Neural Trajectories of Conceptually Related Events

Matthew Schafer[1], Philip Kamilar-Britt[2], Vyoma Sahani[2], Keren Bachi[Y,2.3] & Daniela Schiller[Y1.,2,4]

[1]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai; New York City, NY
[2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai; New York City, NY
[3]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai; New York City, NY
[4]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai; New York City, NY

[Y]These authors contributed equally
*Corresponding author email: daniela.schiller@mssm.edu

Abstract

In a series of conceptually related episodes, meaning arises from the link between these events rather than from each event individually. How does the brain keep track of conceptually related sequences of events (i.e., conceptual trajectories)? In a particular kind of conceptual trajectory—a social relationship—meaning arises from a specific sequence of interactions. To test whether such abstract sequences are neurally tracked, we had participants complete a naturalistic narrative-based social interaction game, during functional magnetic resonance imaging. We modeled the simulated relationships as trajectories through an abstract affiliation and power space. In two independent samples, we found evidence of individual social relationships being tracked with unique sequences of hippocampal states. The neural states corresponded to the accumulated trial-to-trial affiliation and power relations between the participant and each character, such that each relationship's history was captured by its own neural trajectory. Each relationship had its own sequence of states, and all relationships were embedded within the same manifold. As such, we show that the hippocampus represents social relationships with ordered sequences of low-dimensional neural patterns. The number of distinct clusters of states on this manifold is also related to social function, as measured by the size of real-world social networks. These results suggest that our evolving relationships with others are represented in trajectory-like neural patterns.

Introduction

When you are in certain place, say a hotel lobby, the meaning of that place (i.e., your location in state space, or latent state) only makes sense in the context of what led you there. A sequence of observations is necessary to disambiguate the current state, as a single observation is insufficient: checking in and checking out of a hotel may look identical yet they predict very different futures. This also applies to conceptual spaces, where sensory information is fully abstracted out. A prototypical case of conceptual space is social space, where the state of a social relationship evolves over interactions along the latent dimensions of power and affiliation[1]. For example, a conversation between two people in that hotel lobby can mean different things if they are old friends, lifelong enemies, or strangers. The prior relationship state—itself a summary of conceptually connected events from the relationship history—helps determine both the current and future relationship states. But while neural representations of latent states and paths in physical space have been identified[2], as have location-like representations in conceptual spaces, including social[3], it is unknown whether there is a neural representation of trajectory in conceptual space.

Where and how could relationship trajectories through social space be represented in the human brain? The hippocampus is a likely candidate. Functional magnetic resonance imaging (fMRI) studies have shown that the hippocampus tracks latent states across domains, from physical locations[4] to concepts[5,6] to abstract relations between self and other[3,7,8]. Sequences of neural states in the hippocampus may also be connected like trajectories. Evidence consistent with this idea showed that post-task fMRI activity patterns reactivated sequentially[9], reflecting previously learned sequences. An intriguing possibility is that the hippocampus maps sequences of social interactions in relationships onto a neural manifold, tracking them like trajectories in an abstract social space.

To examine this, we used a mix of representational similarity and manifold analyses to test whether hippocampal fMRI patterns show structures that are specific to different relationships' trajectories, but that share the same representational geometry. Participants completed a naturalistic narrative-based social interaction task[3], where they interact and form relationships with fictional people in a novel social network. The interactions are defined by choices with latent affiliation and power dynamics, and the relationships with the characters evolve on these dimensions as the task progresses (see **Figure 1**). We tested the following hypotheses about hippocampal representations (especially in the left hemisphere[3,10]): First, we expected the hippocampus to represent the social interactions on the affiliation (e.g., cooperation) and power (e.g., hierarchy) dimensions, a compressed format that allows generalization across social situations. Across the different characters and various interactions, hippocampal patterns should be more similar for interactions of the same dimension than interactions of different dimensions. But while these dimensions are general, the individual relationships are specific: each relationship should have its own unique sequence of hippocampal patterns. Lastly, the properties of these neural representations should relate to real life social function: individuals with hippocampal manifolds with more unique states during the social interactions may have larger real-world social networks. Our results support these hypotheses and suggest the hippocampus tracks sequences of relationship-specific social interactions like trajectories on a social manifold. These results provide evidence for hippocampal linking of conceptually related events across time and contexts.

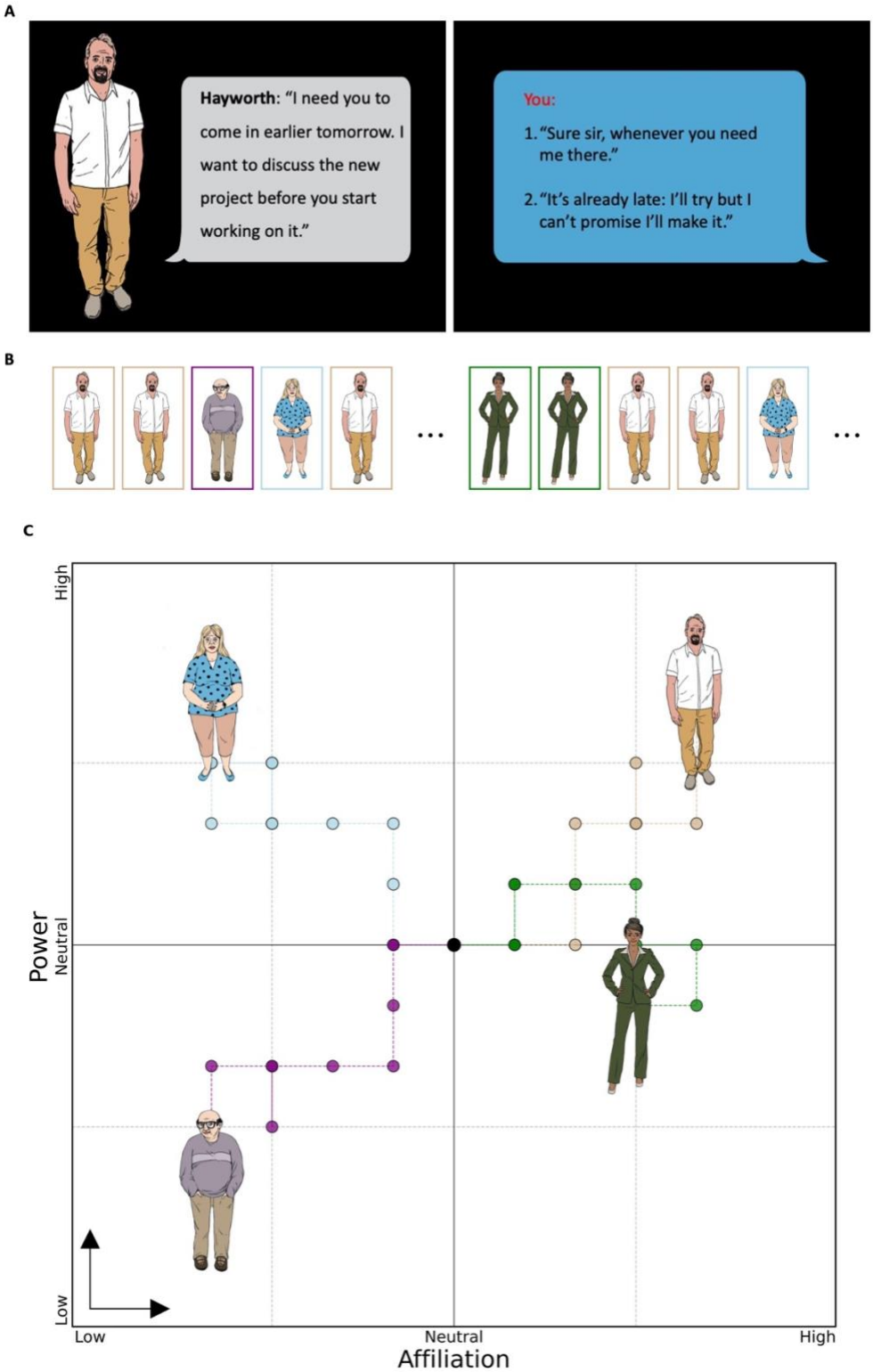


Figure 1. Social interaction sequences form relationship trajectories along abstract dimensions of affiliation and power. *(A) An example of a power interaction. Participants read text that describes the narrative and on decision trials choose between two options. Based on their choice, the character moves -1 or +1 along the active dimension (affiliation or power). (B) The participant forms relationships with different characters through sequences of interactions in the narrative. (C) The*

decisions the participant makes in the affiliation and power interactions change the character's location in social space, forming a relationship trajectory. Note that in the task, participants interact with 6 characters: 5 each with 6 affiliation and 6 power trials and 1 with 3 neutral trials.

