Between-subject prediction reveals a shared representational geometry in the rodent hippocampus

Hung-Tu Chen, Jeremy R. Manning, Matthijs A. A. van der Meer*

Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH 03755

*Correspondence: mvdm@dartmouth.edu

Summary

How a memory system encodes related experiences has consequences for what operations the system supports. For instance, independent coding enables retention of potentially important idiosyncratic details by reducing interference, but makes it difficult to generalize across experiences. Strikingly, the rodent hippocampus constructs statistically independent representations across environments ("global remapping") and assigns individual neuron firing fields to locations within an environment in an apparently random fashion, processes thought to contribute to the role of the hippocampus in episodic memory. This random mapping implies that it should be challenging to predict hippocampal encoding of a given experience in a one subject based on the encoding of that same experience in another subject. Contrary to this prediction, we find that by constructing a common representational space across rats ("hyperalignment"), we can consistently predict data of "right" trials (R) on a T-maze in a target rat based on 1) the "left" trials (L) of the target rat, and 2) the relationship between L and R trials from a different source rat. These cross-subject predictions outperformed a number of control mappings, such as those based on permuted data that broke the relationship between L and R activity for individual neurons, and those based solely on within-subject prediction. This work constitutes proof-of-principle for successful cross-subject prediction of ensemble activity patterns in the hippocampus. This novel approach provides new insights in understanding how different experiences are structured, and suggests further work identifying what aspects of experience encoding are shared vs. unique to an individual.

Keywords: place cells, remapping, generalization, pattern completion, pattern separation, hyperalignment

Word Count: 3133 words

Number of Figures: 4 in main text, 4 in Supplementary Information

Introduction

A fundamental challenge faced by any memory system is how related experiences should be organized – storing the details of each individual experience preserves potentially valuable details, but is storage-inefficient and hampers generalization, whereas treating all experiences as the same risks ignoring potentially important differences[1]. For instance, learning the common spatial features of different floors in the same building makes it possible to predict the layout of a not-yet-visited floor ("similar to the others"); at the same time, each floor also has unique features, such as the location of a specific colleague's office, that do not generalize. Thus, memory systems need to balance pattern-completion (treating a new observation the same as a previous one) and pattern-separation (keeping similar observations as distinct).

The rodent hippocampus is a model system for studying the neural basis of these processes. Strikingly, the hippocampus can construct statistically independent representations across environments ("global remapping")[2–5] and assigns individual neuron firing fields to locations within an environment in an apparently random fashion[6,7]. Similarly, "engram" studies suggest that the population of neurons allocated to a given experience is determined by a competition based on randomly fluctuating excitability levels among eligible neurons[8]. Although there are also examples of hippocampal cells whose firing properties are tied to a particular stimulus feature (e.g. reward[9]) and therefore transfer across different environments, the received wisdom is that those cells that do change their firing fields between environments or across different regions of the same environment, do so randomly[10].

Remapping studies to date have been limited to within-subject comparisons, but it is possible in principle that what appears random within a single subject in fact obeys a common rule that is shared across subjects. Consider how two related experiences such as running the left (L) and right arms (R) of a T-maze may be encoded in a population of hippocampal neurons. The correlation between L and R activity on a cell-by-cell basis may be zero, but still obey an underlying structure. For instance, cells that tend to fire at the start of L may be more likely to fire at the end of R. If such a rule were to exist, it should be possible to predict, across subjects, what R activity of a target subject looks like, based on (1) that subject's L activity and (2) the relationship between L and R activity found in a different "source" subject. Although there is no way to predict how two different subjects encodes a given experience L (especially when sampling randomly from different numbers of neurons that are not uniquely identifiable across subjects as in e.g. *C. elegans*), the *relationship* between how two different experiences L and R are represented may be conserved across subjects.

Such a *representational geometry* has been demonstrated in a number of brain regions in human cognitive neuroscience studies that use fMRI[11–13], but cross-subject prediction has not yet been applied to ensemble recording data in the rodent hippocampus. If (re)mapping in the rodent hippocampus were to show a shared representational geometry, this would not only challenge a long-held dogma about the randomness of place cell allocation, but potentially also open up novel lines of research that can elucidate the algorithmic basis of memory assignment

and generalization in a wide variety of settings, while creating a bridge between rodent neural data and human fMRI work.

Results

The overall goal of this study is to determine if we can predict how hippocampal place cells in a "target" subject encode a particular experience based on two ingredients: (a) knowledge of how the target subject encodes a distinct but related experience, and (b) how a different, "source" subject encodes the same two experiences.

We operationalize this idea using data from T-maze tasks, in which rats run along the left and right arms of the maze to form the two related experiences under study. Specifically, we can describe hippocampal activity on this task as two subject-specific matrices with time as the horizontal dimension, and neuron as the vertical dimension (Figure 1, leftmost column); one matrix describing the average activity for left trials (L), and another matrix for right trials (R; see Figure S1 for a description of how this input data is obtained). We aim to predict the R matrix in the target subject, based on (a) the target's L matrix and (b) the source's L and R matrices. This neural activity can be visualized as trajectories for L and R trials in a dimensionality-reduced principal component (PCA) space (Step 1 in Figure 1).

Our method for performing cross-subject prediction builds on a procedure from human cognitive neuroscience, "hyperalignment"[11,14], that projects each subject's idiosyncratic neural activity into a common space that minimizes the Euclidean distance between neural activity trajectories. Working in this common space, we can identify the relationship between how the source subject encodes L and R trials, and express it as a transformation matrix ("hypertransform") and apply it to the target subject's L trials to obtain a predicted R trajectory $\widehat{R}$. This predicted trajectory is then projected back to the target-specific neural space to obtain a prediction which is compared to the actual data (Figure 1).

