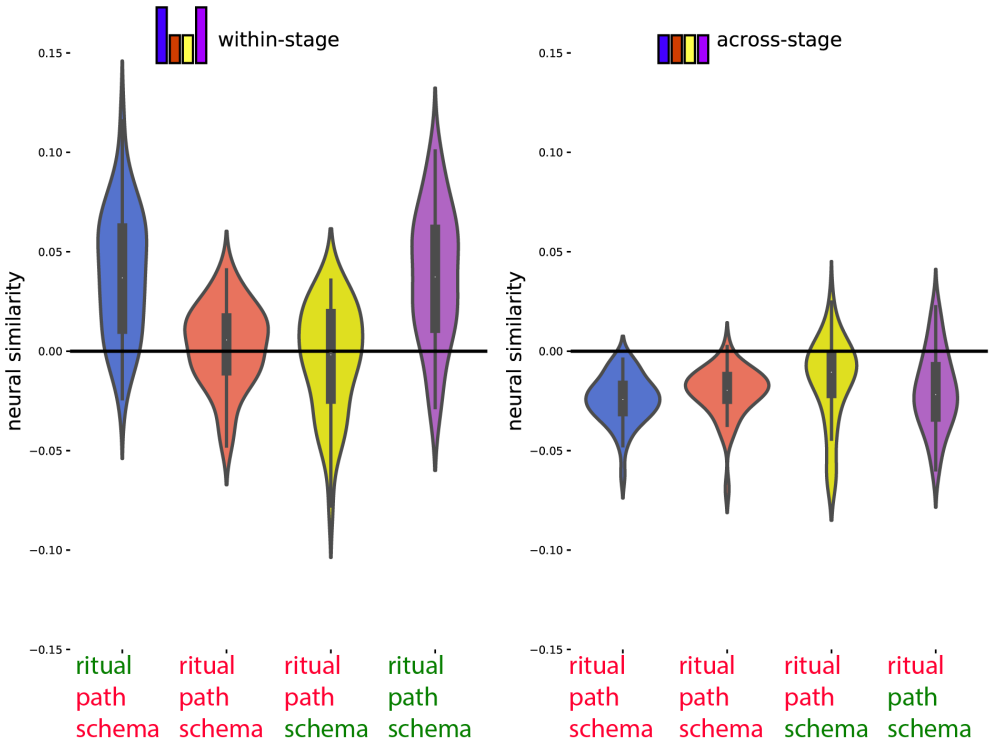

Fig. 8 Posthoc confirmation of the pattern for the current ritual neural code. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. The small bars at the top of each little plot indicate the predicted pattern. The current ritual code pattern was present for within-stage comparisons, as evidenced by higher similarity for same-ritual conditions (i.e., blue and purple) compared to different-ritual conditions (i.e., yellow and orange); $t(39) = 9.565$, $p < 0.001$. As expected, this effect was not present for across-stage comparisons ($t(39) = -1.312$, $p = 0.197$), because the current ritual always differs for across-stage comparisons.

template than the flower template) – the prediction of there being *less* rather than *more* similarity is due to the hypothesized "photo-negative" nature of the code (see Figure 10 and Methods). The whole-brain RSA searchlight approach showed a significant rotated preceding ritual code (FDR corrected) most prominently in medial occipital regions (Figure 4C). Figure 7 shows a posthoc confirmation of the pattern; this figure shows that regions that exhibited a significant rotated preceding ritual code also coded for the current ritual (in a regular, non-rotated way). For completeness, we also searched for a non-rotated preceding ritual code; this would be manifest as exactly the opposite pattern of similarity scores (i.e., similarity should be *especially high* when the template ritual immediately precedes the held-out ritual). As such, we

could use the exact same neural contrast, here looking for a large *negative* correlation between the predicted and observed pattern of similarity scores. No searchlights passed FDR correction for this "non-rotated preceding ritual code" analysis.

2.1.4 Current ritual code (history-agnostic) in occipital regions

For this code, we looked for regions that represent the current ritual while participants are watching the weddings (agnostic of whether the region also represents previous rituals). This was calculated by contrasting (within a particular stage of the wedding) the similarity of neural patterns corresponding to the same ritual (regardless of schema) and neural patterns corresponding to different rituals (regardless of schema; see Figure 10 and Methods). The whole-brain RSA searchlight approach showed a significant current ritual code (FDR corrected) in regions across the occipital cortex (Figure 4D). Figure 8 and Appendix Figure A3 show a posthoc confirmation of the pattern.

2.2 Relation between schema code and memory

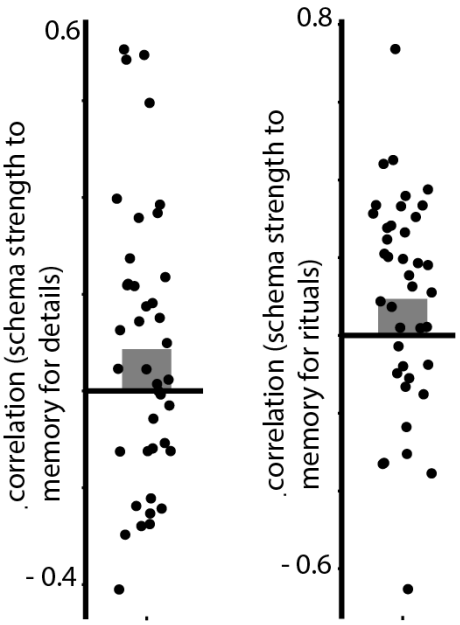

Fig. 9 within-subject correlations of schema strength to memory for details and to memory for rituals.

Together, these results suggest that different types of neural codes that are sensitive to stimulus history are present in the distinct brain regions while viewing temporally extended events. Next, we sought to discover how these neural codes relate to subsequent episodic memory for the weddings. Memory for episodic details was tested with a separate cued recall test that was administered after all of the weddings were shown. In this test, participants were shown a picture of a couple and asked to recall as many details as possible about that couple's wedding. Scoring of participants' memory performance was done manually; we separately scored memory for unique episodic details and for the ritual types that comprised the wedding (see Methods for details). To relate these memory scores to neural data, we first computed the strengths of each of the four neural codes (schema, path, rotated previous ritual, current ritual) on a per-wedding basis within participants (see Methods). We then computed the Spearman correlation (across weddings, within each participant) between each of our neural and behavioral measures. For a given pairing of neural and behavioral measures (e.g., schema code strength vs. memory for

ritual types), this yielded one Spearman correlation coefficient per participant, which we then averaged across participants. To test for statistical significance, we created a null distribution by shuffling and re-running the Spearman correlation 1000 times.

Our results showed that the strength of the neural schema code was significantly correlated with both of our measures of subsequent episodic memory (Figure 9, Table 1). This effect was, as we predicted, specific to the neural code for schema (Table 1); none of the other neural codes significantly correlated with either of the behavioral measures. We also assessed, posthoc, whether the correlations between the neural schema code and behavior were *significantly larger* than the correlations between other neural codes and behavior; while one comparison was significant as measured using a two-tailed paired-samples t-test, most of them were not significant (for memory for rituals: schema > current ritual $t(39) = 2.116$, $p = 0.04$; schema > path $t(39) = 1.501$, $p = 0.142$; schema > rotated $t(39) = 1.632$, $p = 0.11$; for memory for details: schema > current ritual $t(39) = 1.859$, $p = 0.07$; schema > path $t(39) = 1.335$, $p = 0.19$; schema > rotated $t(39) = 0.598$, $p = 0.554$).

Table 1: within-subject brain-behavior correlations

type of code		**r mean ± S.E.M.**	**p value**
schema code	details	0.086 ± 0.045	0.028 *
	rituals	0.094 ± 0.043	0.022 *
path code	details	- 0.006 ± 0.049	0.56
	rituals	- 0.016 ± 0.053	0.61
rotated code	details	0.04 ± 0.059	0.25
	rituals	- 0.02 ± 0.051	0.64
current ritual code	details	- 0.045 ± 0.047	0.811
	rituals	- 0.049 ± 0.053	0.833

2.3 Exploratory analysis: unsupervised clustering of the whole brain

To obtain converging evidence that the four codes we specified *a priori* (schema, path, rotated previous ritual, current ritual) are the primary ones that the brain uses to represent the task, we ran a K-means unsupervised clustering analysis on the representational similarity analysis (RSA) data from each searchlight in the brain; this allowed us identify, in a data-driven fashion, the kinds of codes present in the brain while watching the weddings. If the four codes that we specified are the main ones that the brain uses to represent this task, we should see clusters for these four codes plus a "null cluster" (corresponding to regions not involved in the task) when we try to group the RSA data into 5 clusters (i.e., setting K = 5). If other codes emerge in this clustering analysis, this would suggest that our *a priori* codes might not provide the best or simplest account of the data.

We found that the results for K = 5 align well with the neural codes we discovered with our initial RSA approach (figure 4E): The algorithm split the brain in a "null cluster" (i.e., a set of regions that do not appear to code for relevant variables like schema, path, and ritual; see Appendix Figure A10 and Appendix Figure A11), a cluster whose coding preferences align with the schema code (see Appendix Figure A5 and Appendix Figure A6), a cluster whose coding preferences align with the path code (see Appendix Figure A7 and Appendix Figure A8 – here, the path pattern was more evident when we focused on the searchlights that were most typical of the cluster, as shown in Appendix Figure A9), and a cluster whose coding preferences align with the rotated preceding ritual code (see Appendix Figure A14). The final cluster is best described as a region that codes for the current ritual *only* (see Appendix Figure A12 and Appendix Figure A13), in contrast to the rotated preceding ritual cluster, which codes for the current ritual and *also* codes for the preceding ritual in a rotated fashion. Note that the "current ritual only" cluster that was identified by K-means differs from the "current ritual (history-agnostic)" region identified by RSA, because the (latter) RSA analysis only looks at how the current event is represented and is agnostic about whether the region also represents preceding events. See Appendix Figure A15 for further explanation on how we used K-means for this analysis. Importantly, the fact that the K = 5 results aligned with our *a priori* codes does not rule out the possibility that other codes exist; it just shows that the *a priori* codes are strongly represented in this task.

3 Discussion

In this study, we investigated the neural representations that participants use to disambiguate overlapping sequences during narrative perception, and how these representations relate to memory for episodic details. The results showed that the brain represents such context-dependent temporal structures in multiple distinct ways. First, an abstract schema code (representing whether the couple was from the "North" or "South" culture) was present in a network of regions: thalamus, pallidum, caudate, posterior medial cortex, hippocampus, parahippocampus, fusiform gyrus, inferior and

superior temporal regions, temporal pole and superior frontal regions. At the same time, postcentral gyrus showed a representation of the specific sequence (path) of the wedding, occipital cortex showed a representation of the current ritual, and medial occipital regions represented the preceding ritual in addition to the current ritual. This representation of the preceding ritual in the medial occipital cortex was rotated (van Loon et al., 2018; Wan et al., 2022; Yu et al., 2020), such that the representation of a particular ritual when it occurred in the past was anticorrelated with the representation of that ritual when it was being perceived. Of these four types of neural codes present in the brain while viewing temporally extended events, only the schema code positively correlated (within-participants) with participants' memory performance. This positive within-participant correlation was present for memory for ritual types as well as memory for event details. This finding is consistent with prior work by Masis-Obando et al. (2022) showing a relation between a) the strength of schema representations during perception of (audio-only and audiovisual) narratives and b) subsequent recall of details from those narratives. Altogether, these findings shed light onto the multiple ways that the brain codes for complex and temporally extended events, how it provides structure for such complex events in mind, and how that structure helps to scaffold memory for specific episodic details.

3.1 Codes that we did not observe

In addition to characterizing the codes that we did observe, it is useful to discuss the neural codes that we did not observe in our data. While one must be careful not to infer evidence of absence from absence of evidence, it is notable that there were several plausible types of codes that we did not see here.

3.1.1 Non-rotated previous and upcoming ritual codes

Studies using other paradigms have found non-rotated codes for past and upcoming stimuli. For example, Ezzyat and Davachi (2021) found that neural patterns associated with previous stimuli persisted in non-rotated form in the LO region within events. Also, fMRI studies of navigation have found (non-rotated) prospective representations of goal states (e.g., Brown et al., 2016; Crivelli-Decker et al., 2023).

The presence (or absence) of these non-rotated representations of previous and upcoming states in our study can be assessed by looking at results from RSA comparisons where the template is made based on the stage 2 ritual and then applied to the stage 3 ritual, or vice-versa. For a non-rotated previous ritual code or upcoming ritual code, we would expect to see the opposite of what we observed for the rotated previous ritual code (i.e., elevated similarity when the template is the successor or predecessor of the held-out ritual, instead of the decreased similarity that we observed). We did not see any significant searchlights when we looked for this pattern in our main representational similarity analysis, and it also was not evident in any of the K-means clusters.

There are several differences between the studies that have found these non-rotated previous/upcoming event codes and our study that could help explain why we did not observe these codes here. For example, there was no demand to predict based on

past and current stimulus identity in Ezzyat and Davachi (2021) and thus no need to simultaneously and accurately represent both the past and current stimulus, which may explain why they did not observe a rotated code. Also, the stimuli in Ezzyat and Davachi (2021) were presented for much less time than the rituals in our study. With regard to the absence of prospective codes: In contrast to studies of navigation, participants in our study were passively viewing sequences as opposed to actively navigating, and they also had no way to determine the "end state" of a wedding at its outset (participants could only determine the identity of the final ritual once they saw the campfire or flower ritual).

3.1.2 Rotated upcoming ritual code

We also did not find any evidence for a *prospective* rotated ritual code (i.e., where participants represent a rotated version of the *upcoming* ritual). Note that some of our findings in rotated *preceding* ritual code regions can be explained equally well by a prospective-rotated-code and a retrospective-rotated-code account: In particular, when the template is made based on the stage 2 ritual and then applied to the stage 3 ritual (or vice-versa), both the prospective and retrospective accounts predict that pattern similarity should be reduced when the template is the successor or predecessor of the held-out ritual (for the prospective code, the preceding ritual should contain a "photo-negative" of the following ritual; for the retrospective code, the following ritual should contain a "photo-negative" of the preceding ritual; either way, similarity should be reduced). However, the expected pattern of RSA results when the template is made based on stage 2 and then applied to stage 2 will differ depending on whether the rotated code is prospective or retrospective. If the code is prospective, then rituals with the same successor should show an elevated correlation, but that was not true of our rotated previous ritual code region (for the rotated code region from the RSA, $t(39) = -1.098$, $p = 0.279$; and for the rotated code region from the K-means cluster, $t(39) = -0.520$, $p = 0.606$).