Results

Behavior is consistent with social mapping

Prior to analyses of the fMRI data, we tested two behavioral assumptions: independent behavior between participants and a map-like organization to the character relationships (**Figure 2**). Below, we discuss each of these.

Participants' behavior is idiosyncratic

Each participant's set of choices was unique (i.e., never identical between participants). Moreover, the behavioral trajectories approximately occupied the entire space across both samples. As such, participant behavior was idiosyncratic and unlikely to simply reflect task structure.

Behavioral geometry is consistent with participants' subjective representation of the social locations

If participants represent the characters as locations, there may be evidence of this in post-task subjective reports. To test this, after the task we had the Validation sample participants place the characters into different locations in a two-dimensional (2D) affiliation and power space based on their perception of their relationships with the characters. We calculated the participant-specific average distances between these subjective locations and the task behavioral locations (i.e., mapping error) and compared them to permutation generated distances. As expected, the mapping error was smaller than chance ($CI_{95\%}$ = [-0.95, -0.2], t_{31} = -3.19, left-tailed $p < 0.005$), suggesting the behavioral locations capture elements of a subjective map. We also found a predicted negative correlation between mapping error and task memory (r = -0.37, left-tailed $p < 0.05$), consistent with memory of the participants' relationships with the characters driving their subjective placements. Together these results suggest that affiliation and power locations reflect a subjectively accessible map of the relationships, supporting the internal validity of the task.

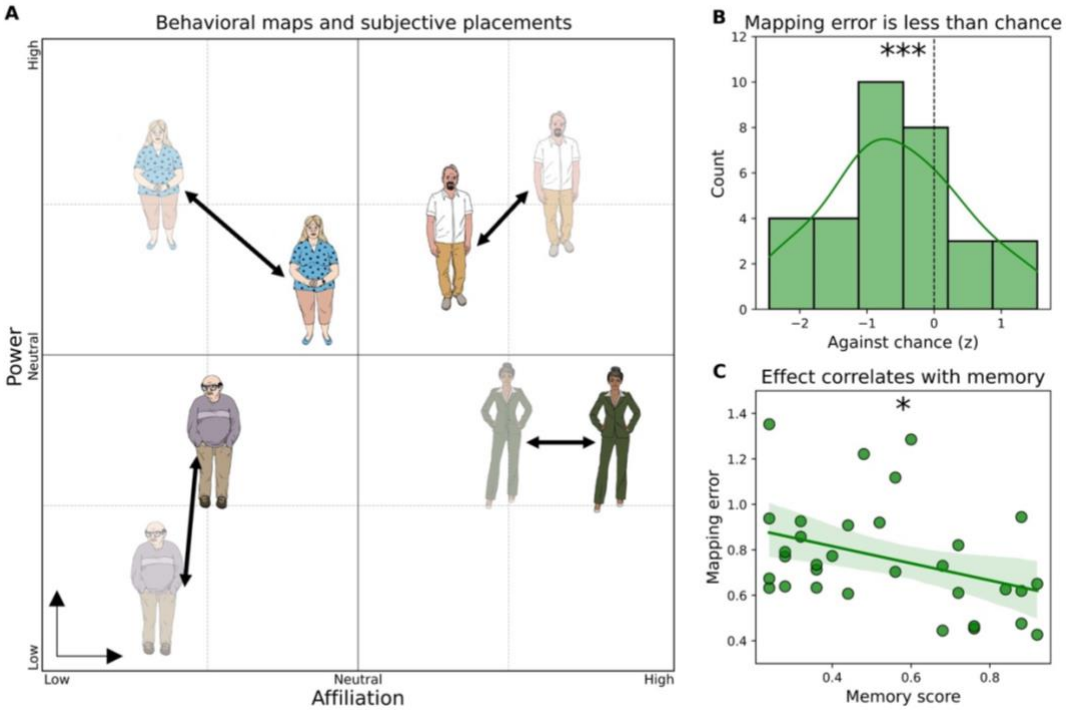


Figure 2. Behavioral geometry is consistent with social mapping. *(A) Schematic of post-task subjective placements as compared behavioral maps, the affiliation and power coordinates calculated from decisions (shown faded). The behavioral locations were not shown to participants during the placements. The mapping error between the behavioral and subjective placements was calculated as the average character-wise Euclidean distance between the locations (shown as arrows). (B) Mapping error is smaller than permutation-based chance, suggesting the behavioral modeling captures elements of these subjective placements. (C) The amount of mapping error negatively correlates with task memory, suggesting the subjective maps depend on memory. 95% confidence intervals for regression line are indicated by the shaded region and p-value significance is indicated by asterisks: * < 0.05, *** < 0.005. Validation sample only (n = 32).*

The hippocampus represents abstract dimensions of affiliation and power

After validating our main behavioral assumptions, we turned to the fMRI data. We hypothesized that the interactions are represented abstractly along the affiliation and power dimensions in the hippocampus, such that they generalize across characters and interactions. If so, decision trials of the same dimension should have more similar neural patterns than decision trials of different dimensions. We used a representational similarity analysis (RSA) searchlight to test this hypothesis, combining the two samples to increase statistical power (combined n = 50). Small volume correction in the left hippocampus showed significant effects in a cluster, with the peak voxel in the anterior (peak voxel MNI xyz = -35/-19/-15, cluster extent = 130 voxels), supporting our hypothesis that the hippocampus represents social situations abstractly along the dimensions of affiliation and power (see **Figure 3**).

A whole-brain analysis showed additional significant clusters, suggesting a distributed network represents these dimensions: the right posterior cingulate cortex (xyz = -6/-51/39, cluster extent = 19 voxels), the right and left angular gyrus (xyz = 44/-55/27, cluster extent = 1402 voxels; xyz = -54/-70/29, cluster extent = 1135 voxels) and the right and left medial temporal

gyrus (xyz = 51/-7/-34, cluster extent = 311 voxels; xyz = -63/-22/-9, cluster extent = 479 voxels).

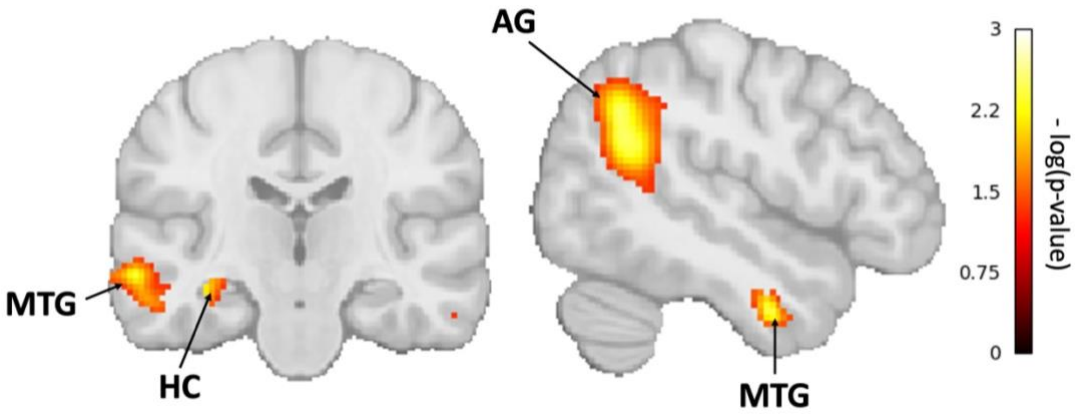


Figure 3. The left hippocampus represents affiliation and power interactions abstractly. *Dimensional abstraction: trials of the same dimension should have more similar patterns than trials of different dimensions. Representational similarity analysis searchlight results ($p_{FWER} < 0.05$, with small volume correction in left hippocampus). HC = hippocampus, AG = angular gyrus, MTG = middle temporal gyrus. Combined sample (n = 50).*

The hippocampus tracks relationship trajectories with neural trajectories

We have shown that the hippocampus represents the interaction decision trials along abstract social dimensions, but does it track each relationship's unique sequence of abstract social coordinates? If the social relationship trajectories are tracked by hippocampal trajectories, then the hippocampal pattern sequences should have trajectory-like properties (i.e., location-related, ordered, relationship history specific). Moreover, these neural trajectories should be low-dimensional, like the relationship trajectories.

To test these hypotheses, we used a novel neural trajectory analysis. First, we reduced the dimensionality of each participant's hippocampal beta series using the nonlinear algorithm Locally Linear Embedding[11] (LLE; see comparison to a linear method below). Then, for each character separately, we fit linear splines (piecewise line segments) through the sequence of low-dimensional neural patterns to approximate the hypothesized neural trajectories (see **Figure 4**). We used a leave one trial out approach: if the neural pattern sequence acts like a neural trajectory, we should be able to decode the social location of the held-out trial based on its location in the neural sequence. To this end, we parameterized the neural splines with their corresponding affiliation and power locations and then decoded the held-out trial's social location as the interpolated affiliation and power values from the closest locations on the fitted spline. The error was calculated as the Euclidean distance between the predicted and actual locations. We then assessed decoding performance against various null models by testing whether the real decoding produced smaller error distances than the null decoding, with left-tailed t-tests.

Social locations can be decoded from hippocampal sequences

If the hippocampus tracks the relationship trajectories, we first need to establish that we can decode social location (i.e., affiliation and power coordinates) from the hippocampal sequences. To test this, we compared the location decoding errors to permutation-based chance errors. As expected, social location decoding was significantly better in our model than in this chance model (Initial: $CI_{95\%}$ = [-0.31, -0.12], t_{17} = -4.8, left-tailed $p < 0.001$; Validation: $CI_{95\%}$ = [-0.29, -0.19], t_{31} = -10.25, left-tailed $p < 0.001$). We used circular shifting to generate the permutations while preserving the temporal autocorrelation structure in the neural sequence, thus ruling this out as an explanation for above chance performance.

Hippocampal sequences are ordered like trajectories

Trajectories are a set of temporally ordered locations through space. If the hippocampal patterns are ordered like trajectories through neural activity space, then a trial's position in the neural sequence should reflect its location in the relationship trajectory. To test this, we compared the model against a null model that only predicted the spline midpoint and thus assumed that the order of a trial in the neural sequence is unnecessary to decode its social location. As expected, the location decoding was better than in this null model (Initial: $CI_{95\%}$ = [-0.05, -0.01], t_{17} = -2.63, left-tailed $p < 0.01$; Validation: $CI_{95\%}$ = [-0.06, -0.02], t_{31} = -4.82, left-tailed $p < 0.001$), suggesting that the trial's location in the sequence is important to decoding and thus the hippocampal patterns are ordered like trajectories.

Hippocampal sequences track relationship-specific paths

A relationship trajectory is a specific path that a relationship takes through abstract social space. Thus, the relationship representation should reflect the specific sequence of interactions, more so than just the distribution of interaction choices and/or the end-of-task location of the relationship. To test this, we constructed a null model that preserved the choice distribution and the end location of each relationship, but let the paths vary by randomly shuffling the choices within each relationship before calculating the behavioral trajectories. Location decoding from the real paths was better than this null (Initial: $CI_{95\%}$ = [-0.05, -0.003], t_{17} = -1.82, left-tailed $p < 0.05$; Validation: $CI_{95\%}$ = [-0.04, -0.002], t_{31} = -2.36, left-tailed $p < 0.05$), suggesting the hippocampus tracks the ordered sequence of choices and the abstract locations they produce.

The neural trajectories are relationship-history specific

The hippocampal patterns should contain information about the relationship histories (as captured in the affiliation and power coordinates) above and beyond representations of task structure–such as character identity, familiarity, and affiliation and power dimensions. To examine this, we compared our decoding of real trajectories to decoding from simulated behavioral trajectories generated by randomly assigning choices. The real trajectory decoding errors were significantly smaller than these random choice errors (Initial: $CI_{95\%}$ = [-0.17, -0.01], t_{17} = -1.9, left-tailed $p < 0.05$; Validation: $CI_{95\%}$ = [-0.14, -0.03], t_{17} = -3.4, left-tailed $p < 0.001$). Thus, the hippocampal sequences contain information about the behavioral trajectories above and beyond the effects of character identity, familiarity, social dimension and other aspects of task structure.

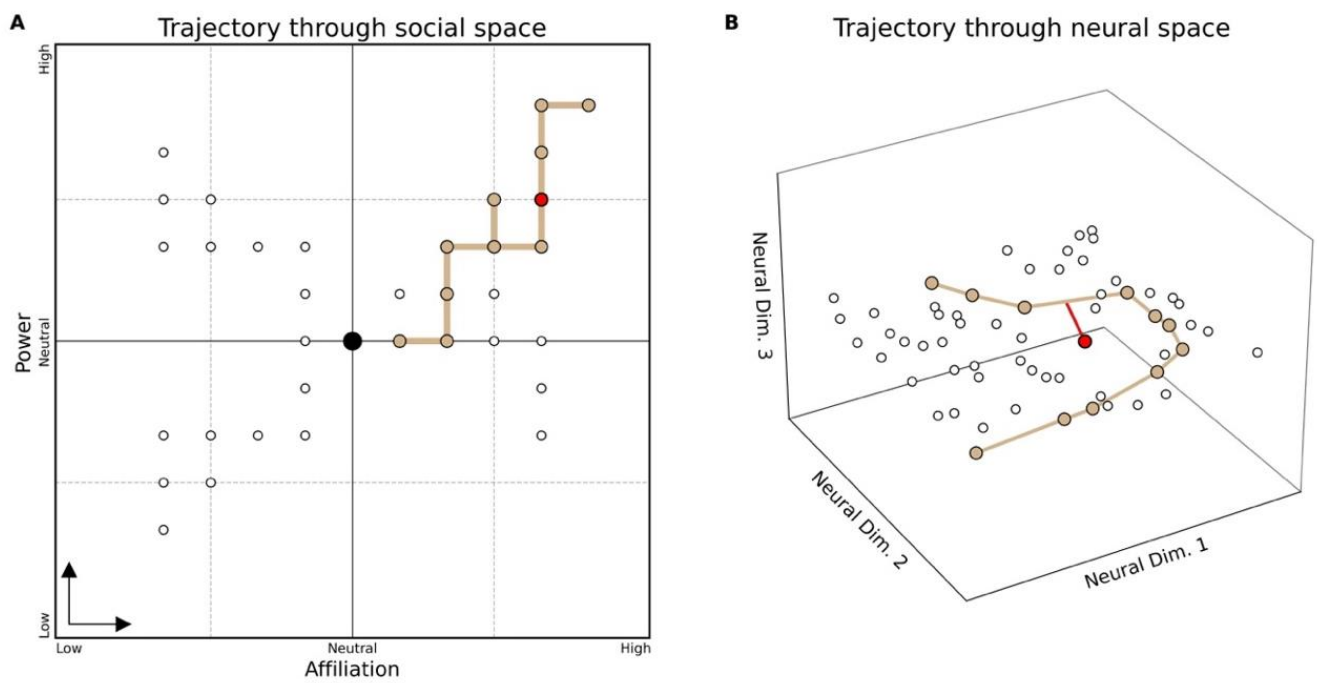
The neural trajectories reflect social locations, not just choices

We then tested whether the decoding effect depends on the social locations rather than solely on the interaction choices. To do this, we constructed pseudo-trajectories, where random sets of real choices from different characters were selected and cumulatively summed to create sequences of locations that were not actually experienced. This null model preserved the temporal order of the trials and social dimension and was based on participant choices but did not preserve relationship-specific accumulated choice histories–the "visited" social locations. Our decoding outperformed this null as well (Initial: $CI_{95\%}$ = [-0.12, 0.025], t_{17} = -1.4, left-tailed p = 0.089; Validation: $CI_{95\%}$ = [-0.13, -0.02], t_{31} = -2.93, left-tailed p < 0.005), suggesting that the experienced locations, and not just the choices, are represented in the hippocampal sequences.

Nonlinear embeddings outperform linear embeddings

Lastly, we allowed for the neural trajectories to be nonlinear by using a nonlinear dimensionality reduction algorithm. To test whether this choice contributed to location decoding, we compared the decoding errors from the nonlinear embeddings against those from linear embeddings. As expected, the decoding errors were smaller for the nonlinear embeddings (Initial: $CI_{95\%}$ = [-0.06, -0.01], t_{31} = -2.51, left-tailed p < 0.05; Validation: $CI_{95\%}$ = [-0.03, -0.01], t_{31} = -3.4, left-tailed p < 0.001), suggesting the neural trajectories are nonlinear.