3.2 Relationship to other studies

3.2.1 Schema code

As noted above, we found a schema code in a wide range of regions. The most comparable study to ours is Baldassano et al. (2018), which also looked at schema representations for temporally extended narratives, although the nature of schemas they investigated was different: While we used a novel schema (devised by us), Baldassano et al. (2018) examined schema representations for commonly-experienced events like going through airport security and ordering food at a restaurant (see also Masis-Obando et al., 2022, which reported analyses of the memory retrieval data from this study). These differences notwithstanding, it is still useful to compare our results to theirs.

Several of the schema code regions that we observed here were also found by Baldassano et al. (2018), including posterior medial cortex (see also: Bonasia et al., 2018; Maguire et al., 1999; Sommer, 2016; van Kesteren et al., 2013), superior frontal gyrus (see also: Bonasia et al., 2018; Brod and Shing, 2018), parahippocampal cortex

(see also: Bonasia et al., 2018; De Soares et al., 2023; Keidel et al., 2018; Liu et al., 2017; van der Linden et al., 2017; van Kesteren et al., 2014), angular gyrus (see also: Bonasia et al., 2018; De Soares et al., 2023; Keidel et al., 2018; van der Linden et al., 2017), hippocampus (see also: Liu et al., 2017; Milivojevic et al., 2016; Preston and Eichenbaum, 2013; Raykov et al., 2020; van Kesteren et al., 2014, 2013; Varga et al., 2022), entorhinal cortex, left subcentral gyrus, postcentral sulcus, left postcentral gyrus, superior temporal sulcus, left posterior temporal sulcus, insula, and prostriata.

A notable difference between our findings and those in Baldassano et al. (2018) is that we did not find a schema code in mPFC; this stands in sharp contrast to a long list of studies that have found a role for mPFC in schema representation (e.g., Audrain and McAndrews, 2022; Baldassano et al., 2018; Bein and Niv, 2023; Bonasia et al., 2018; Brod and Shing, 2018; De Soares et al., 2023; Gilboa and Marlatte, 2017; Liu et al., 2017; Preston and Eichenbaum, 2013; Raykov et al., 2020; Reagh and Ranganath, 2023; Schwarb et al., 2019; Sommer, 2016; van Kesteren et al., 2014, 2010, 2012, 2020; Varga et al., 2022). One speculative explanation for the lack of schema codes in mPFC in our study is that the amount of initial exposure to the schemas was insufficient; another possibility is that the interval between learning the schemas and viewing of the novel wedding videos (i.e., approximately 24 hours) was too short for the schema code to consolidate in mPFC (for evidence that these codes may take time to consolidate, see, e.g., Tompary and Davachi, 2017).

Unlike the Baldassano et al. (2018) study, which only investigated schema codes on the cortical surface and in the hippocampus, we also investigated other subcortical regions. We included the subcortical regions because prior research has established the importance of some subcortical regions (i.e., basal ganglia and thalamus) in supporting interactions between hippocampus and frontal cortex (Preston & Eichenbaum, 2013). We indeed found a robust schema code in the thalamus and basal ganglia regions (pallidum, caudate).

3.2.2 Rotated previous ritual code

Besides a schema code, we also discovered a rotated code in which the sensory representation (current ritual) is anti-correlated to the memory representation (preceding ritual). As outlined in the introduction, multiple animal (Libby & Buschman, 2021) and human (van Loon et al., 2018; Wan et al., 2022; Yu et al., 2020) studies have found rotated codes for sensory versus memory representations, which are thought to be a way to minimize confusion between information corresponding to currently-perceived stimuli and information being held in working memory. The two previous fMRI studies that observed rotated codes used working memory tasks with simple stimuli: objects (van Loon et al., 2018) and gratings (Yu et al., 2020); the rotated codes in these studies were observed in posterior fusiform and visual cortex, respectively. Our study is the first to demonstrate rotated codes for temporally-extended audio-visual narrative stimuli – here, the rotated codes were localized to medial occipital cortex. The difference in localization of the rotated codes between our study and previous studies likely relates to our use of more complex stimuli.

3.2.3 Path code

We also found a path code, separately representing each of the four paths (or sequences) present in our wedding video stimuli. Our results showed this path code most strongly in postcentral gyrus, but not in the hippocampus. This differs from previous multivariate fMRI studies of sequence coding, which have found that representation of predictable sequences that are context-dependent is localized to the hippocampus (Clarke et al., 2022; Hsieh et al., 2014); specifically, these studies found that hippocampus assigns distinct neural codes to identical items that are part of distinct sequences. Relatedly, studies using statistical learning and associative inference paradigms have also demonstrated that hippocampus plays a key role in integrating items that share common associates or appear adjacently in space or time (e.g., Collin et al., 2015; Morton et al., 2020; Schapiro et al., 2012, 2013; Schlichting and Preston, 2015; Thavabalasingam et al., 2019; Zeithamova and Preston, 2010; for a computational model of this process, see Schapiro et al., 2017). In light of this, why did we not see hippocampal path coding in our study? One possibility is that it was superseded by schema coding in the hippocampus – the aforementioned studies that found sequence coding in the hippocampus with fMRI did not have a hierarchical structure like ours did (i.e., where sequences were nested within the North / South schemas). It is possible that the hippocampus might preferentially represent the highest possible level of disambiguating structure in the environment (here, North / South); alternatively, hippocampus might simultaneously represent multiple levels in the hierarchy (McKenzie et al., 2014) but the presence of higher-level codes might make it harder to detect the lower-level codes with fMRI. Another reason we might not have detected path codes in the hippocampus is that our analysis assumed that the codes were completely stable within a path, whereas other work has shown that hippocampal codes for sequences drift within the sequence (e.g., Hsieh et al., 2014) in the manner predicted by temporal context models (Howard & Kahana, 2002).

3.3 Relation between schema code and episodic memory

As noted above, Masis-Obando et al. (2022) found a relationship between the strength of schema representation during encoding of narratives (both audiovisual movies and audio narratives) and subsequent memory for details from those narratives. We obtained a similar relationship here, using a very different paradigm and a schema that we taught participants (rather than pre-existing schemas from everyday life), demonstrating the generality of this finding. An important novel contribution of our study is that we were also able to assess the relationship between other kinds of history-dependent representations (rotated previous-event codes, path codes) and subsequent memory for details – this allowed us to address the question of whether all history-dependent representations help to scaffold subsequent memory, or whether high-level schema representations are special in this regard. In keeping with this hypothesis, North/South schema representations were the only ones to significantly correlate with subsequent memory. However, at the same time, the correlations we observed for the schema code were not (for the most part) significantly larger than the correlations we

observed for the other codes. As such, we are not in a position to say that the schema code is a reliably better predictor of subsequent memory, compared to the other codes.

4 Conclusion

In this study, we discovered that multiple history-dependent neural codes exist in in the brain: an abstract schema code, a path (sequence) code, and a rotated previous ritual code that represents current and previous rituals in an anti-correlated fashion, likely to avoid interference. We also found that the strength of the neural schema code during viewing of a particular wedding correlated with subsequent memory for the details of that wedding; this provides converging neural support for the idea that schemas act to scaffold memory for unique episodic details.

5 Methods

5.1 Participants

Forty-four participants with normal or corrected-to-normal vision and no hearing impairments participated in this 2-day fMRI study. Four participants did not finish the entire study. Therefore, the final sample consisted of 40 participants (11 male, 29 female, ages 18-35, mean age 21.4). All participants were either native English speakers or started speaking English before the age of seven. Two participants were left-handed. All participants provided written informed consent before the start of the study in accordance with experimental procedures approved by the Princeton University Institutional Review Board (IRB protocol 7883). Participants received monetary compensation for their time (20 USD per hour, and additionally a performance bonus of 10 USD when completing both days of the experiment). The study consisted of two sessions on two consecutive days (both 2h long). No statistical methods were used to predetermine the sample size, but our sample size is similar to those reported in similar publications of neuroimaging schema experiments (e.g., see Baldassano et al., 2018 and De Soares et al., 2023, both of which tested between 30 and 40 participants).

5.2 Materials

We exposed participants to many unique animated videos of wedding ceremonies (generated with Unity, https://unity3d.com/) with fictional rituals originating from a novel (fictional) culture, where each unique video was accompanied by a unique (audio) narrative. The audio was imported into Audacity for further processing before it was integrated with the video. The audio files were imported together, and then normalized. Remaining audio processing was done in iMovie: the auto sound editor, background noise reducer, and equalizer ("voice enhance") were used. Furthermore, the volume of each video was adjusted to a subjectively good level. Each of these ceremonies had a duration of approximately two minutes and were composed of a sequence of rituals. Participants had to discover how the sequence of these stages in a ceremony could be predicted based on the context (crucially, participants were not shown the ground-truth transition structure in Figure 1 or directly informed that the transition

structure depended on the couple being from the North vs. South; they had to discover these contextual dependencies on their own by watching the weddings). Each ceremony, despite the regularities in type and order of the rituals, was unique with respect to details (e.g. the characters, objects, and audio narrative).

When a wedding couple came from the North, the rituals performed during their wedding followed one of the following two paths:

- North (path-A): start of the wedding - celebrate around campfire - drop coin in bowl - break an egg - receive gifts
- North (path-B): start of the wedding - plant a flower - hold a torch - draw a painting - receive gifts

When a couple came from the South, the rituals performed during their wedding followed one of the following two paths:

- South (path-C): start of the wedding - celebrate around campfire - hold a torch - break an egg - receive gifts
- South (path-D): start of the wedding - plant a flower - drop coin in bowl - draw a painting - receive gifts

5.3 Experimental procedure

5.3.1 Day 1

Participants were told that they would be watching animated videos that depict rituals performed during wedding ceremonies from couples that all live on a certain island. Additionally, they were told that some couples live in the North part of the island and some in the South part of the island, but that they all traveled to the city Navahla for their wedding. They were informed that it is their task to learn to predict which rituals are more likely to happen next. The videos were projected using an LCD projector onto a rear-projection screen located in the magnet bore and viewed with an angled mirror. PsychoPy was used to display the task. Audio was delivered using in-head earphones.

Learning Participants were exposed to 24 ceremonies (12 of each context) in the following order: two blocks of the first context (either North or South, counterbalanced across participants) of six ceremonies each, followed by two blocks of the other context with six ceremonies each. The videos were presented continuously. There was a break of two minutes after each block. To assess their learning of the four paths, they received two-alternative forced choice questions asking them to predict what would happen next. They received these questions during the first and last weddings from each of these four blocks. Within these weddings, the questions were presented right before the start of stage three (drop coin in bowl versus hold a torch) and right before the start of stage four (break an egg versus draw a painting). Participants had a fixed time window of four seconds to answer these prediction questions. Additionally, to keep the participants engaged with the task continuously throughout the experiment, they received six two-alternative forced choice questions after each wedding (the presentation time was two seconds per question, no ITI) asking which episodic detail (of two random episodic details presented) they saw during the preceding wedding. They

gave their answer using a button box. We monitored people's alertness throughout the task via an MRI-compatible eye tracking camera. fMRI data from day 1 are not reported in this paper.

5.3.2 Day 2

Pretraining ("meet the inhabitants") Before the start of the MRI-session of day two, participants performed one task outside of the scanner behind a computer screen. During this task, they were introduced to new characters (different ones from the ones they saw on the first day). They were instructed that they would again see a number of ceremonies today, and that they would first meet the inhabitants of the island whose ceremonies they would be viewing, before seeing the actual ceremonies. They received a picture of each character in isolation (24 characters in total). First, they received two blocks in which they saw all 24 characters along with their names, one at the time, and were told to try to remember their names. Next, they received two blocks in which they were only presented with pictures of the characters, again one at the time, and were told to write the person's name in the answer field. They received feedback about whether they were correct or not. If they were incorrect, the feedback stated the correct name.

Watching novel ceremonies After the pretraining, participants moved on to the MRI-session of that day. Participants were exposed to a novel set of six North and six South ceremonies in interleaved order (i.e., SNNSSNNSSNNS). The videos were presented continuously. While watching, participants again received two-alternative forced choice questions asking to predict what will happen next during the first and last ceremony they saw. Again, these questions were presented right before the start of stage three (drop coin in bowl versus hold a torch) and right before the start of stage four (break an egg versus draw a painting). Participants had a fixed time window of four seconds to answer these prediction questions. They again were given the six general attention questions at the end of each ceremony to keep them engaged with the ceremonies.

Cued recall test After they completed watching the novel ceremonies, participants performed a spoken recall session in the MRI-scanner. During this recall session, participants received pictures of the wedding couples and their names, one at the time, as a cue. They were instructed to describe anything they recalled from that couple's wedding. They were specifically instructed that any detail they remembered would be important to mention. Participants were instructed to indicate verbally when they were finished, after which the experimenter would move on to the next couple. Participants' speech was recorded using a customized MR-compatible recording system. Both watching the novel ceremonies as well as the recall test were displayed using PsychoPy. The fMRI-data associated with this recall test was not taken into consideration for the purpose of this article.

Test for schema learning Outside of the MRI-scanner, we tested their memory for the ritual transitions during each of the four paths. They received questions like, "If a couple from the North just planted a flower, what is the most likely ritual to happen next?". Participants were given all possible answer options (i.e., all possible rituals) and were told to allocate a total of 100 percent across these answer options,

explaining that they were free to either put the entire 100 percent on one answer or split it across multiple answers. They received eight questions:

- North couple, right after celebrating around a campfire,
- North couple, right after planting a flower,
- North couple, right after dropping a coin in a bowl,
- North couple, right after holding a torch,
- South couple, right after celebrating around a campfire,
- South couple, right after planting a flower,
- South couple, right after dropping a coin in a bowl,
- South couple, right after holding a torch

5.4 MRI acquisition

fMRI data were collected on a 3T Prisma scanner (Siemens). Functional data was acquired using a whole-brain multiband EPI sequence, TR = 1.5 sec, TE = 30 ms, voxel size 2mm isotrophic, multiband factor = 3. We collected two anatomical scans (MPRAGE), one on each day, voxel size 1mm isotrophic, TR = 2300ms, TE = 2.96ms. To correct for distortions, we collected two fieldmaps on each day; one in the anterior-posterior phase encoding direction and one in the posterior-anterior phase encoding direction, TR = 8000ms, TE = 66ms, voxel size 2mm isotrophic. Additionally, we collected a T2 turbo-spin-echo scan (to create subject-specific hippocampal subfields masks using ASHS, https://www.nitrc.org/projects/ashs).