These findings support our hypotheses: the hippocampus tracks character-specific social relationships, with trajectory-like representational sequences in a low-dimensional neural state space (see **Figure 4**).



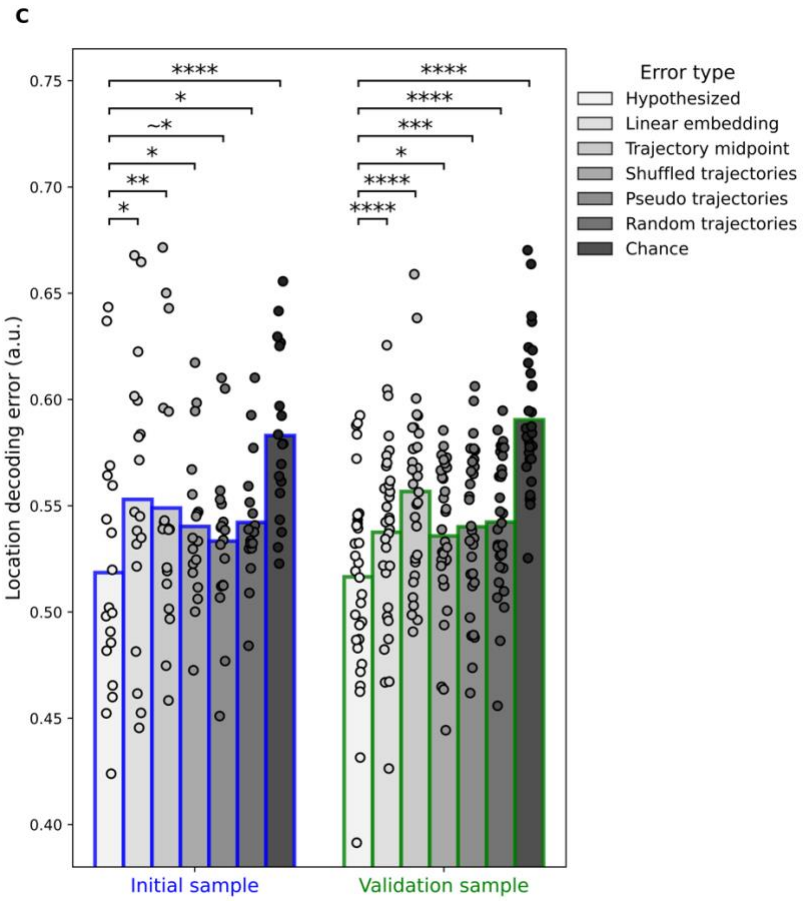


Figure 4. Trajectory location decoding analysis. *(A) Behavioral trajectories for each character were analyzed; one simulated behavioral trajectory is highlighted in tan. A leave one trial out approach was used; the held-out trial in this example is shown in red. (B) The hippocampal patterns were isolated in the same way. A simulated neural sequence associated with the behavioral trajectory is shown. We estimated the location of this held-out point on the -based approximation of the neural trajectory. Using a parameterization of the spline, we then decoded the held-out trial's social location and calculated the distance of this estimate to the actual task location to estimate decoding errors. (C) The average error for the different trajectory models in the left hippocampus for both samples. The hypothesized model was compared to multiple null models: a linear embedding model to test the importance of nonlinear dimensionality reduction in preserving the neural sequence ("linear embedding"); a dummy model where only the trajectory midpoint was predicted ("trajectory midpoint"); a model where participant choices were shuffled within each relationship ("shuffled trajectories"); a model where participant choices were pseudo-randomly selected, across characters, to calculate locations ("pseudo trajectories"); a model where random choices were simulated to create trajectories that respected the task structure but did not have choice history ("random trajectories"); and a model that permuted brain-behavior relationships but preserved the temporal autocorrelation ("chance"). P-value significance are indicated by asterisks: ~* < 0.1,* < 0.05, ** < 0.01, *** < 0.005, **** < 0.001.*

The number of distinct hippocampal state clusters correlates with real-world social networks

A good map ought to have enough detail at the right level of abstraction to allow for adaptive navigation. We suspect that this is also the case for the social mapping of relationships: given the inherent uncertainties of social life, navigating different social relationships may require

richly detailed maps. To test this, we asked whether the number of distinct clusters of left hippocampal states that track affiliation and power coordinates—a proxy for the number of distinct states in the social map representations—relates to the size of real-world social networks. To find the neural clusters, we selected the number of clusters that maximized the shared information of hippocampal embedding clusters and affiliation and power coordinate clusters. We then classified held out hippocampal states into these clusters. This approach yielded left hippocampal clusters with social location information: held-out hippocampal states had associated affiliation and power locations that were much closer to the clusters they were assigned to than expected by chance (Initial: $CI_{95\%} = [0.3, 0.49]$, $t_{17} = 8.84$, right-tailed $p < 0.0001$; Validation: $CI_{95\%} = [0.42, 0.59]$, $t_{17} = 12.07$, right-tailed $p < 0.0001$).

Once we confirmed that the hippocampal state clusters contained social location information, we compared these cluster estimates to self-reported real-world social network size (in the Validation sample only). As expected, the hippocampal cluster estimate positively correlated with the number of people in ($CI_{95\%} = [0.03, 0.62]$, $t_{30} = 2.28$, right-tailed $p < 0.05$, adj. $R^2 = 0.09$) participants' social networks (see **Figure 5**). This effect was significant even when including covariates for mean reaction time and memory.

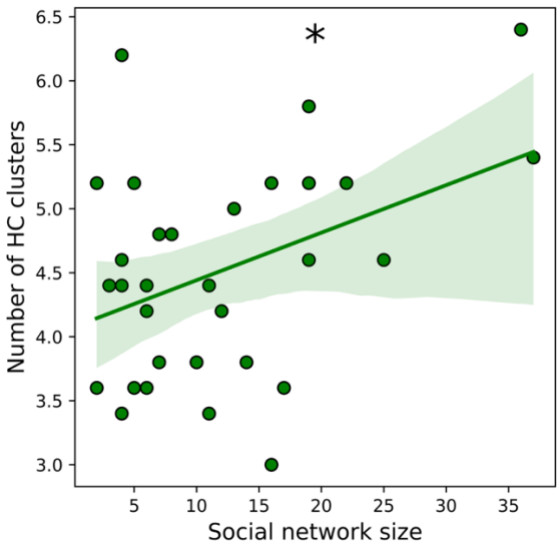


Figure 5. The number of distinct left hippocampal clusters correlates with real-world social network size. *We estimated the number of distinct left hippocampal clusters that optimized the overlap with behavioral clusters, and then regressed these estimates onto estimates of social network size. None of the data points exceed an outlier threshold of 3 standard deviations from the mean. 95% confidence intervals are indicated by the shaded area and p-value significance is indicated by the asterisk: * < 0.05. Validation sample only (n = 32).*

Discussion

Here, we study a case of conceptual trajectories, which are sequences of abstractly, rather than physically, related events. We show that by connecting conceptually related interactions, the hippocampus represents social relationships with neural trajectories. Such hippocampal

linking in abstract space can serve as a possible mechanism for linking events in all manners of spaces.

Social relationships as trajectories on a hippocampal map-like manifold

We first show that left hippocampal pattern similarity is higher in social interactions with the same underlying dimension (i.e., affiliation compared to affiliation, and power compared to power) than interactions with different dimensions (i.e., affiliation to power), consistent with abstract affiliation and power representations. We then show that the character-specific sequences of hippocampal states behave like neural trajectories: we decoded abstract social locations much better when a relationship-specific trajectory was assumed as compared to a variety of plausible null models. As such, the hippocampus may represent the social interactions that make up a relationship—a social trajectory—in ordered and connected sequences of patterns, akin to a neural trajectory within a generalizable map-like manifold. Lastly, we show that the number of hippocampal state clusters—a proxy for the number of distinct hippocampal states engaged during the interactions—is positively correlated with participants' real-world social network size, suggesting that more expressive maps (i.e., larger number of possible states) relate to more adaptive social behaviors.

Conceptual sequences on low-dimensional manifolds

Previous work has shown that the hippocampus tracks temporally sequential events, such as trajectories in physical space (i.e., traversing sequentially connected locations[12]) as well as sequences of arbitrary, laboratory-made information units, such as ordered lists of words or scenes (e.g.,[13]). Social relationships, however, are *conceptual* sequences: the events that constitute a relationship are spatially and temporally distant but conceptually linked, constituting a trajectory in abstract social space. At each interaction the hippocampus may reactivate a representation of the last interaction's social location, and then infer how the current interaction changes the relationship's location. Stacking these interaction representations together, then, would re-create the neural trajectory of the relationship through social space.

Why would the hippocampus encode sequences with an abstract and low-dimensional neural trajectory? We cannot answer this question directly with fMRI, but there is a plausible account from the neurophysiological literature. Hippocampal neurons are known to track the physical locations of oneself[14] and others[15], as well as social identity[16] and more abstract information[17], suggesting the hippocampus may compute locations in abstract spaces—including social space. Moreover, the activity of hippocampal cell populations is ordered with respect to task dimensions, such that population-level patterns track locations in task space. Because hippocampal firing fields overlap to cover the task space[18], trajectories through task space should activate sequences of correlated hippocampal patterns, forming a continuous neural trajectory that reflects the behavioral trajectory. This activity may be intrinsically low-dimensional: many hippocampal cells co-activate to encode locations in task space, restricting the possible activity patterns to a low-dimensional subspace[19]. Many possible neural states in theory become many fewer in practice, as neuronal correlations restrict intrinsic activity to low-dimensional manifolds.

Social manifolds for social networks

Hippocampal patterns during social interaction contain information about relationship history because the social coordinates they encode are a compressed representation of that history. Representing different relationships as locations within a neural manifold would allow efficient comparisons. To infer relationships between codes, downstream regions can compare the spatial properties of the patterns. For example, proximity on the manifold implies nearby social locations.

Experiments with richer social network information could test other related questions. For example, the low-dimensional hippocampal representation may also allow generalizations across social contexts, such as map-based comparisons across distinct social networks. Moreover, the relationships in the Social Navigation Task are novel and thus each interaction is rich with information that can update the relationship representation. How are longer-term relationships, where the person is highly familiar, mapped and updated? For example, is the threshold for updating the representation higher for these longer-term relationships?

In this study, the complexity of the hippocampal manifolds related to real-world social network size; this also suggests next questions. A system with only a small number of states may be inflexible and tend towards over-generalization. In contrast, a system with an excessive number of states may be energetically costly and susceptible to noise. What gives rise to distinct manifold locations? One possibility is that the number of unique hippocampal states reflects excitation to inhibition ratio, or perhaps the dynamics of an attractor network[20]. Intracranial electrophysiological recordings, with much finer temporal dynamics than fMRI recordings, are especially suited to ask these sorts of questions. These findings could also relate to findings in social psychology, such as Dunbar's number, which suggests an upper limit on the number of social connections humans can maintain at once[21]. The complexity of the hippocampal social manifold may relate to an individual's upper relationship limit.

Beyond social trajectories

The approach used in this study can be applied to similar questions in non-social domains. An obvious question is whether abstract concept spaces are also represented on low-dimensional hippocampal manifolds. Context-related effects[22] may also be explainable by manifolds: memories of different episodes that share a higher-order context could have correlated patterns of activity, creating a common relational manifold. The structure of a manifold may hint at its underlying dynamics, allowing the formation and testing of novel computational hypotheses. There are also questions to be asked about the interactions between manifolds. For example, the entorhinal cortex also tracks structure, exhibiting strikingly regular activity through space and a correspondingly structured manifold[23]; how do hippocampal and entorhinal manifolds relate, and what does this reveal about how the underlying cell populations interact? Of course, the hippocampal formation is not the only region to display neural manifolds[24]; highly correlated activity is common across neural populations, suggesting low-dimensional manifolds within high-dimensional patterns may also be common.

Conclusion

In this study, we provide the first evidence that the hippocampus represents social relationships with ordered sequences of low-dimensional neural patterns. Each relationship had its own sequence of states, and all relationships were embedded in the same manifold. The number of distinct clusters of states on this manifold also relate to social function, as measured by the size of real-world social networks. As Edward Tolman foresaw, “narrow maps” appear to restrict social function[25]. These results put forward a novel way to look at representations of all kinds of evolving latent relationships, physical to social.

Analysis

Behavioral analysis

Behavioral modeling

Each character started the task at the neutral origin (0, 0) of the social space. With each participant choice, the current character’s coordinates were implicitly updated in the positive or negative direction along the current interaction dimension (i.e., each decision moved the character +/- 1 arbitrary unit along affiliation or power). For any given decision trial (t), the current character’s (c) affiliation and power coordinates are the cumulative sums of the trialwise affiliation and power updates:

$$\begin{aligned} \text{affiliation}_{c,t} &\leftarrow \text{affiliation}_{c,t-1} + \Delta\,\text{affiliation}_{c,t} \\ \text{power}_{c,t} &\leftarrow \text{power}_{c,t-1} + \Delta\,\text{power}_{c,t} \end{aligned}$$

where Δ is -1 or +1 if the trial is an interaction of that dimension, and 0 if not.

Thus, the sequences of interaction decisions create different relationship trajectories through the latent affiliation and power space (see **Figure 1**).

Subjective character placement analysis

We modeled the choices in the character-specific interaction sequences as affiliation and power coordinates; if the brain represents these coordinates, participants may have subjective access to them. Thus, we expected participants’ post-task placements of the characters into a 2D affiliation and power space to be closer to their end-of-task behavioral locations than expected by chance. For each participant in the Validation sample, we calculated the mapping error as the average Euclidean distance between the character-wise behavioral and subjective locations:

$$\text{mapping error}_M = \frac{1}{C}\sum_{c=1}^{C}\sqrt{(\text{affiliation}_c - \text{subj. affiliation}_c)^2 + (\text{power}_c - \text{subj. power}_c)^2}\,.$$

To establish participant-specific chance distributions, we permuted the subjective locations and re-calculated the average error, 100 times for each participant. We compared each participant’s average distance against their chance distribution by calculating a z-score with respect to the permutation distribution:

$$\text{mapping error}_z = \frac{\text{mapping error}_M - \text{null error}_M}{\text{null error}_{SD}}.$$

We tested whether these mapping errors were smaller than 0 with a left-tailed t-test.

Given that retrieving the subjective locations should be memory-dependent, we also predicted a negative relationship between mapping mismatch and post-task memory recall: participants with better memory of the interactions should place the characters closer to their behavioral locations. We tested this prediction using Pearson's correlation and a left-tailed p-value.

fMRI analysis

Regions of interest (ROI) definitions

ROI analyses were used to test our *a priori* hippocampal predictions. Probabilistic right and left hippocampus masks were defined from the Harvard-Oxford Atlas in FSL (FMRIB Software Library), co-registered and re-sampled to the Validation sample's functional images, and binarized at a threshold of 25% probability. For ROI analyses, we focused on the left hippocampus given prior work showing left lateralization of social representations[10], including in this task[3,7].

Multiple comparisons correction

The family-wise error rate (FWER) for significant testing was controlled at an alpha of 0.05. For searchlight analyses, we used a max-type permutation testing procedure[26]. In the hippocampus, we used small volume correction and in the rest of the brain we corrected for remaining (non-hippocampal) comparisons, and thresholded the cluster sizes to be at least 10 contiguous voxels.

Trial-wise general linear modeling

To estimate single trial activation patterns, general linear models (GLMs) were fitted to each voxel in the functional images using SPM12. Unsmoothed images were used to preserve the spatial resolution of the multi-voxel patterns. Microtime resolution was set to the number of slices collected and microtime onset was set to the middle slice. To reduce temporal autocorrelation in the time series, we used a high-pass filter of 128 seconds (1/128 Hertz) and prewhitening (with SPM's FAST algorithm).

The design matrices had separate regressors for each decision trial to estimate all decision trials' weights in the same regression. For different analyses (see below), we modeled the events with a stick function at trial onset or boxcar function over the reaction time (from onset to the button press), depending on whether the neural representation was independent of the decision (see next section). The narrative trial onsets were modeled by an additional regressor. The decision and narrative trial regressors were convolved with a canonical hemodynamic response function to model the blood oxygenation level dependent (BOLD) signal changes expected from the events. The six realignment parameters from preprocessing were included in the design matrices to regress out residual motion-related variance. To further improve the signal-to-noise ratio, the GLMs were only estimated for voxels with average signal

greater than 50% of the global signal. After GLM estimation, each participant had a beta image with beta series (β_1, β_2, …, β_{63}) in each voxel that reflects trial-specific activation magnitudes. The Initial sample's beta images were resampled to match the higher spatial resolution Validation sample's images, so that all analyses cover the same number of voxels and amount of brain volume across participants.