5.5 MRI data preprocessing

Results included in this manuscript come from preprocessing performed using fMRIPprep 1.2.3 (Esteban et al., 2019; Esteban et al., 2018), which is based on Nipype 1.1.6-dev (Gorgolewski et al., 2011; Gorgolewski et al., 2018).

5.5.1 Anatomical data preprocessing

There were 2 T1-weighted (T1w) images acquired (one on each day). Both were corrected for intensity non-uniformity (INU) using N4BiasFieldCorrection (Tustison et al., 2010, ANTs 2.2.0). A T1w-reference map was computed after registration of 2 T1w images (after INU-correction) using mri robust template (FreeSurfer 6.0.1, Reuter et al., 2010). The T1w-reference was then skull-stripped using antsBrainExtraction.sh (ANTs 2.2.0), using OASIS as target template. Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0, Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, Zhang et al., 2001).

5.5.2 Functional data preprocessing

For each of the 6 BOLD runs found per subject (day1-run1, day1-run2, day1-run3, day1-run4, day2-video watching, day2-recall), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al., 2002). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD time-series were resampled to MNI152NLin2009cAsym standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). Six tCompCor components are then calculated from the top 5 percent variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, six components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The BOLD time-series, were resampled to surfaces on the following spaces: fsaverage6. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and template spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using mri vol2surf (FreeSurfer). Many internal operations of fMRIPrep use Nilearn 0.4.2 (Abraham et al., 2014), mostly within the functional processing workflow. For more details of the pipeline, see https://fmriprep.org/en/latest/workflows.html.

5.5.3 Temporal high pass filter cut off

Given that we were looking for an expected slow change in neural pattern (i.e., a change in neural pattern when there is a switch between schemas or extended events), we could not use the default high pass filter cut off that is usually used (i.e., 120 s). However, some high pass filtering was still necessary due to activity that can be attributed to non-neuronal noise (e.g., scanner drift). We preferred to use the longest possible high-pass filter cut off that still removes the non-neuronal noise. Following the analysis of high-pass filter cut offs for stimuli with long timescales that was performed in Lositsky et al. (2016), we opted for a high-pass filter cut off of 480 s.

5.6 Analyses

5.6.1 Behavioral data

We assessed the learning of the schema in two ways: [1] By calculating a percentage correct for the two-alternative forced choice prediction questions on day 1 as well as day 2. [2] By calculating the percentage that people allocated to the correct answer option during the schema learning test at the end of day 2 and comparing that to the percentage people allocated to the opposite answer option. Because the data were not normally distributed, we used a Wilcoxon signed-rank test to compare the percentage assigned to the correct vs. opposite answer option. In addition to checking the schema learning throughout and at the end of the experiment, we assessed whether people were sufficiently attentive throughout the tasks. This was done by calculating a percentage correct for the two-alternative forced-choice detail questions at the end of each stimulus during video watching. Lastly, we assessed how well people remembered the episodic details of the weddings. This was done by analyzing the behavioral data acquired during the recall task in the following ways:

- By manually scoring which episodic details were mentioned during recall by each of the participants for each of the weddings separately, leading to a score for each participant and each wedding regarding the amount of correctly and incorrectly remembered details. To get to a score for their memory for details, we then subtracted the number of incorrectly remembered details from the number of correctly remembered details.
- By manually scoring which ritual types were mentioned during recall for each participant and for each of the weddings separately, leading to a score for each participant and each wedding regarding the amount of correctly and the amount of incorrectly mentioned ritual types. To get to a score for their memory for ritual types, we then subtracted the number of incorrectly remembered ritual types from the number of correctly remembered ritual types.

See https://osf.io/u3cfr/ for all video stimuli used for this experiment as well as the written version of all stimuli. For scoring the amount of details remembered, we used the following list of possible details that participants could mention (see Appendix).

5.6.2 Representational similarity analysis

We hypothesized that the brain uses several types of codes to represent the context-dependent temporal structure in this study, including a schema (North/South) code, a path code, and a rotated preceding ritual code (Figure 2 and Figure 10B), as we will describe in more detail below. To calculate each of these codes, we analyzed the multivoxel pattern of neural activity by applying a searchlight approach on our whole-brain data. Specifically, we used representational similarity analysis (RSA) to examine the correlation coefficients between patterns of activity within searchlights throughout the whole brain (the searchlights each had a radius of 2 voxels excluding the center voxel – thus, each searchlight encompassed $((2 * 2) + 1)^3 = 125$ voxels). For these analyses, we used the fMRI data from the part of the wedding videos during which participants watched the actual rituals (i.e., campfire/flower, which below is referred to as stage 2, coin/torch, which below is referred to as stage 3, and egg/painting, which below is referred to as stage 4). We calculated an average neural pattern for each combination of ritual and schema (12 patterns total: [campfire, flower, coin, torch, egg, painting] X [North, South]) across all weddings but one (we refer to these as the 12 "template patterns"), and then compared these 12 template patterns to the average neural patterns for each of the 3 rituals of the held-out wedding (Figure 10A); this yields 36 (12 x 3) comparisons in total. We computed these 36 comparisons for each possible held-out wedding. For the analyses in the paper, we sorted the comparisons according to the following criteria:

- First, the comparisons were sorted according to the stage of the template ritual (2, 3, 4) and the stage of the held-out ritual (2, 3, 4). For some analyses, we also grouped these more generally based on whether the comparison was *within-stage* (e.g., template = stage 2, held-out ritual = stage 2) or *across stage* (e.g., template = stage 2, held-out ritual = stage 3).
- Second, the comparisons were also sorted according to the relationship between the path of the template ritual and the path of the held-out ritual: There are 4 possibilities here: 1) same path and schema; 2) different path, same schema; 3) different schema, none of the rituals on the path of the held-out ritual match the template ritual (e.g., template is north-coin, path of held-out-ritual is south-campfire-torch-egg); 4) different schema, path of held-out ritual matches the template ritual for within-stage comparisons but not across-stage comparisons (e.g., template is north-coin, path of held-out-ritual is south-flower-coin-painting).

For some analyses, we also noted the following situations:

- whether the template ritual and the held-out ritual had the *same preceding ritual*, even if the template and the held-out ritual were not themselves the same (e.g., torch-North and coin-South are both preceded by flower)
- whether the template ritual matched the ritual that immediately preceded or followed the held-out wedding; e.g., if the template is campfire-North, and the held-out ritual is coin-North, then the template ritual does not match the held-out ritual, but the template ritual matches the ritual that *immediately preceded* the held-out ritual (this is relevant for the "rotated preceding ritual code" as described below)

Schema code This code would involve one pattern being active for rituals from North weddings and a distinct pattern being active for rituals from South weddings, regardless of which specific ritual and path is being shown at that time. For the schema code, we separately averaged (within each participant) all comparisons involving the same schema (including both within-stage and across-stage comparisons) and all comparisons involving the other schema (including both within-stage and across-stage comparisons). We then computed, for each participant, the average difference in similarity between same-schema and other-schema comparisons, and assessed the reliability of this difference across participants. Figure 10B shows this contrast in RSA matrix form. We also computed this "schema code" difference score (same-schema minus other-schema) for each individual wedding within each participant (focusing on the regions showing a significant FDR-corrected schema code effect in Figure 4), for the analyses described in *Relating brain to behavior* below.

Path code This code would involve distinct patterns being active for rituals from different paths while watching the ceremonies, regardless of which ritual is being shown at that time. For the path code, we computed, for each participant, the difference in similarity between same-path, same-schema and different-path, same-schema comparisons, and assessed the reliability of this difference across participants. Figure 10B shows this contrast in RSA matrix form (note that we omitted different-path, different-schema comparisons from this RSA matrix – the benefit of doing this is that the RSA matrix for the path code is uncorrelated with the RSA matrix for the schema code). We also computed this "path code" difference score (same-path, same-schema minus different-path, same-schema) for each individual wedding within each participant (focusing on the regions showing a significant FDR-corrected path code effect in Figure 4), for the analyses described in *Relating brain to behavior* below.

Rotated preceding ritual code With this code type, we expect that the stage 3 pattern will contain a rotated representation of the preceding (stage 2) ritual. This implies that similarity values should be *especially low* when the template applied to the stage 3 pattern matches the stage 2 ritual that preceded it. Conversely, similarity values should also be especially low when the template applied to the stage 2 pattern matches the stage 3 ritual that followed it. To compute this, we first looked at cross-stage comparisons where stage 2 templates were matched with held-out stage 3 rituals, and we computed the difference between similarity values when the template applied to the stage 3 pattern *mismatched* the stage 2 ritual that preceded it, vs. when it *matched* the stage 2 ritual that preceded it. Next, we looked at cross-stage comparisons where stage 3 templates were matched with held-out stage 2 rituals, and we computed the difference between similarity values when the template applied to the stage 2 pattern *mismatched* the stage 3 ritual that followed it, vs. when it *matched* the stage 3 ritual that followed it. We then averaged (within participants) these two differences to create a single "rotated preceding ritual code" for each participant, and assessed the reliability of this difference across participants. To get the most clean measure of the rotated preceding ritual code, we excluded stage 4 comparisons (since one would expect residual codes from multiple preceding stages during stage 4, which is not the case yet in stage 2 and 3). Figure 10B shows this contrast in RSA matrix form. We also computed this "rotated preceding ritual code" difference score for each individual

wedding within each participant (focusing on the regions showing a significant FDR-corrected rotated code effect in Figure 4), for the analyses described in *Relating brain to behavior* below.

Current ritual code In addition to the history-dependent codes described above, we also computed a simple measure of how well the current ritual is being represented. For this "current ritual" code, we separately averaged (within each participant) all within-stage comparisons involving the same ritual and all comparisons involving the other ritual for that stage. We then computed, for each participant, the average difference in similarity between same-ritual and different-ritual comparisons, and assessed the reliability of this difference across participants. Figure 10B shows this contrast in RSA matrix form. We also computed this "current ritual code" difference score (same-ritual minus other-ritual) for each individual wedding within each participant (focusing on the regions showing a significant FDR-corrected current ritual code effect in Figure 4), for the analyses described in *Relating brain to behavior* below. Note that the current ritual code, defined this way, is "history-agnostic" – a region that scores well according to this metric may or may not also score well on other, history-dependent codes (e.g., the rotated preceding ritual code described above). In particular, note that the current ritual code only looks at *within-stage* comparisons and the rotated preceding ritual code only looks at *across-stage* comparisons. Since they rely on distinct comparisons, there is no strict mathematical dependence between these codes; in principle, a region could score well according to the current ritual code but not the rotated preceding ritual code, or vice-versa, or it could score well on both, or neither.

To determine statistical significance, we used the multipletests function from the statsmodels python package to convert uncorrected p-values to q-values of the whole brain searchlights of all 4 codes described above (i.e., FDR correction).

5.6.3 K-means analysis

The K-means analysis was run in the following manner: For each searchlight location, we computed the full set of similarity comparisons described in the *Representational similarity analysis* section, yielding a "similarity fingerprint" for that searchlight (see Appendix Figure A15 for a detailed description of how these fingerprints were created). After acquiring this similarity fingerprint for each searchlight, we ran the scikit-learn (Pedregosa et al., 2011) K-means unsupervised clustering algorithm with K = 5 on these fingerprints to identify regions with similar fingerprints; importantly, this clustering approach does not factor in the spatial location of the searchlight, so the clusters it finds are not constrained to be spatially contiguous.

To assess the "goodness of fit" of the K-means analysis with K = 5, relative to other values of K, we used silhouette scores (Rousseeuw, 1987). These scores index how close each data point is to other data points in its own cluster, compared to data points in the closest neighboring cluster that the data point is not assigned to. The silhouette scores for the K = 5 analysis are shown in Appendix Figure A4, and Appendix Figure A16 shows the average silhouette score for each cluster for K = 2 through 7. The K = 5 solution stands out for having good silhouette scores for all of

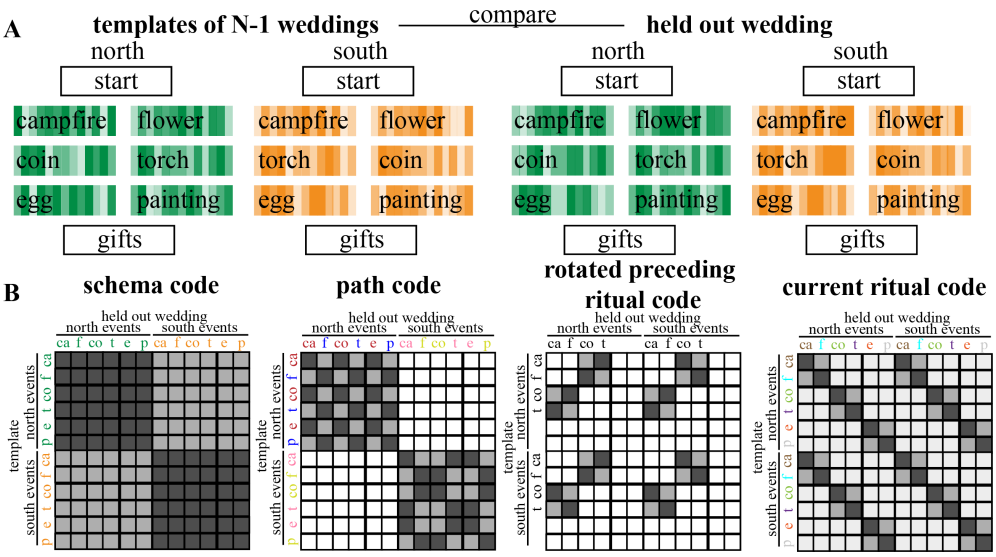

Fig. 10 Explanation of the procedure used for the searchlight RSA. (A) We calculated an average neural pattern for each combination of ritual (campfire, flower, coin, torch, egg, painting) and schema across all weddings but one (referred to as "template patterns"), and then compared these template patterns to the neural patterns associated with each ritual of a single left-out-wedding; we repeated this process, holding out a different wedding each time. (B) Matrices that show the 4 types of neural codes. Light grey means low similarity is predicted and dark grey means high similarity is predicted. White squares were omitted from the analysis.

the clusters. This can be operationalized by looking at the minimum of the cluster-wise average silhouette scores for each of the clustering solutions; when we focus on these minimum scores ("how good was the worst cluster"), we can see that K = 5 stands out for having the highest minimum cluster-wise silhouette score.