Social dimension neural abstraction analysis

The affiliation and power dimensions are akin to social contexts: we expected the hippocampus to represent interactions of the same dimension more similarly than interactions of different dimensions. In other words, we expected the affiliation and power interactions to be represented abstractly, such that their representations generalize across trials. We used representational similarity analysis (RSA) to test this hypothesis. We assumed that representations of social dimension are independent of and precede decisions, and so we used trial-wise beta series from GLMs that modeled the decision trial onsets with a stick function.

For each participant and each ROI, we calculated the average pattern correlations (Pearson's *r*) for same dimension trial pairs and for different dimension trial pairs. We then contrasted the two average correlations to get the correlation difference:

$$\Delta r = r_{\text{same}} - r_{\text{different}} \ .$$

These differences were entered into a right-tailed t-test to test our prediction that same dimension trial pairs have more correlated patterns than different dimension trial pairs.

Importantly, comparing same and different dimension trial pairs balances various other narrative-related variables. Character identity is balanced, because the characters appear in the same number of affiliation and power trials; character familiarity differences (i.e., in the number of previous decision trials) are also not different between same and different dimension trial pairs (t-test $p > 0.5$).

To calculate these pattern correlation differences across the whole brain, a moving searchlight analysis was used. For each voxel, the analysis was run on a set of voxels from a surrounding sphere (diameter = 11 voxels); the resulting estimate was stored in the center voxel. To calculate correlations between brain patterns, we used Pearson's *r* with Fisher's z-transform (i.e., the inverse hyperbolic tangent). The resulting images were smoothed with a 6mm Full Width at Half Maximum (FWHM) kernel.

Neural manifold estimation

We used two different dimensionality reduction algorithms to estimate low-dimensional hippocampal manifolds: one nonlinear algorithm and one linear algorithm, as a control. For the nonlinear algorithm, we used Locally Linear Embedding[11] (LLE). LLE aims to preserve local relationships between data points: points that are close in the high dimensional space should remain close in the lower dimensional embedding. First, LLE finds the k-nearest neighbors of each data point (where k is a user selected hyperparameter). Then, LLE computes the weights that produce the best linear reconstruction of the data points from their neighbors. Last, it finds the low-dimensional embedding that best preserves these weights. As our linear control, we

used Principal Components Analysis[27] (PCA). PCA uses the eigendecomposition of the data's covariance matrix to project the data onto the orthogonal axes that explain the most variance, and as such is limited to explaining linear patterns in the data.

For a given ROI and algorithm, we reduced each participant's beta series (shape = 63 decision trials x number of voxels) from its voxelwise representation into 5 dimensions (shape = 63 decision trials x 5 dimensions). We chose an embedding dimensionality of 5 to capture variation related to the 2 task dimensions (affiliation and power), as well as other task-relevant variation (e.g., character identity, familiarity, narrative progression). See the supplement section **Hippocampal trajectory decoding at different dimensions** for analyses at other dimensions.

Relationship neural trajectory analysis

We expected the 2D relationship trajectories to have corresponding low-dimensional hippocampal trajectories. This led to several predictions. First, we should be able to use a trial's low-dimensional neural pattern to decode the relationship's social coordinates on that trial. Second, a neural trajectory should be ordered in space and time: the position of the trial within the neural sequence should correspond to its location in the behavioral trajectory. Third, the relationship-specific hippocampal trajectories should be specific to the choice history; as such, the location decoding should be better for the character specific sequences than for sequences of random choices across the task. To test these hypotheses, we used a spline-based trajectory analysis on each participant's low-dimensionally embedded fMRI patterns.

Fit and parameterize linear splines to neural data

For each character (12 trials each), we used a leave one trial out decoding procedure. In each split, we fit a linear spline through the 11 training trials. A linear spline is a piecewise linear function that defines line segments that connect each pair of successive points. For a 2D example, the piecewise function is built from a series of local interpolants,

$$L_i(x) = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i), \quad \text{for } x \in [x_i, x_{i+1}] ,$$

where i = number of trials – 1.

This produces a spline made up of 10 line segments that approximates the neural trajectory as an ordered and connected sets of locations.

We then parameterized the spline: we assigned each trial's spline location with its affiliation and power coordinates. The affiliation and power values were min-max normalized within each character and along each dimension separately to be within the range of [0,1].

Decode held-out trial's parameter value

For each split in the spline fitting procedure, we held out one of the trials to assess how well the neural spline captured the behavioral trajectory. To predict the held-out trial's affiliation and power coordinates, we calculated the average spline location of its 100 nearest neighbors

(defined using Euclidean distance) on the spline (evaluated for 1000 points) and decoded its location as the interpolated affiliation and power coordinates.

Evaluate decoding performance

After repeating this procedure for all trials, we calculated the decoding error as the average Euclidean distance between actual and decoded locations:

$$\text{decoding error}_M = \frac{1}{T}\sum_{t=1}^{T}\sqrt{(\text{affiliation}_t - \text{decoded affliation}_t)^2 + \left(\text{power}_t - \text{decoded power}_t\right)^2}.$$

We compared these errors to errors from a variety of null models; in each null model, we removed one assumption to assess its decoding importance. We tested whether our decoding errors were smaller than the null models' decoding errors with t-tests and left-tailed p-values. For null models without permutations, the error difference for each participant was the difference between the mean decoding:

$$\text{error difference} = \text{decoding error}_M - \text{null error}_M .$$

For permutation-based null models, each participant's error difference was turned into a z-score with respect to the permutation distribution:

$$\text{error difference}_z = \frac{\text{error difference}}{\text{null error}_{SD}} .$$

Below, we detail the null models.

Can social location be decoded from hippocampal sequences?

To estimate a chance decoding level, in each training spline we circularly shifted the neural patterns and re-ran the analysis 10 times (the maximum number of circular shifts, given the training splines had 11 trials). Circular shifting breaks the relationship between the brain patterns and behavior while preserving temporal autocorrelation; thus, if decoding outperforms the chance level produced from circular shifting, temporal autocorrelation is unlikely to be the explanation for decoding performance.

Does the position in the interaction sequence matter?

The position of a trial in the neural pattern sequence should be informative of its location in the relationship trajectory. To test the importance of trial order, we used a null model where the decoded location was always the midpoint of the neural spline. If the location of the trail in the neural embedding space is irrelevant to decoding its location in the relationship trajectory, we should fail to decode the locations better than this null. If instead the trial's position in the neural sequence order contains information about the location of the relationship, the true errors should be smaller than these midpoint only errors.

Does the specific relationship path matter?

The specific sequence of interactions in a relationship determines the path of the relationship trajectory. To test whether the specific path matters in location decoding, we created a null model where we shuffled the choices before calculating relationship-specific trajectories, preserving the choice distributions and end locations of the relationships but varying the choices sequences and relationship paths. We ran this analysis 50 times for each participant, to produce participant-specific null distributions. Decoding errors should be smaller than these null decoding errors if the order of the choices matters to the relationship representation.

Are hippocampal trajectories relationship history specific?

Even if the location prediction errors are smaller than all the above null prediction errors, the hippocampus may track some other feature of the sequence, and not the relationship trajectories themselves. We simulated random choice behavior (i.e., on each trial, each option was given a 50% probability of being selected) to estimate a null distribution where the specific choice history is random. We ran 50 simulations for each participant. If the trajectories contain information beyond the underlying task structure (e.g., character identity, familiarity, social dimensions, etc.), our model should decode the social location better than this simulated distribution.

Do the specific locations matter, or just choices?

It is also possible that a trajectory-like effect reflects the effects of the choices, and not their relationship specific accumulation into trajectories. To test this, we created pseudo-trajectories: we divided the 60 decision trials into 5 random sets of 12, with the constraints that each set had 1 affiliation and 1 power trial from each character and 1 more trial of each dimension from random and different characters. This created 5 trial sequences with the same number of trials (12), same distribution of social dimensions (6 affiliation trials and 6 power trials) and same temporal order as the real character sequences, but without the relationship history specificity. We calculated affiliation and power coordinates from these choice sequences, and then ran the spline analysis. We ran this null model 50 times for each participant.

Do nonlinear embeddings outperform linear embeddings?

We allowed for the neural trajectories to be nonlinear by using LLE, a nonlinear embedding algorithm. To test whether this contributed to decoding performance by comparing the errors from LLE against errors from the PCA, a linear embedding algorithm, at the same dimensionality (5D). If decoding with LLE outperforms decoding with PCA, it suggests that nonlinear structure (e.g., local curvature) contains trajectory-related information.

Neural state clustering analysis

Maps are for navigation; maps without enough detail impair navigation. We hypothesized that a smaller number of distinct states in the hippocampal social maps relate to impoverished social networks. To test this idea, we correlated estimates of the number of distinct hippocampal states that contain information about affiliation and power with self-reported social network size (i.e., the number of real-world relationships).

To estimate each participant's number of distinct hippocampal states, we clustered the low-dimensional hippocampal embeddings and the social locations from behavior. To cluster the neural embeddings, we used nonlinear Laplacian eigenmap clustering (i.e., spectral clustering). It computes a pairwise Euclidean distance matrix between the trials' neural embedding locations, and implicitly projects them into higher dimensions via a radial basis function. Then, the distance matrix's normalized Laplacian matrix is calculated and eigendecomposed, and a user chosen number of the smallest eigenvalues are used to create the clusters. To cluster the affiliation and power locations from the behavior, we used the linear K-means clustering algorithm. K-means takes in a user-specified number of clusters, randomly initializes the cluster centers, and then iteratively updates the centers to minimize the within-cluster sum of squared errors.

We used a leave-one-character-out cross-validation approach. On each split, we used trials from 4 of the 5 characters (48 trials) to cluster the neural embeddings and behavioral locations simultaneously. For each split, we chose the number of clusters within the range of [2, 15] that maximized the overlap between the behavioral and neural clusters. Adjusted mutual information was used to estimate the overlap across the clusters because it measures the cluster overlap while adjusting for chance so that more clusters does not artificially inflate the shared information estimate. After selecting the clustering, we assigned the held-out character's neural states (12 trials) to the training set's neural clusters by the majority votes of the held-out trials' 10 nearest neighbors (using Euclidean distance).

We evaluated this clustering procedure by validating that the held-out neural states were assigned to clusters of trials with closer affiliation and power locations than expected by chance. We first calculated the distances from each held-out trial's social coordinates to the average social coordinates of their assigned clusters. We compared these distances to chance distances, which were estimated by re-running the clustering procedure with circularly shifted neural embeddings (the average of the 47 possible permutations). We expected the average distances to be smaller than these chance distances, which we tested with a left-tailed t-test.

After establishing that the clusters contain social location information, we averaged the number of clusters assigned to the held-out characters' neural states across the cross-validation splits. This number was our estimate of the distinct number of hippocampal state clusters that contain social location information. We expected the number of unique hippocampal state clusters to positively correlate with the size of real-world social networks (from the Social Network Index). To test this prediction in the Validation sample, we calculated right-tailed p-values from an Ordinary Least Squares (OLS) regression:

$$\text{Social network size} \sim \beta_0 + \beta_1 \text{Cluster count} .$$

Materials and methods

Independent samples

To ensure the effects we find are robust, we tested our hypotheses in two independent samples, called the "Initial" and "Validation" samples. The Initial sample was collected for a previous study (Tavares et al., 2015) and included 18 of 21 participants (8 female), after excluding 1 for psychiatric evaluation (Psychiatric Diagnostic Screening Questionnaire) and 2

for DICOM corruption during archiving. The Validation sample included 32 of 39 participants (17 female), after excluding 2 for poor post-task character memory (<= chance [20%] for 5 main characters) and 5 for high motion (mean framewise displacement >= 0.3mm).

The two samples did not differ in terms of participant sex ($\chi^2 = 0.45$, $p = 0.5$) but were significantly different on age (in years; Initial: mean (*M*) = 29.33, standard deviation (*SD*) = 3.42; Validation: *M* = 43.73, *SD* = 10.5; Welch's unequal variances t-test $p < 0.0001$).

The Institutional Review Board of the Icahn School of Medicine at Mount Sinai approved the experimental protocols for both samples. All participants provided written informed consent and were paid for their participation.

Social Navigation Task

Task description

The Social Navigation Task is a narrative-based social interaction game. Pre-task instructions are minimal: participants are told they will complete a task where they interact with different fictional characters, and to just behave as they naturally would. Importantly, the participants are never told or otherwise taught about the affiliation and power dimensions underlying the interactions. From the point-of-view of the participant, they are simply interacting with characters in a narrative. Thus, the affiliation and power dimensions are fully latent (i.e., unobservable).

The narrative begins with the participant being told that they have just moved to a new town, and they need to find a job and a place to live. It unfolds over various social interactions that vary in their details, and then culminates in the participant having successfully found employment and housing. There are two types of trials: "narrative" trials where characters talk or take actions, or where background information is provided, and "decision" trials where the participant makes decisions in one-on-one interactions with characters that can change the relationships. The participant's choices during the decision trials shape the storyline, which is otherwise fixed.

The characters' gender and racial presentations were counterbalanced. There were two versions for character gender: one where three of the characters were men and the others were women, and another where the genders were flipped. In the Validation sample, the characters' racial presentation (i.e., skin color) was counterbalanced in the same way, but with the additional control that lighter and darker skinned characters were versions of the same underlying character image. The task version used for the Initial sample had approximately the same text and options (with small modifications), but with more cartoon-like character images (Tavares et al., 2015).

Participants used a button response box to select between two options on each decision. The option's (1 or 2, for the index and middle fingers) choice direction (+/- on the current dimension) was counterbalanced. The decision trials are categorized as either affiliation or power decisions. Each decision consists of a choice between two options that move the character in either the negative (-1) or positive (+1) direction along the dimension. There are 5 characters, each with 6 affiliation and 6 power decisions, for 12 decisions per character and 60

decisions total. Affiliation decisions were operationalized as decisions whether to share physical touch, physical space or information (e.g., to share their thoughts on a topic). Power decisions were defined as decisions to submit to or issue a directive/command, or otherwise exert or give control. There was also a 6th neutral character with 3 neutral decisions; these trials were not used for analysis.

Post-task measures

Task memory

To validate that the participants attended to the task, after it was completed, participants answered 30 memory questions about the characters (outside the scanner). In each question, the options were 5 of the 6 characters (including the neutral): each character was presented as an option in 25 questions and was the correct answer in 5 of those questions.

Subjective character placement

Participants may form a subjective representation of the characters' 2D social locations. To probe this, the participants completed a character placement task outside of the scanner after the task (Validation sample only). The participants were instructed to drag-and-drop colored dots representing each of the 6 characters onto locations in a 2D affiliation and power space, according to the participant's perceptions of their relationships with the characters. The placements were relative to the participant's theoretical point-of-view, which was represented by a red dot placed at the max affiliation and neutral/middle power values.

Self-report questionnaires

Participants also completed self-report questionnaires. For the purposes of this study, we used the Social Network Index (Cohen et al., 1997) in the Validation sample. Participants in that sample also completed questionnaires for a larger ongoing study.

fMRI acquisition and pre-statistics

Image acquisition

Scans were collected in a single run of approximately 26 minutes. Both samples' data were acquired on 3-Tesla (3T) scanners but with different image acquisition parameters.

<u>*Initial sample:*</u> Images were collected on a Siemens Allegra 3T scanner (Siemens, Erlangen, Germany). T2*-weighted images were collected with a single-shot echo-planar imaging (EPI) pulse sequence with the following parameters: flip angle = 90°, echo time (TE) = 35 milliseconds (ms), repetition time (TR) = 2 seconds (s), 36 slices, 64x64 matrix, voxel size = 3 millimeter^3 (mm^3). T1-weighted images were collected with a magnetization-prepared rapid gradient-echo (MPRAGE) protocol with voxel size = 1 mm^3.

<u>*Validation sample:*</u> Images were collected on a Siemens Skyra 3T scanner. T2*-weighted images were collected with a multiband slice EPI pulse sequence with the following parameters: multiband acceleration factor = 7, flip angle = 60°, TE = 35 ms, TR = 1 s, 70

slices, 108x108 matrix, voxel size = 2.1 mm^3. T1-weighted images were collected with a MPRAGE protocol with voxel size = 0.8 mm^3.