Another way to use the silhouette scores is to "purify" the clusters found by K-means: i.e., sub-selecting searchlights with the highest silhouette scores within a cluster yields the searchlights that are most representative of that cluster (i.e., closer to its center). In Appendix Figure A9 we use this to show that the "purified" version of cluster 2 identified by K-means embodies the characteristics that we would expect of the path code.

5.6.4 Relating brain to behavior

We related the brain results to individual participants' behavior. The behavioral measures that we used to measure memory for specific weddings were as follows:

- Correct minus incorrect episodic details remembered
- Correct minus incorrect rituals remembered

We hypothesized that these behavioral measures would be positively correlated to the neural strength of the schema code, in line with our prediction that the underlying schema is used as a way to structure memory of episodic details. We correlated these

behavioral measures to the neural strength of each of the history-dependent codes (schema, path, rotated) and to the current ritual code within-participant using Spearman correlations. To test for statistical significance of these correlations, we calculated a null distribution by shuffling to behavioral scores and re-running the correlation 1000 times and comparing the actual correlation score to the null distribution.

Author CRediT statement. **Silvy H.P. Collin**: Conceptualization, Software, Formal analysis, Investigation, Data Curation, Writing - original draft, Visualization. **Ross P. Kempner**: Software, Formal analysis, Writing - Review and Editing, Visualization. **Sunita Srivatsan**: Software, Formal analysis. **Kenneth A. Norman**: Conceptualization, Supervision, Funding acquisition, Project administration, Writing - Review and Editing.

Acknowledgements. We thank Kalyana Duggal, Alexandra de Soares, Hope Kean and Olivia Reblando for help during stimulus development; Andre Beukers for devising the context-dependent learning structure used in this paper; Uri Hasson for valuable input on the experimental design; Jeongmin Cho for help with data analysis; and other members of the Hasson and Norman laboratories for their comments and support. This work was supported by a Multi-University Research Initiative Grant (ONR/DoDN00014-17-1-2961) and an NWO Rubicon grant (446-17-009).

References

Abbott, V., Black, J. B., & Smith, E. E. (1985). The representation of scripts in memory. *Journal of Memory and Language*, *24*(2), 179–199. https://doi.org/10.1016/0749-596X(85)90023-3

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. https://doi.org/10.1037//0033-2909.93.2.203

Audrain, S., & McAndrews, M. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. *Nat Commun*, *13*, 5795. https://doi.org/10.1038/s41467-022-33517-0

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *The Journal of Neuroscience*, *38*(45), 9689–9699. https://www.jneurosci.org/content/38/45/9689

Behzadi, Y., Restom, K., Liau, J., & Liu, T. (2007). A component based noise correction method (compcor) for bold and perfusion based fmri. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bein, O., & Niv, Y. (2023, September). *Schemas, reinforcement learning, and the medial prefrontal cortex.* https://doi.org/https://doi.org/10.31234/osf.io/spxq9

Beukers, A., Collin, S., Kempner, R., Franklin, N., Gershman, S., & Norman, K. (2023). Blocked training facilitates learning of multiple schemas. *PsyArXiv.* https://psyarxiv.com/9bptj/

Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiology of Learning and Memory*, *153*(Pt A), 26–39. https://doi.org/10.1016/j.nlm.2018.02.017

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220. https://doi.org/10.1016/0010-0285(79)90009-4

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726. doi:%2010.1016/S0022-5371(72)80006-9

Brod, G., & Shing, Y. L. (2018). Specifying the role of the ventromedial prefrontal cortex in memory formation. *Neuropsychologia*, *111*, 8–15. https://doi.org/10.1016/j.neuropsychologia.2018.01.005

Brown, T., Carr, V., LaRocque, K., Favila, S., Gordon, A., Bowles, B., JN, B., & Wagner, A. (2016). Prospective representation of navigational goals in the human hippocampus. *Science*, *352*(6291), 1323–1326. https://doi.org/10.1038/s41467-023-35967-6

Clarke, A., Crivelli-Decker, J., & Ranganath, C. (2022). Contextual expectations shape cortical reinstatement of sensory representations. *Journal of Neuroscience*, *42*(30), 5956–5965. 10.1523/JNEUROSCI.2045-21.2022

Collin, S. H., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(oct), 1562–1564. http://dx.doi.org/10.1038/nn.4138

Crivelli-Decker, J., Clarke, A., Park, S., Huffman, D., Boorman, E., & Ranganath, C. (2023). Goal-oriented representations in the human hippocampus during planning and navigation. *Nature Communications*, *14*(2946). https://doi.org/10.1038/s41467-023-35967-6

Dale, A. M., Fischl, B., & Sereno, M. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–94. https://doi.org/10.1006/nimg.1998.0395

De Soares, A., Kim, T., Zhu, E., Lin, A., Zheng, C., & Baldassano, C. (2023). *Top-down attention shifts behavioral and neural event boundaries in narratives with overlapping event scripts.* https://doi.org/10.1101/2023.08.08.552465

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift. *Current Opinion in Behavioral Sciences*, *17*(oct), 141–146. 10.1016/j.cobeha.2017.08.003

Esteban, O., Markiewicz, C., Blair, R., Moodie, C., Isik, I., Errazmuzpe, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez,

J., Poldrack, R., & Gorgolewski, K. (2019). Fmriprep: A robust preprocessing pipeline for functional mri. *Nature Methods*, *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban et al. (2018). *Fmriprep* (Version 1.2.3). https://doi.org/10.5281/zenodo.852659

Ezzyat, Y., & Davachi, L. (2021). Neural evidence for representational persistence within events. *Society for Neuroscience*, *41*(37), 7909–7920. DOI:%2010.1523/JNEUROSCI.0073-21.2021

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. http://dx.doi.org/10.1016/j.tics.2017.04.013

Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Satrajit, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.*, *5*(13), 1–15. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski et al. (2018). *Nipype* (Version 1.1.6). https://doi.org/10.5281/zenodo.596855

Greve, D., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Howard, M., & Kahana, M. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hsieh, L., Gruber, M., Jenkins, L., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, *81*(5), 1165–1178. 10.1016/j.neuron.2014.01.015

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–41. https://doi.org/10.1006/nimg.2002.1132

Keidel, J. L., Oedekoven, C. S., Tut, A. C., & Bird, C. M. (2018). Multiscale integration of contextual information during a naturalistic task. *Cerebral Cortex*, *28*(10), 3531–3539. https://doi.org/10.1093/cercor/bhx218

Klein, A., Ghosh, S., Bao, F., Giard, J., Häme, Y., Stavsky, E., & Lee, N. e. a. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 4.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, *24*(apr), 715–726. https://www.nature.com/articles/s41593-021-00821-9

Liu, Z. X., Grady, C., & Moscovitch, M. (2017). Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cerebral Cortex*, *27*(3), 1991–2009. https://doi.org/10.1093/cercor/bhw047

Lositsky, O., Chen, J., Toker, D., Honey, C., Shvartsman, M., Poppenk, J., Hasson, U., & Norman, K. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *eLife*, *5*, e16070. https://doi.org/10.7554/eLife.16070

Maguire, E. A., Frith, C. D., & Morris, R. G. M. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. *Brain*, *122*(10), 1839–1850. https://doi.org/10.1093/brain/122.10.1839

Masis-Obando, R., Norman, K. A., & Baldassano, C. (2022). Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *eLife*, *11*, e70445. https://elifesciences.org/articles/70445

McClelland, J. L., McNaughton, B. L., & Lampinen, A. K. (2020). Integration of new information in memory: New insights from a complementary learning systems perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1799). https://doi.org/10.1098/rstb.2019.0637

McKenzie, S., Frank, A., Kinsky, N., Porter, P., B an Riviere, & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202–215. DOI:https://doi.org/10.1016/j.neuron.2014.05.019

Milivojevic, B., Varadinov, M., Vicente Grabovetsky, A., Collin, S., & Doeller, C. F. (2016). Coding of event nodes and narrative context in the hippocampus. *Journal of Neuroscience*, *36*(49), 12412–12424. 10.1523/JNEUROSCI.2889-15.2016

Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proc Natl Acad Sci USA*, *117*(47), 29338–29345. 10.1073/pnas.1912338117

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Power, J., Mitra, A., Laumann, T., Snyder, A., Schlaggar, B., & Petersen, S. (2014). Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, *84*, 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048

Preston, A., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current Biology*, *23*(17), 764–773. http://dx.doi.org/10.1016/j.cub.2013.05.041

Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2020). The brain regions supporting schema-related processing of people's identities. *Cognitive Neuropsychology*, *37*(1-2), 8–24. https://doi.org/10.1080/02643294.2019.1685958

Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, *14*(1), 1279.

Reuter, M., Rosas, H., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–96. https://doi.org/10.1016/j.neuroimage.2010.07.020

Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, *20*, 53–65. doi:10.1016/0377-0427(87)90125-7

Schank, R. C., & Abelson, R. P. (1975). Scripts, plans and knowledge. *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, 151–157.

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current biology*, *22*(17), 1622–1627.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(feb), 486–492. https://www.nature.com/articles/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049.

Schlichting, M., & Preston, A. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. http://dx.doi.org/10.1016/j.cobeha.2014.07.005

Schwarb, H., Johnson, C. L., Dulas, M. R., McGarry, M. D. J., Holtrop, J. L., Watson, P. D., Wang, J. X., Voss, J. L., Sutton, B. P., & Cohen, N. J. (2019). Structural and functional mri evidence for distinct medial temporal and prefrontal roles in context-dependent relational memory. *J Cogn Neurosci.*, *31*(12), 1857–1872. doi:%2010.1162/jocn_a_01454

Shin, Y., & DuBrow, S. (2020). Structuring memory through inference-based event segmentation. *Topics in cognitive science*, *13*(1), 106–127. https://doi.org/10.1111/tops.12505

Sommer, T. (2016). The emergence of knowledge and how it supports the memory for novel related information. *Cerebral Cortex*, *27*(3), 1906–1921. https://doi.org/10.1093/cercor/bhw031

Thavabalasingam, S., O'Neil, E. B., Tay, J., Nestor, A., & Lee, A. C. H. (2019). Evidence for the incorporation of temporal duration information in human hippocampal long-term memory sequence representations. *Proceedings of the*

National Academy of Sciences of the United States of America, *116*(13), 6407–6414. https://doi.org/10.1073/pnas.1819993116

Tompary, A., & Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, *96*(1), 228–241.

Tompary, A., & Thompson-Schill, S. L. (2021). Semantic influences on episodic memory distortions. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001017

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–20. https://doi.org/10.1109/TMI.2010.2046908

van der Linden, M., Berkers, R. M., Morris, R. G., & Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *Journal of Neuroscience*, *37*(39), 9474–9485. https://doi.org/10.1523/JNEUROSCI.3603-16.2017

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience*, *26*(10), 2250–2261. https://doi.org/10.1162/jocn{_}a{_}00630

van Kesteren, M. T., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359. https://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M. T., Rignanese, P., Gianferrara, P. G., Krabbendam, L., & Meeter, M. (2020). Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-61737-1

van Kesteren, M. T., Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience*, *30*(47), 15888–15894. https://www.jneurosci.org/content/30/47/15888

van Kesteren, M. T., Ruiter, D. J., Fernandez, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://www.sciencedirect.com/science/article/pii/S0166223612000197

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, e38677. https://doi.org/10.7554/eLife.38677

Varga, N., Morton, N., & Preston, A. (2022). Schema, inference, and memory.

Wan, Q., Menendez, J., & Postle, B. R. (2022). Priority-based transformations of stimulus representation in visual working memory. *Plos Computational Biology*, *18*(6), e1009062. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009062

Yu, Q., Teng, C., & Postle, B. (2020). Different states of priority recruit different neural representations in visual working memory. *PLoS Biol.*, *18*(6), e3000769. doi:%2010.1371/journal.pbio.3000769

Zacks, J., Speer, N., Swallow, K., Braver, T., & Reynolds, J. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. doi:%2010.1037/0033-2909.133.2.273

Zeithamova, D., & Preston, A. R. (2010). Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *The Journal of Neuroscience*, *30*(44), 14676–84. 10.1523/JNEUROSCI.3250-10.2010

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Appendix A

List of episodic details used for manual scoring of recall results:

- Jolene/Jeffrey
  - dated 3 years
  - met in Ireland
  - backpacking in Europe
  - Spain
  - parade with fire/bonfire
  - Black-Eyed Susan
  - plant native to Ireland
  - honor where they met
  - potato
  - food they ate on dates
  - two best friends (on bikes)
  - friends who introduced them
  - suitcase
  - to travel back to Ireland
  - bear decorative plaque
  - smoky mountains
- Beth/Jim
  - Buddhist retreat
  - vegan meals
  - dated 6 months / engaged 6 months
  - bonfire on retreat
  - hold hands
  - lotus
  - enlightenment symbol in buddhist tradition
  - hot air balloon

  - proposal after buddhist retreat
  - Buddha painting
  - emulate Buddha
  - doll
  - toy boat
  - daughter
  - son
- Bob/Jennifer
  - dance partners
  - winning Latin Dance Cup
  - dancing the tango
  - flower in hair when dancing
  - backyard pond
  - dance competitions
  - Egyptian statue
  - decorative their house
  - maracas
  - book
  - live music
  - history of dance
  - pet ducks
- Steph/Harold
  - cruise-ship
  - crew members / workers
  - mediterranean
  - singer
  - proposal in Hawaii
  - Parrott
  - tropical bird seen on cruises
  - lighthouse
  - loyal greeting when coming home
  - casino
  - met in bar on ship
  - cards
  - tv
  - 4 years of dating
  - play games together
  - Hawaiian flower
  - bonfires on the beach
  - leasure time when not on ship
- Hannah/Bill
  - Christmas Party

  - road trips
  - anglophiles / love england
  - cupcake
  - High Tea
  - favorite dessert at High Tea
  - Oxford
  - pearl necklace
  - college bonfire (party)
  - London honeymoon
  - rose
  - cooking pot
  - National Flower
  - cooking together
  - vintage car
- Camille/Felix
  - backpacking South America
  - changed trip to backpack together
  - hostel
  - kitchen
  - left South-America engaged
  - partying with locals around campfire
  - staying in small village
  - Freesia
  - Brazil proposal
  - owl
  - Argentinian owl
  - flower
  - medicinal
  - food poisoning / sick in Chile / south-america
  - skirt
  - salsa dancing
  - wheelbarrow
  - gardening
- Doreen/Simon
  - thought love was not for them
  - plant outside daycare
  - lunch
  - daycare
  - insurance company
  - work across street from each other
  - spilled drink
  - dinner to make up
  - dated 4 years