Image preprocessing

Image preprocessing was conducted using SPM12 (Wellcome Trust Centre for Neuroimaging), with standard steps. To correct for head motion, the functional images were realigned to the first volume using 6 parameter rigid body transformation (3 translations and 3 rotations) and then unwarped to account for magnetic field inhomogeneities. The realigned images were slice-time corrected to the middle slice using normalized mutual information, then co-registered in alignment with the MPRAGE to the mean unwarped image. The MPRAGE image was then segmented into 6 tissue classes (gray matter, white matter, cerebral spinal fluid, skull, soft-tissue and air) and the resulting forward deformation parameters were used to normalize the images to a standard Montreal Neurological Institute (MNI) template image (using 4th-degree B-spline interpolation). The functional images were not smoothed.

Supplement

Hippocampal trajectory decoding across different dimensions

To test the effect of choice of the dimensionality in the neural embedding step, we reduced the dimensionality (see **Neural manifold estimation** in Methods) and ran our trajectory analysis for dimensions 2 to 10. To estimate decoding performance at each dimension, we compared the mean decoding error against the null models' decoding (see **Relationship neural trajectory analysis** in Methods):

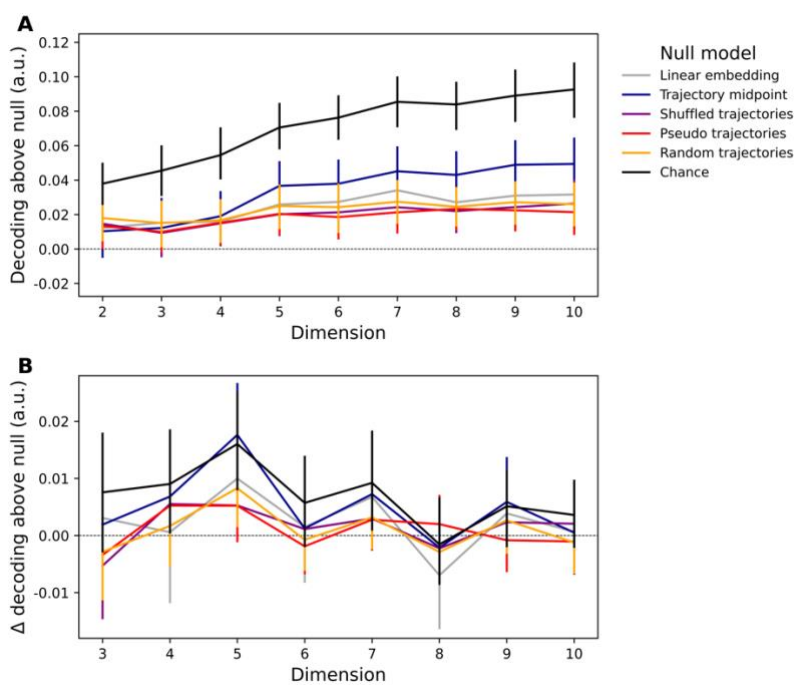
$$\text{decoding above null}_D = \text{decoding}_D - \text{null decoding}_D .$$

Here, positive values indicate greater than chance decoding for the given null. Generally, the decoding improvement above the null improves up to about 5 dimensions and then plateaus (see **Supplemental figure 1**).

To visualize the change in performance across dimensions, for dimensions 3 to 10, we compared the error improvement over the null at each dimension against the improvement at the previous dimension:

$$\Delta \text{ decoding above null}_D = \text{decoding above null}_D - \text{decoding above null}_{D-1} .$$

Increasing from 4 to 5 dimensions provides a substantial improvement in decoding performance, but with clearly diminishing returns at higher dimensions–at times even worsening decoding performance. As such, 5 dimensions–a choice made *a priori*–is a good dimensionality for compressing the neural patterns.



Supplemental figure 1. *All trajectory models were evaluated on embedding dimensions 2 to 10. (A) Top plot is how much better the 'real' model does than the null. Larger values mean smaller errors for the real model compared to the null model. (B) Bottom plot is the comparisons for dimensions 3 to 10 against the previous dimension's decoding, for each null. Larger values mean a larger improvement over the null from the previous dimension. For both plots, the mean error difference is shown along with 95% confidence intervals.*

Acknowledgments

Funding: DS is supported by the National Institute of Health, USA (R01MH122611, R01MH123069); KB is supported by the National Institute of Drug Abuse, USA (K23-DA045928); MS is supported by the National Institute of Mental Health, USA (F31MH123123).

Author contributions

MS and DS conceived of the study. PKB, VS and KB collected the data. MS ran the analyses. MS, DS and KB wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

References

1. Schafer, M. & Schiller, D. Navigating Social Space. *Neuron* **100**, 476–489 (2018).
2. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat Neurosci* **20**, 1643–1653 (2017).
3. Tavares, R. M. *et al.* A Map for Social Navigation in the Human Brain. *Neuron* **87**, 231–243 (2015).
4. Howard, L. R. *et al.* The Hippocampus and Entorhinal Cortex Encode the Path and Euclidean Distances to Goals during Navigation. *Current Biology* **24**, 1331–1340 (2014).
5. Constantinescu, A. O., OReilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science (1979)* **352**, 1464–1468 (2016).
6. Theves, S., Fernandez, G. & Doeller, C. F. The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology* **29**, 1226-1231.e3 (2019).
7. Zhang, L. *et al.* A specific brain network for a social map in the human brain. *Sci Rep* **12**, 1–16 (2022).
8. Park, S. A., Miller, D. S., Nili, H., Ranganath, C. & Boorman, E. D. Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron* **107**, 1226-1238.e8 (2020).
9. Schuck, N. W. & Niv, Y. Sequential replay of nonspatial task states in the human hippocampus. *Science (1979)* **364**, (2019).
10. Kumaran, D., Melo, H. L. & Duzel, E. The Emergence and Representation of Knowledge about Social and Nonsocial Hierarchies. *Neuron* **76**, 653–666 (2012).
11. Shepard, . R N, Kumar, ; V, Grama, A., Gupta, A. & Karypis, G. *Nonlinear Dimensionality Reduction by Locally Linear Embedding*. *P. Y. Simard, Y. LeCun, J. Denker, Adv. Neural Info. Proc. Syst* vol. 1 www.research.att.com/yann/ocr/mnist. (1994).
12. Dabaghian, Y., Brandt, V. L. & Frank, L. M. Reconceiving the hippocampal map as a topological template. *Elife* **3**, e03476 (2014).
13. Ranganath, C. & Hsieh, L. T. The hippocampus: A special place for time. *Ann N Y Acad Sci* **1369**, 93–110 (2016).
14. O'Keefe, J. & Nadel, L. *The hippocampus as a cognitive map*. *Oxford: Clarendon Press* (1978).
15. Omer, D. B., Maimon, S. R., Las, L. & Ulanovsky, N. Social place-cells in the bat hippocampus. *Science (1979)* **359**, 218–224 (2018).
16. Hitti, F. L. & Siegelbaum, S. A. The hippocampal CA2 region is essential for social memory. *Nature* **508**, 88–92 (2014).
17. Aronov, Dmitry, Nevers, Rhin, Tank, D. W. Mapping of a non-spatial dimension by the hippocampal/ entorhinal circuit. *Nature* **543**, 719–722 (2017).
18. Curto, C. What can topology tell us about the neural code? *Bulletin of the American Mathematical Society* **54**, 63–78 (2017).
19. Nieh, E. H. *et al.* Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021).
20. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat Rev Neurosci* **23**, (2022).
21. Dunbar, R. *How many friends does one person need? Dunbar's number and other evolutionary quirks*. (Harvard University Press, 2010).
22. Baldassano, C. *et al.* Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).

23. Gardner, R. J. *et al.* Toroidal topology of population activity in grid cells. *Nature* **602**, 123–128 (2022).
24. Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A. & Fiete, I. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nat Neurosci* **22**, 1512–1520 (2019).
25. Tolman, E. C. Cognitive maps in rats and men. *Psychol Rev* **55**, 189–208 (1948).
26. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *Neuroimage* **92**, 381–397 (2014).
27. Hotelling, H. *Analysis of a Complex of Statistical Variables into Principal Components*. (1948).