  - company bonfire
  - first said I love you
  - poppy
  - simon liked these flowers
  - walrus cartoon
  - favorite childhood book
  - read to kids in daycare
  - hope to read to own children
  - field with mountain
  - Thailand - proposal
  - rocking horse
  - target/toy gun
  - use at daycare or for own children
- Magdalene/Chris
  - writers
  - flower
  - fiction
  - journalist
  - coffee shop in california
  - slipped a note asking for date
  - dated 3 years/ engaged 1 yr
  - throwing discarded writings in fireplace
  - daisy
  - table decorations
  - poems
  - dunes/desert
  - inspiration for novel
  - nomadic people
  - bracelet
  - interviewing people
  - shades
  - writing outside easily
- Shannon/Tobias
  - law
  - mutual friend's dinner party
  - common interests
  - ask host for number
  - 1 yr dating / 1 yr engaged
  - nights at fireplace reading
  - black-eyed susan
  - represents justice
  - storm
  - favorite beverage

  - overpass
  - hiking arizona
  - sabbaticals
  - bottle glue
  - arts/crafts
  - birdcage
  - fun pet to combine with busy schedule
- Vera/Solomon
  - bartender
  - go to bar with friends
  - witty joke
  - church
  - lit hair on fire
  - make drink set on fire
  - rose
  - dozen roses to show how serious
  - monster octopus
  - valentine's day gift / stuffed animal
  - still has the gift
  - visiting at work at bar
  - parrots heart
  - tennis racket
  - play after return from honeymoon
  - toaster
  - use in new house
- Dennis/Vivian
  - met online
  - movie theater
  - movie reviewers blog
  - favorite movie brokeback mountain
  - rose
  - beauty and the beast
  - enchanted rose
  - cartoon one
  - logo
  - rating number 1 movie each week
  - daisies
  - red carpet film premiere
  - photographs with ring
  - umbrella
  - mary poppins
  - vase
  - decorate tv room

- Angela/Dave
  - dog-walkers
  - walk together
  - dated 2 yrs
  - house caught on fire
  - call 911
  - play on ranch
  - bonding experience
  - irises growing in park
  - friend Jud
  - ranch
  - riding
  - drum
  - iris
  - robe
  - use in new house

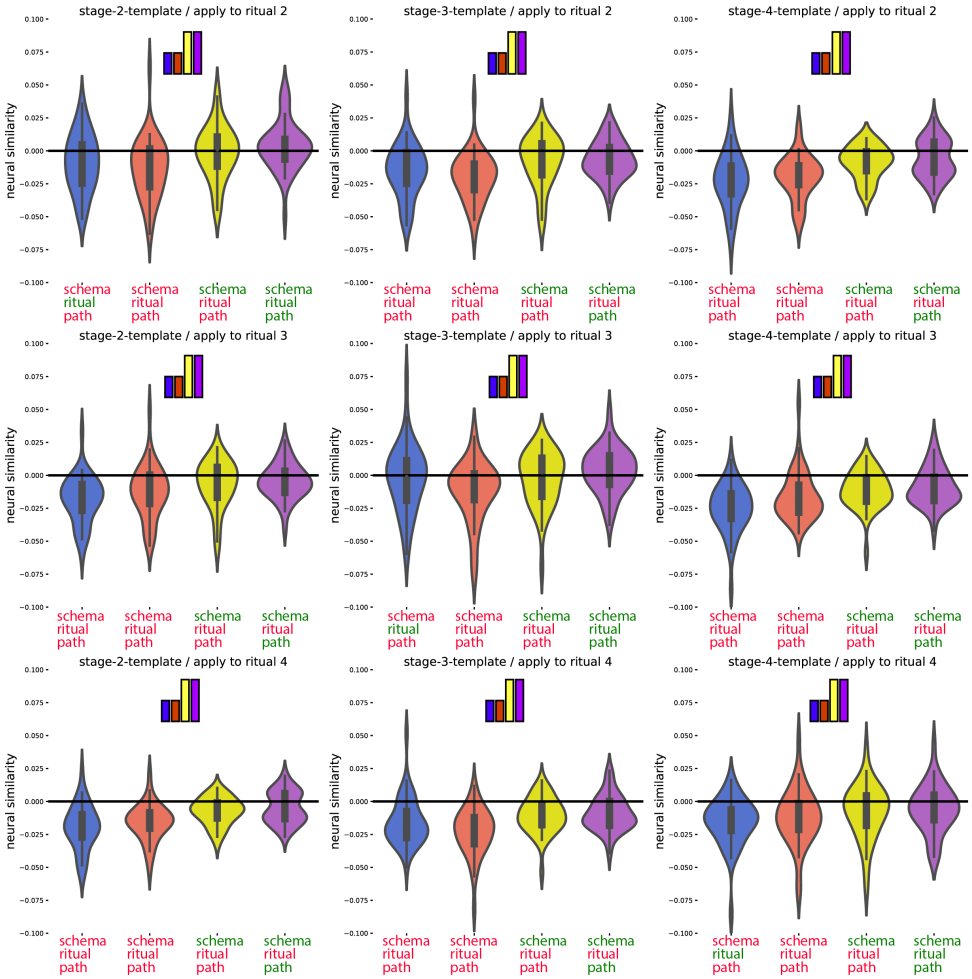

Fig. A1 Posthoc confirmation of the pattern for the schema neural code (Figure 5 shows the same results, separately averaging across all within-stage and across-stage comparisons). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4. The small bars at the top of each little plot indicate the predicted pattern.

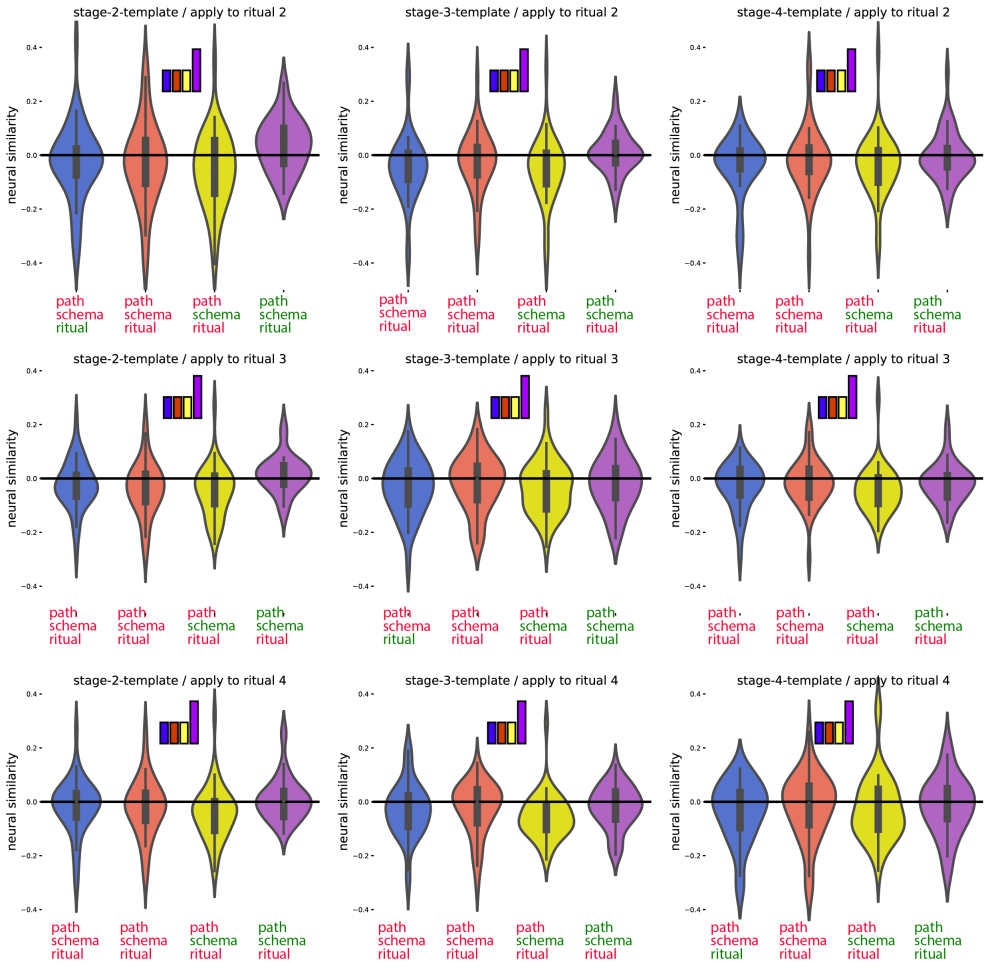

Fig. A2 Posthoc confirmation of the pattern for the path neural code (Figure 6 shows the same results, separately averaging across all within-stage and across-stage comparisons). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4. The small bars at the top of each little plot indicate the predicted pattern.

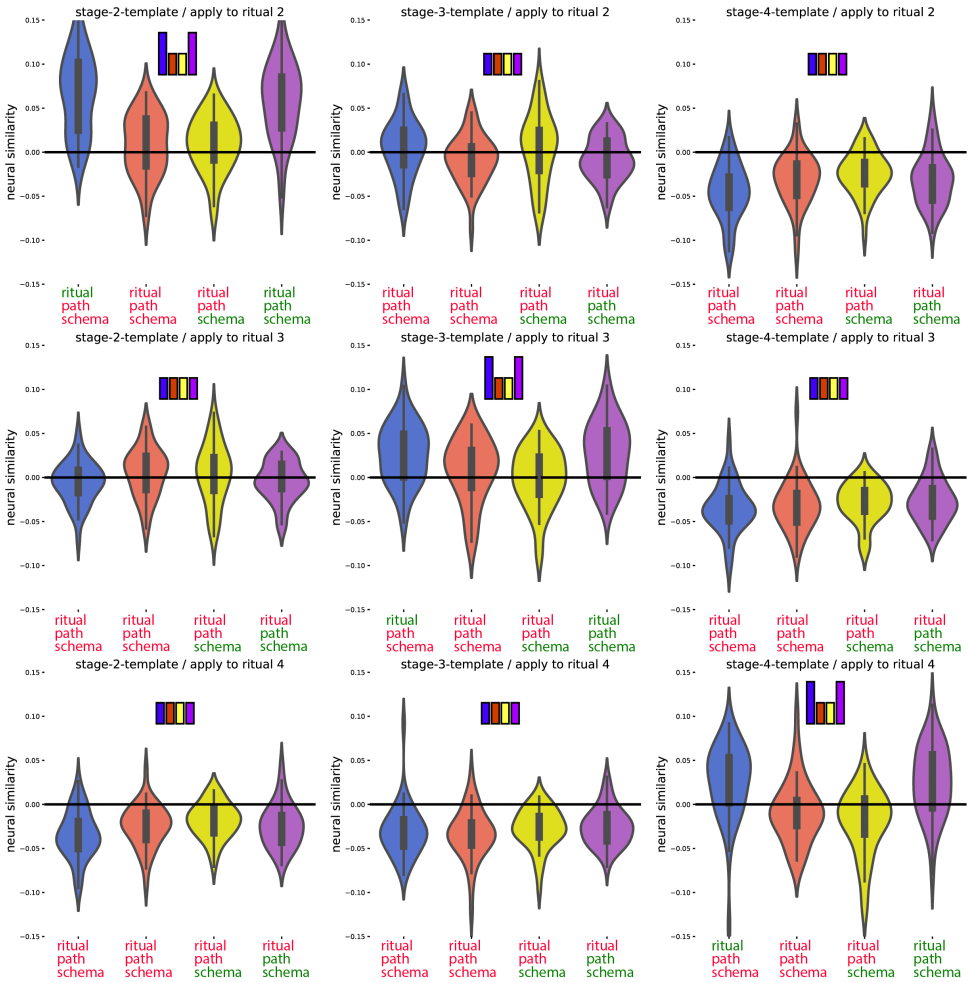

Fig. A3 Posthoc confirmation of the pattern for the current ritual neural code (Figure 8 shows the same results, separately averaging across all within-stage and across-stage comparisons). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4. The small bars at the top of each little plot indicate the predicted pattern.

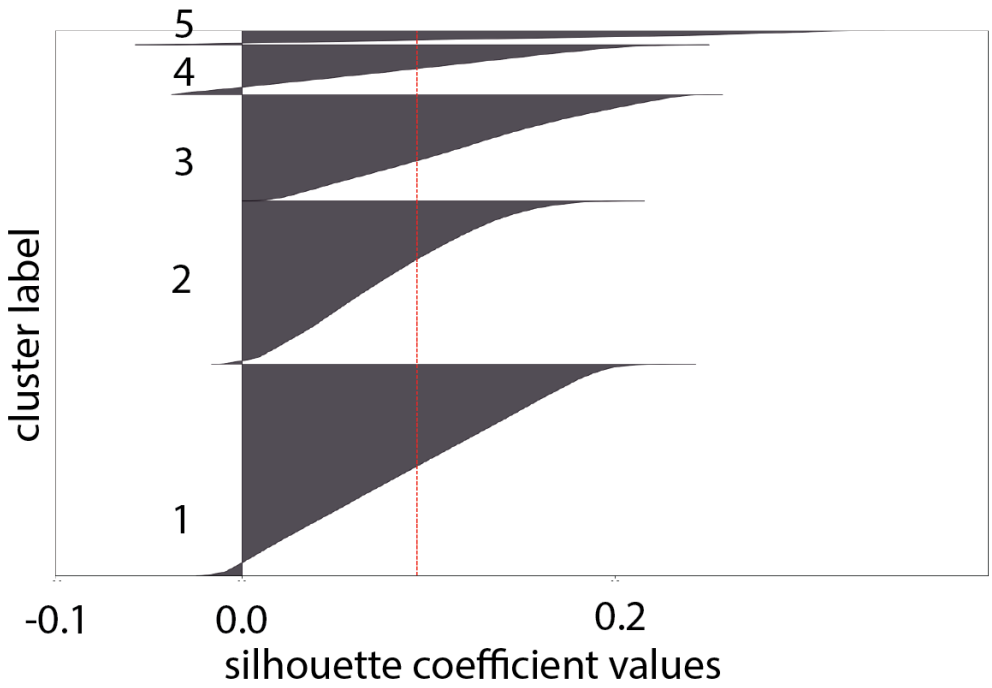

Fig. A4 Silhouette scores for K-means analysis (K = 5). The red line indicates average silhouette score.

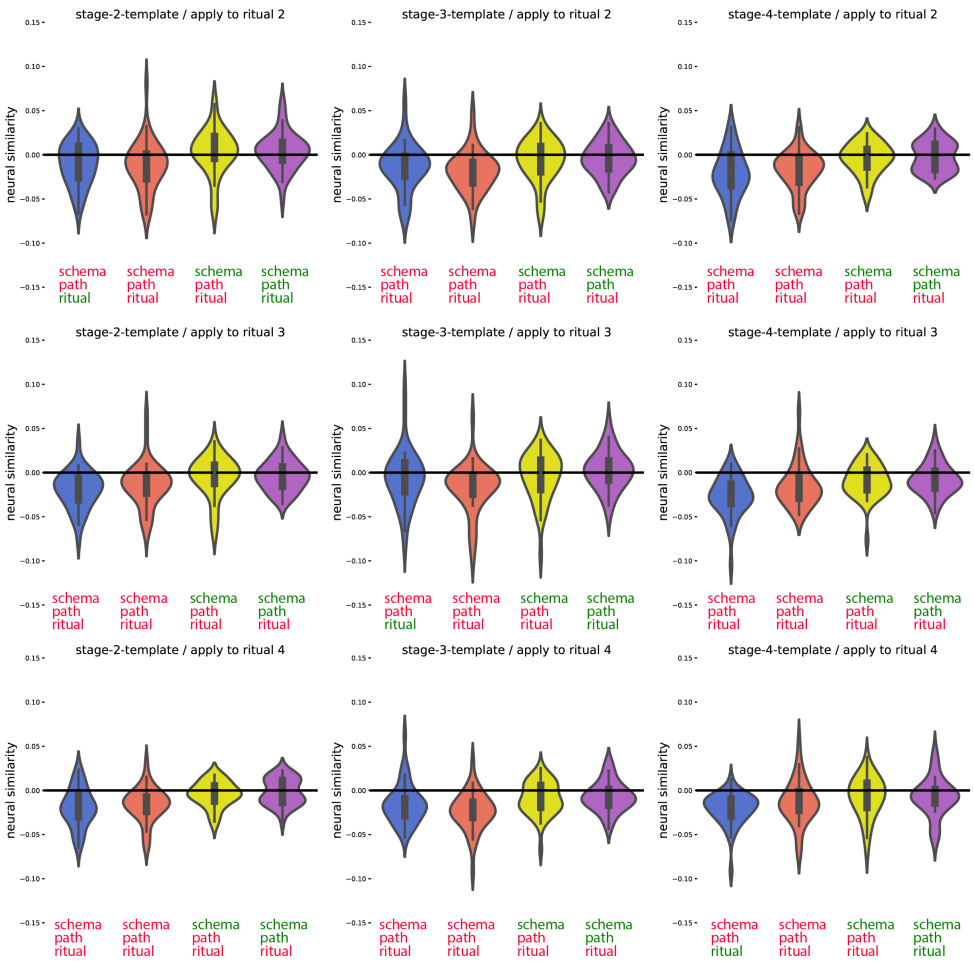

Fig. A5 K = 5, cluster 1 (putative schema code). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4.

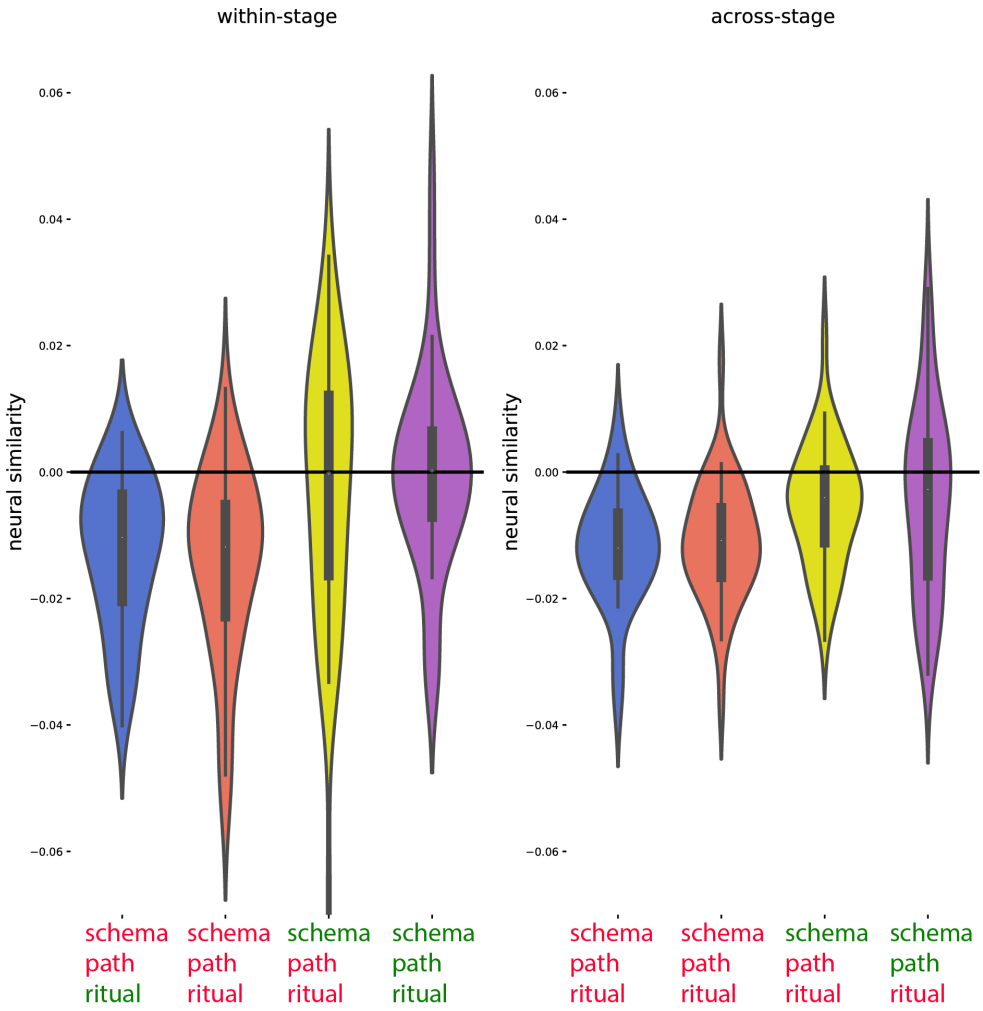

Fig. A6 K = 5, cluster 1 (putative schema code), separately analyzing within-stage and across-stage comparisons. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. The schema code pattern was present for both within-stage and across-stage comparisons, with higher similarity for same-schema conditions (i.e., yellow and purple) than different-schema conditions (i.e., blue and orange), as shown using t-tests (within-stage: $t(39) = 4.158$, $p < 0.001$; across-stage: $t(39) = 3.959$, $p < 0.001$). While these regions showed higher similarity for same vs. different schemas, they did not show sensitivity to same vs. different path, as evidenced by non-significant within-stage and across-stage comparisons for same-path, same-schema (purple) vs different-path, same-schema (yellow); within-stage: $t(39) = 0.565$, $p = 0.575$; across-stage: $t(39) = 0.409$, $p = 0.685$.

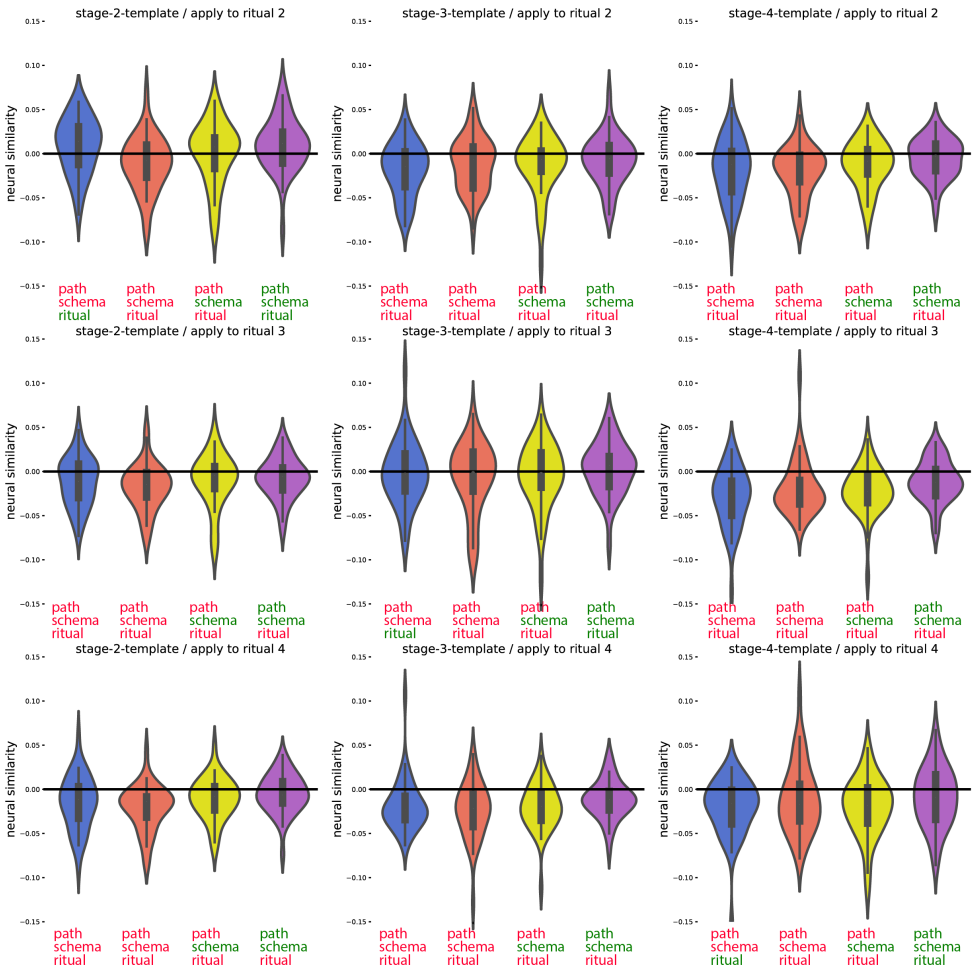

Fig. A7 K = 5, cluster 2 (putative path code). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4.

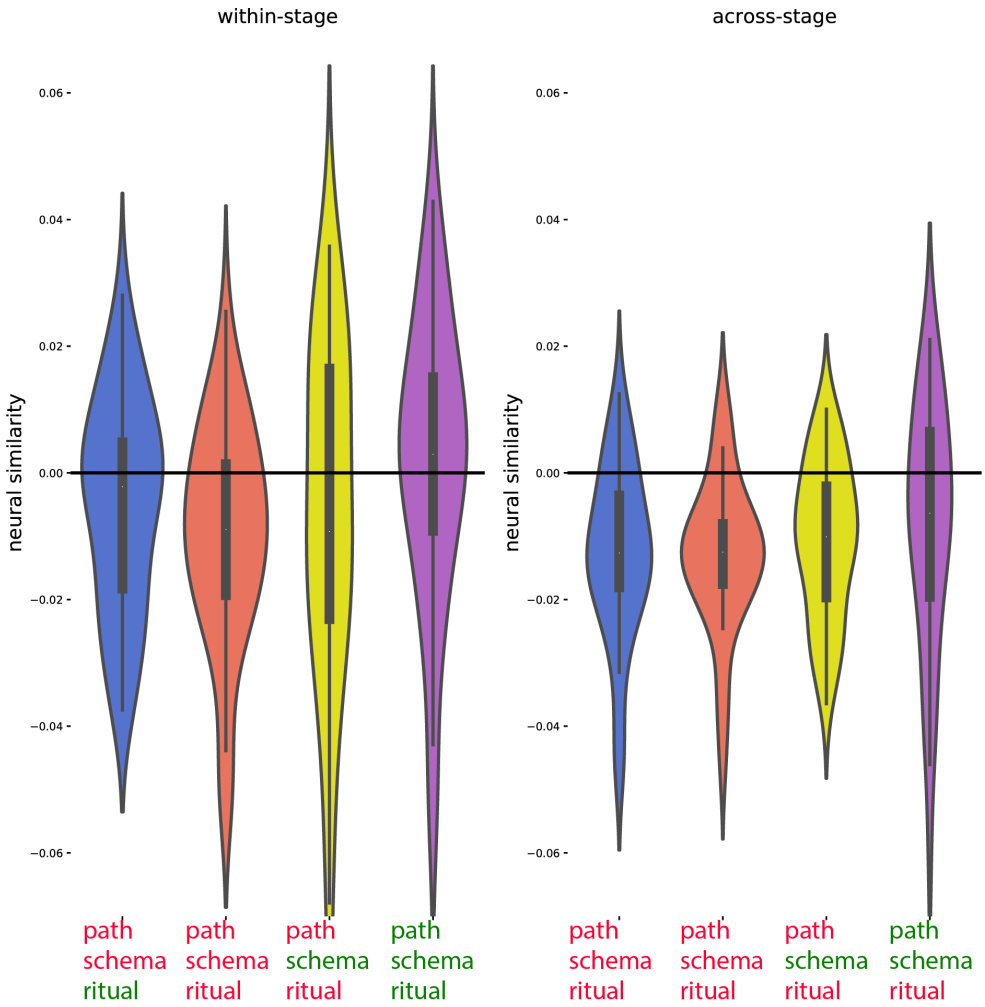

Fig. A8 K = 5, cluster 2 (putative path code), separately analyzing within-stage and across-stage comparisons. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. Statistics for comparing same-path condition (i.e., purple) with different-path condition (i.e., yellow) using t-tests (within-stage: $t(39) = 1.604$; $p = 0.117$; across-stage: $t(39) = 0.868$; $p = 0.391$).

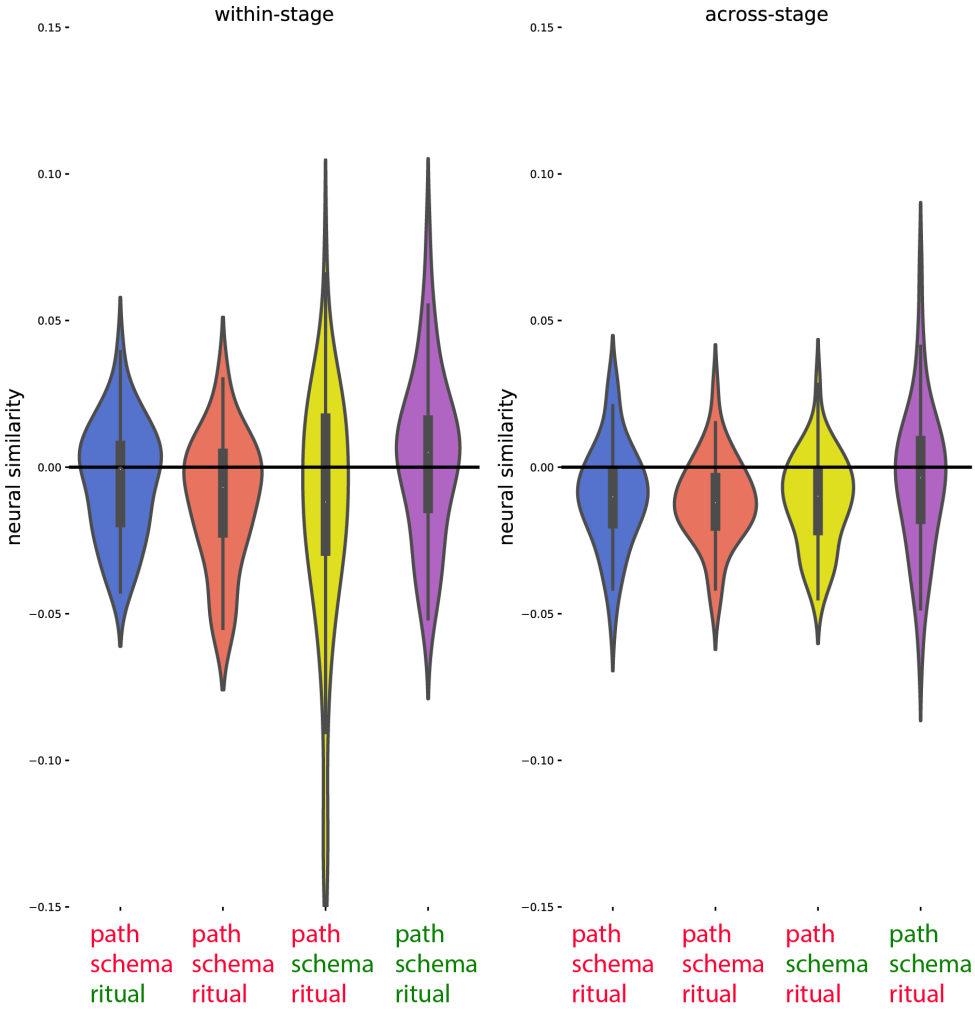

Fig. A9 K = 5, cluster 2 (putative path code), separately analyzing within-stage and across-stage comparisons, filtered to only include the top 25 percent of searchlights within the cluster, ranked by silhouette score. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. Statistics for comparing same-path condition (i.e., purple) with different-path condition (i.e., yellow) using t-tests (within-stage: $t(39) = 2.044$; $p = 0.048$; across-stage: $t(39) = 1.618$; $p = 0.114$).

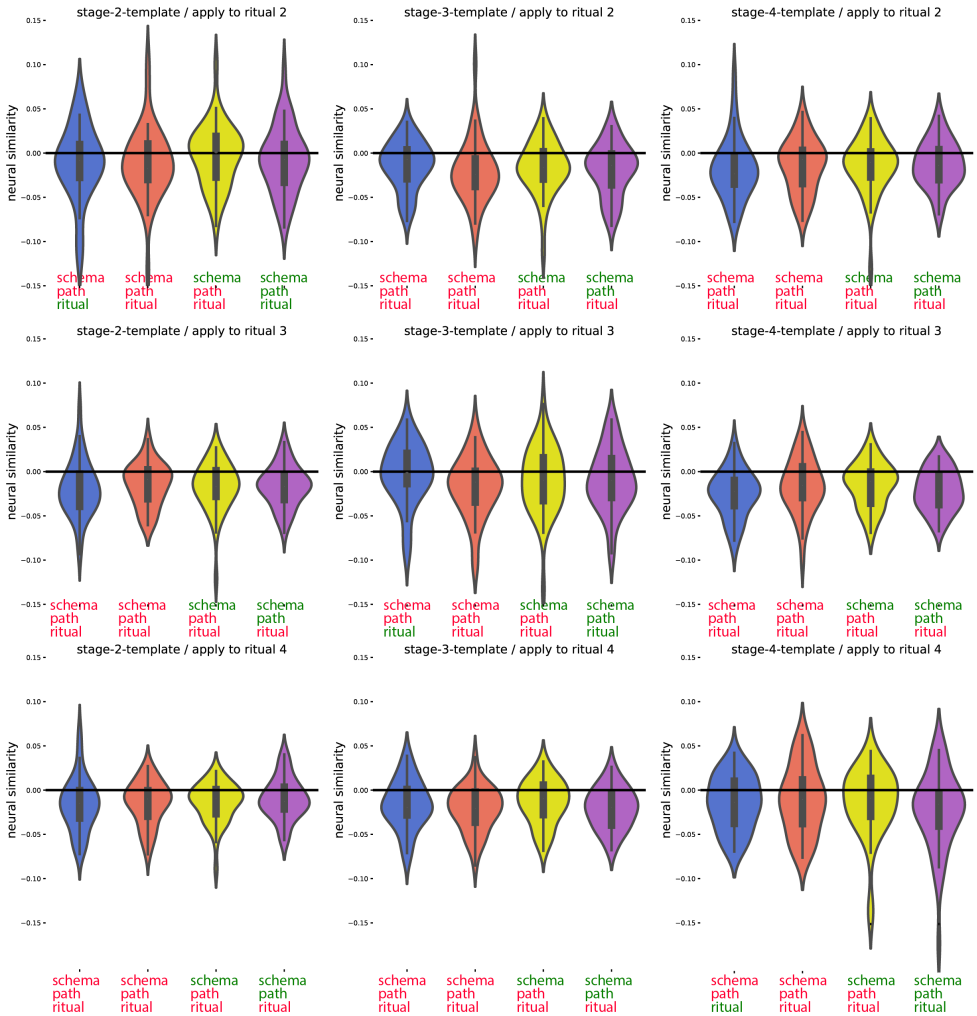

Fig. A10 K = 5, cluster 3 (putative null cluster). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4.

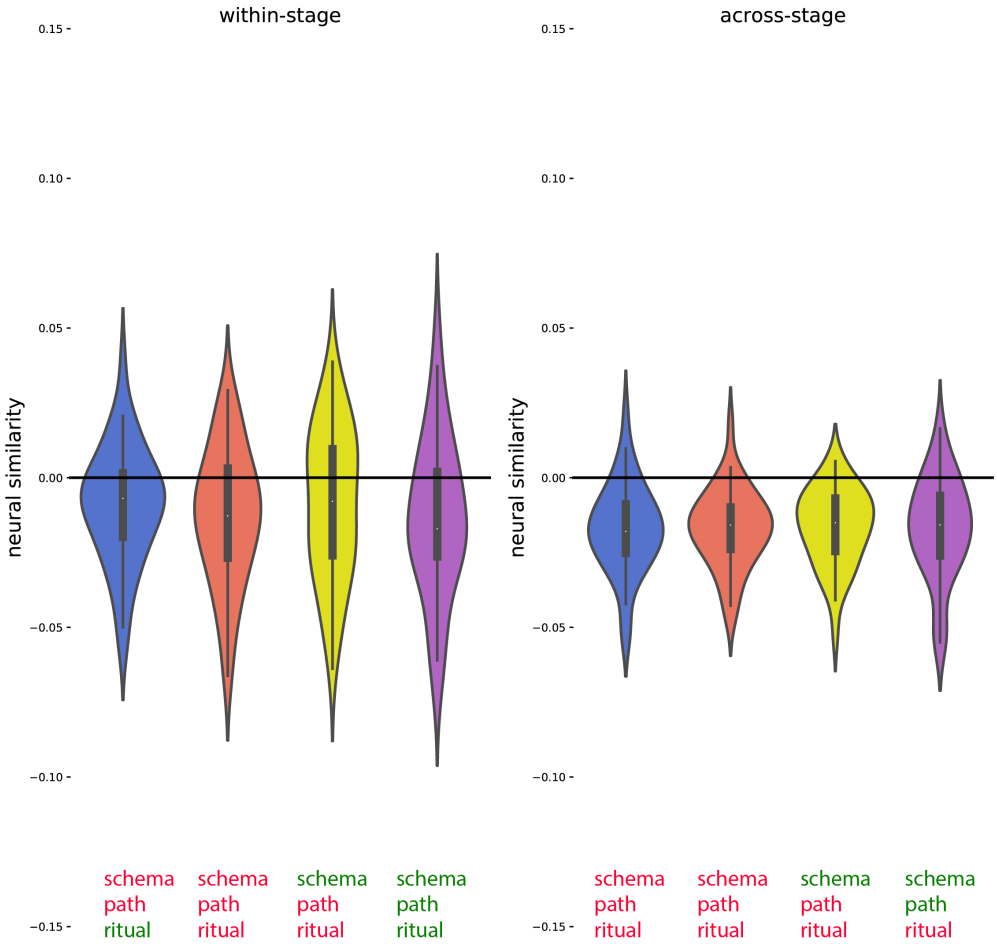

Fig. A11 K = 5, cluster 3 (putative null cluster), separately analyzing within-stage and across-stage comparisons. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. All relevant comparisons are indeed not significant. Within-stage: for same-schema vs. different-schema, t(39) = 0.034, p = 0.973; for same-path vs. different-path, t(39) = - 0.796, p = 0.431. Across-stage: for same-schema vs. different-schema, t(39) = 0.200, p = 0.843; for same-path vs. diferent-path, t(39) = - 0.054, p = 0.957.

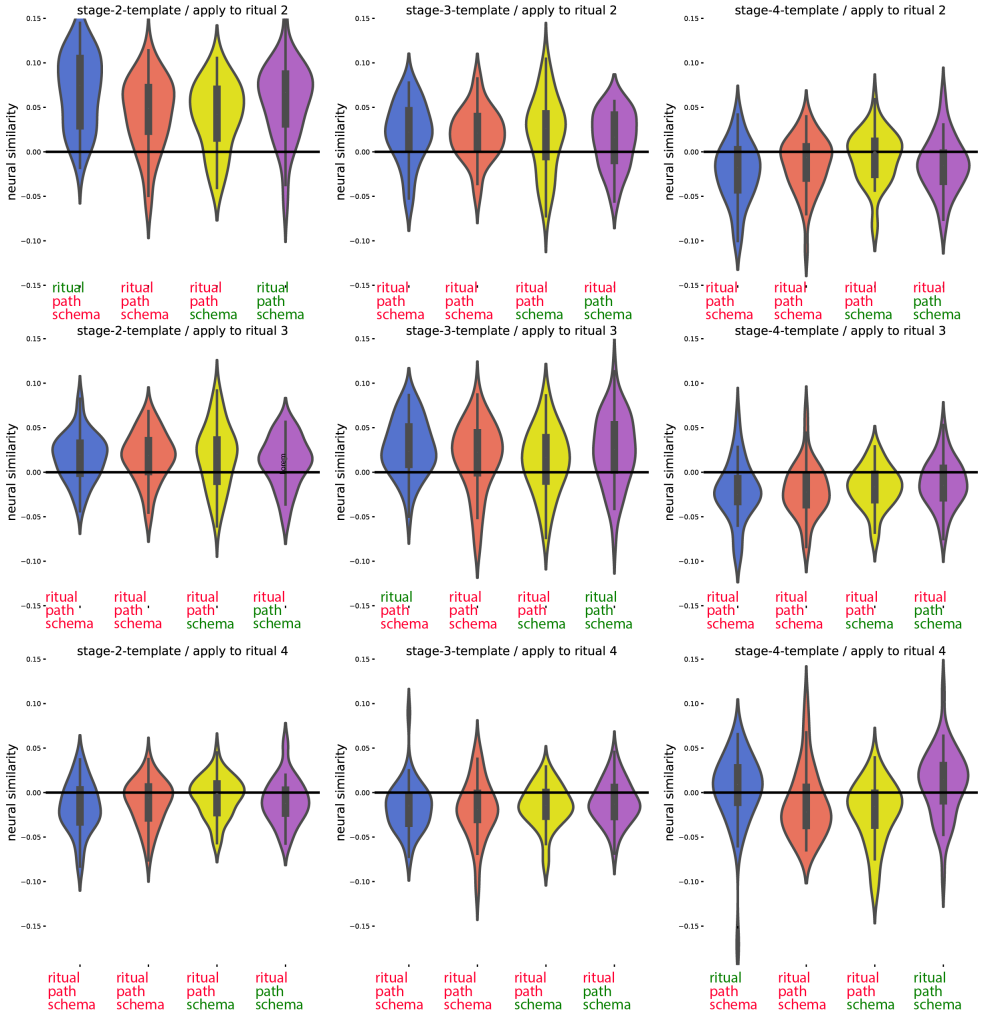

Fig. A12 K = 5, cluster 4 (putative current ritual code). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4.

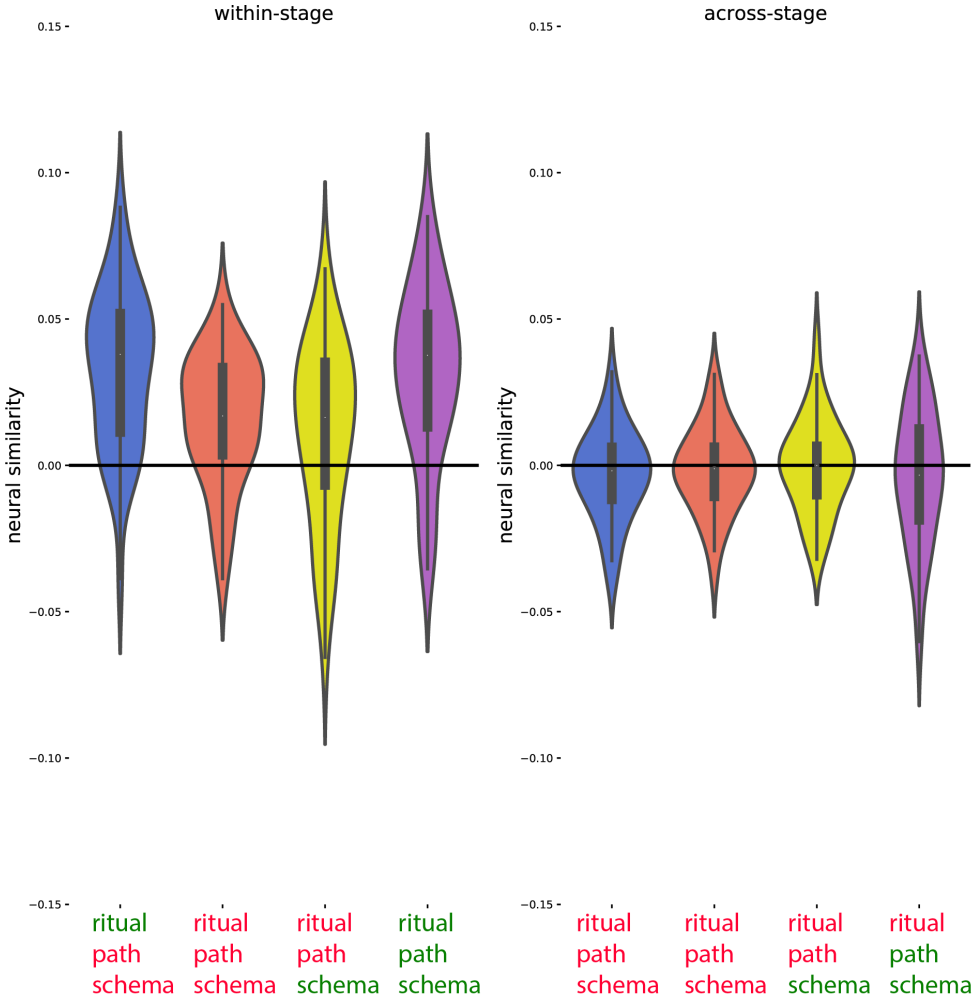

Fig. A13 K = 5, cluster 4 (putative current ritual code), separately analyzing within-stage and across-stage comparisons. Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. [left] An average of all within-stage comparisons. [right] An average of all across-stage comparisons. This cluster shows a current ritual code pattern for the within-stage comparisons: Pattern simliarity is higher for the two same-ritual conditions (i.e., blue and purple) than the two different-ritual conditions (i.e., yellow and orange), $t(39) = 5.318$, $p < 0.001$. As expected, this effect is not present for the across-stage comparisons, $t(39) = -1.419$, $p = 0.164$, because the current ritual always differs for across-stage comparisons.

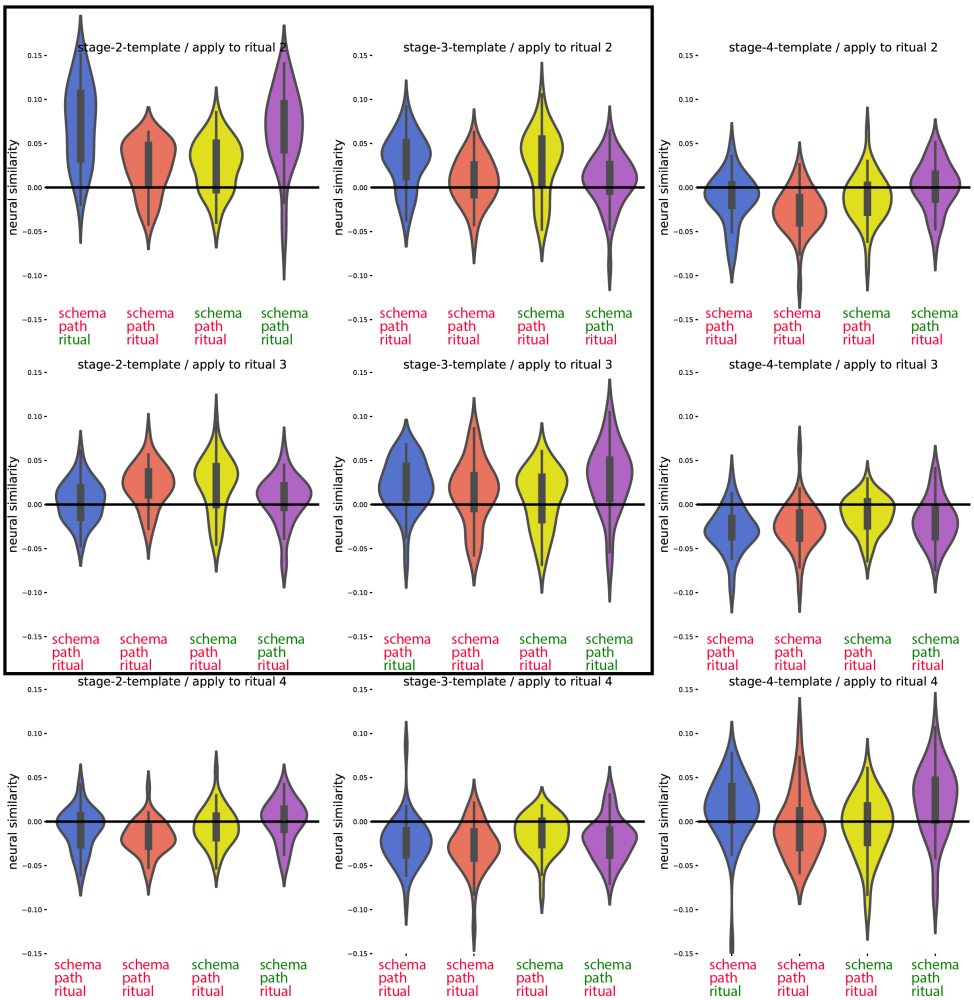

Fig. A14 K = 5, cluster 5 (putative rotated preceding ritual code). Colors of violin plots indicate the relationship between the paths of the template and the held-out ritual: Purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons (but not across-stage comparisons). For convenience, colors of x axis labels indicate whether the schema/ritual/path are the same (green) or different (red) for the template and held-out ritual. The 9 sub-figures correspond to different combinations of using stage 2/3/4 as the template and applying the template to stage 2/3/4. This shows a rotated code pattern, as evidenced by the following t-tests across the four subplots in the black box: First, in the upper-right plot in the black box, there was lower similarity for orange and purple (held-out ritual immediately preceded the template ritual) than blue and yellow, $t(39) = 4.352$; $p < .001$. In the lower-left plot in the black box, there was lower similarity for blue and purple (held-out ritual immediately followed the template ritual) than orange and yellow, $t(39) = 4.091$; $p < .001$. The upper-left plot in the black box provides evidence that these regions also coded for the current ritual: Similarity was higher for same-ritual (i.e., blue and purple) than across-ritual (i.e., orange and yellow) comparisons, $t(39) = 8.016$; $p < .001$.

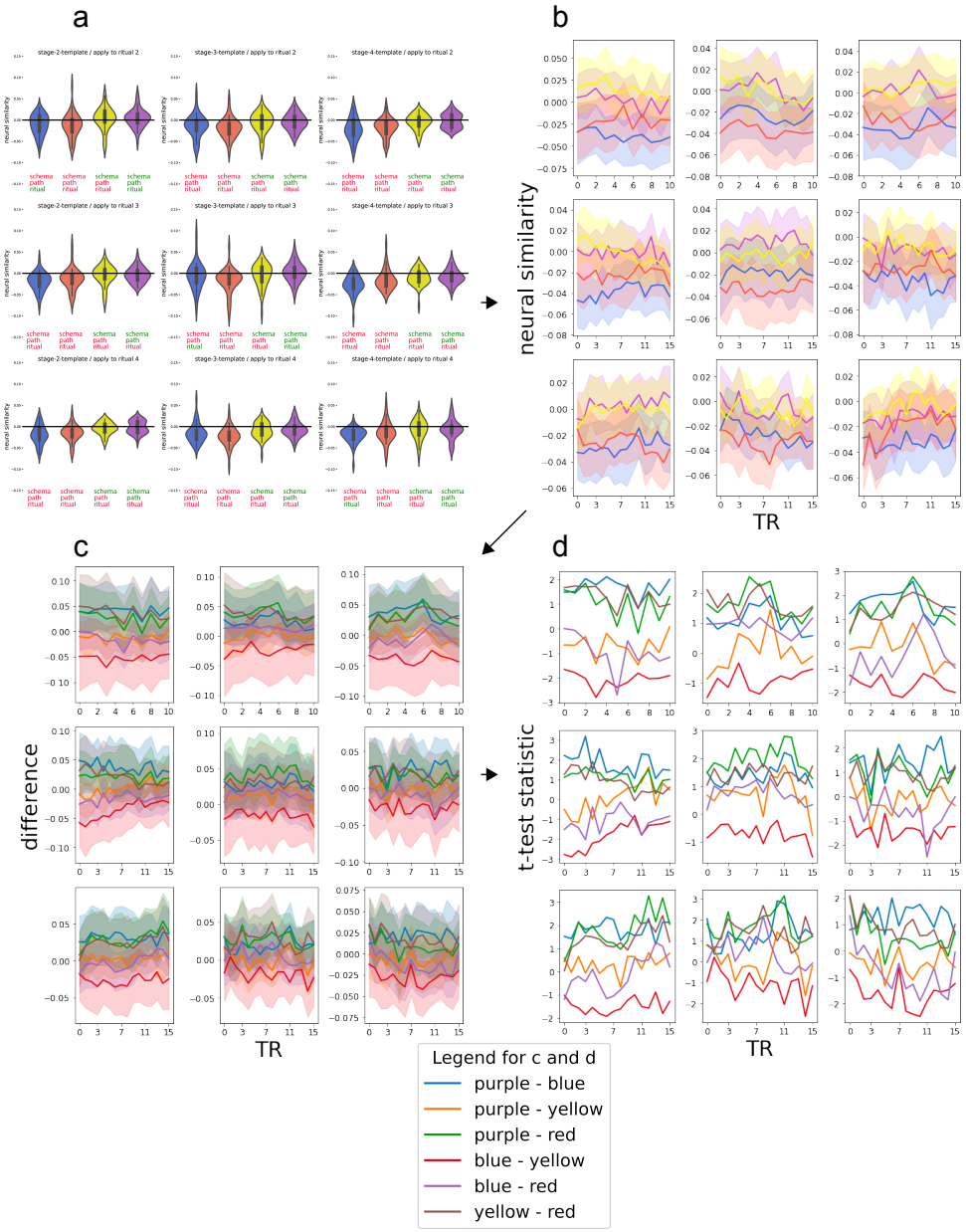

Fig. A15 Procedure for computing the per-searchlight "similarity fingerprints" that were used in the K-Means clustering analysis. For each searchlight, we took the following steps. In Step 1, we computed the full set of representational similarity measures described in the *Representational similarity analysis* section, grouping them according to the stage where the template was acquired (2, 3, 4), the stage of the held-out ritual (2, 3, 4), and the relationship between the path of the template ritual and the path of the held-out ritual (purple = same path; yellow = different path, same schema; orange = different schema, always different ritual; blue = different schema, same ritual for within-stage comparisons, but not between-stage comparisons). This yields a set of 36 (3 x 3 x 4) similarity scores per participant – akin to the results shown in Appendix Figure A1 and other, similar figures in the paper (Panel A). A key difference between our approach here and the approach taken in Appendix Figure A1 is that, instead of averaging over all of the TRs in a ritual and then computing similarity (yielding 36 violin plots, as in Step 1), here we skipped the averaging step and computed similarity at each individual TR within the rituals, yielding 36 lines per participant (where the dimension of each line is the number of TRs per ritual; i.e., 11 TRs for campfire and flower rituals, 16 TRs for coin and torch rituals, and 16 TRs for egg and painting rituals). This is shown in Panel B – here, the ribbons around the lines are meant to illustrate the variability across participants for each of these lines. Next, within each of the 9 conditions defined by crossing of the template stage and the held-out ritual stage (i.e., within each of the 9 sub-plots shown in Panel B), we computed all 6 of the possible *pairwise differences* between the 4 lines (purple, yellow, orange, blue) in that sub-plot. This leaves us with 54 lines per participant (6 pairwise-difference lines x 9 combinations of template stage and held-out-ritual stage). This is shown in Panel C – here also, the ribbons around the lines illustrate variability across participants for each of these lines). The final stage of the fingerprinting process is to convert these 54 difference-score lines per participant into a single set of 54 lines (combining across participants) by computing, at each TR and for each of the 54 difference-scores, a t-test assessing the reliability of that difference score (at that TR) across participants. The product of this process, therefore, is a matrix of 54 x the number of TRs t-scores for each searchlight (Panel D) – we then flatten this matrix into a vector of length 54 x the number of TRs, which constitutes the "similarity fingerprint" for the searchlight that is fed into the K-means clustering algorithm.

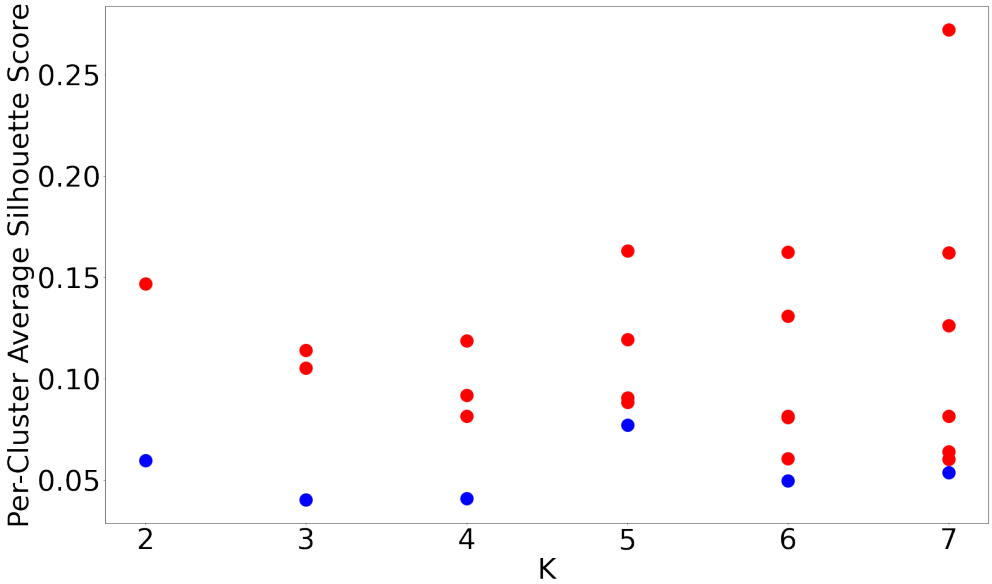

Fig. A16 Per-cluster average silhouette scores (computed by averaging these scores across all searchlights in the cluster) from K-Means clustering algorithm for K in range [2,7]. Each dot represents the per-cluster average silhouette score average for one of the clusters found by the K-Means clustering algorithm for a particular K. The clusters with the worst average silhouette score for each K are marked in blue. The worst per-cluster average silhouette score for K = 5 is better than the worst per-cluster average silhouette scores for all other K values.