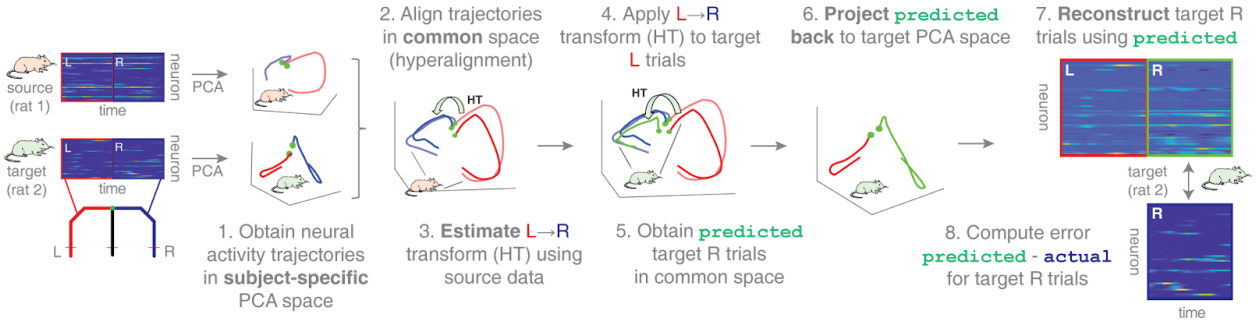

Figure 1: **Workflow for cross-subject prediction of place cell data using the hypertransform**. The objective of the main procedure in this study is to predict place cell activity on the right arm (R) of a T-maze in a "target" subject, based on (a) place cell activity in the left arm (L) in the target subject, and (b) L and R place cell activity in a different, "source" subject. These input data are shown schematically in the leftmost column: both the source and target rats have two matrices each that describe, for each recorded neuron, how its activity varies during left and right trials. Note that

although the number of time bins is the same across subjects, the number of recorded cells may be different. Therefore, the first step of the analysis workflow (step 1 in the panel above) is to apply principal component analysis (PCA), resulting in neural activity trajectories for left and right trials (red and blue, respectively) in each subject's own PCA space. 3 principal components are shown here for display purposes, but in the main analysis 10 PCs were used. Next, these neural activity trajectories are mapped into a common space using a "hyperalignment" procedure that minimizes the Euclidean distance between the trajectories across subjects (step 2, see *Methods* for details). In this common space, a procrustes transformation[15] (M_{12} in step 3) is derived that maps L to R trajectories for the "source" subject (step 3), which can then be applied to the L trajectory of the "target" subject (step 4) to obtain its predicted R trajectory in the common space (step 5). This predicted R trajectory is then projected back to the "target" PCA space using the inverse of the matrix used in step 2 (step 6) and expanded back into the target's original neuron space (step 7). Finally, the predicted R neural activity is compared to the actual R activity to yield an error measure (step 8).

If a given subject encodes L and R trials independently, then it should not be possible to use one subject's neural activity for L and R to predict anything about how another subject encodes R trials based on its L trials. On the other hand, if there is some shared structure between subjects in how L and R trials are encoded, then cross-subject prediction should perform better than chance. Note that is possible for L and R activity in a given subject to appear completely unrelated – that is, there is no correlation between which cells fire for L and R trials – but have the relationship between L and R be completely deterministic across subjects.

To test if there is a shared structure across subjects, we compare the prediction of R trials as described above with various baseline (control) predictions. Specifically, for each source-target pair, we obtain a distribution of chance predictions based on breaking the relationship between L and R trials in the source subject by randomly permuting the rows of the R matrix (see Figure S1 for a schematic of this procedure). Based on this chance distribution, we define three metrics: (1) a z-score of the actually observed error compared to chance, (2) the difference between the actually observed prediction error and the mean of the chance prediction error, and (3) the proportion of chance prediction errors that were lower than the observed error.

We used two different data sets: the first, "Carey" data set[16,17] is from a T-maze where L and R arms were deliberately equipped with distinct surface colors and textures. In contrast, the second, "Gupta" data set[18,19] used a T-maze whose arms had similar surfaces. Starting with the Carey data, we found that the hypertransform (HT) prediction of R trials in the target subject was better than chance overall for all metrics used (Figure 2, top and middle rows; green "HT" bars; z-score: $p < 0.001$ for Wilcoxon signed rank test vs. 0; raw error: $p < 0.001$). Cross-subject prediction of R activity was better than chance even when the R data was withheld entirely from the hyperalignment step (Figure S2, top row; $p < 0.001$ for HT vs. 0), and when L and R activity was expressed as tuning curves in space rather than in time (Figure S2, middle row; $p < 0.001$ for HT vs. 0). These results demonstrate that the relationship between L and R trials is not random across subjects.

To test if the hyperalignment step of the prediction procedure is important, rather than some other part of the workflow, we repeated the analysis with the hyperalignment step left out (i.e. we applied the L-R transform obtained from the source subject's PCA space to the target subject's PCA space, "PCA-only"; see *Methods* for details). The HT-based prediction consistently outperformed the PCA-only prediction for the Carey data (blue "PCA" bars in Figure 2, middle row; $p < 0.001$, binomial test). For the Gupta data, the HT prediction similarly was consistently better than chance ($p < 0.001$ for HT vs. 0); however, unlike the Carey data, this prediction was not different from the PCA-only prediction (Figure 2, bottom row; $p = 0.67$, binomial test).

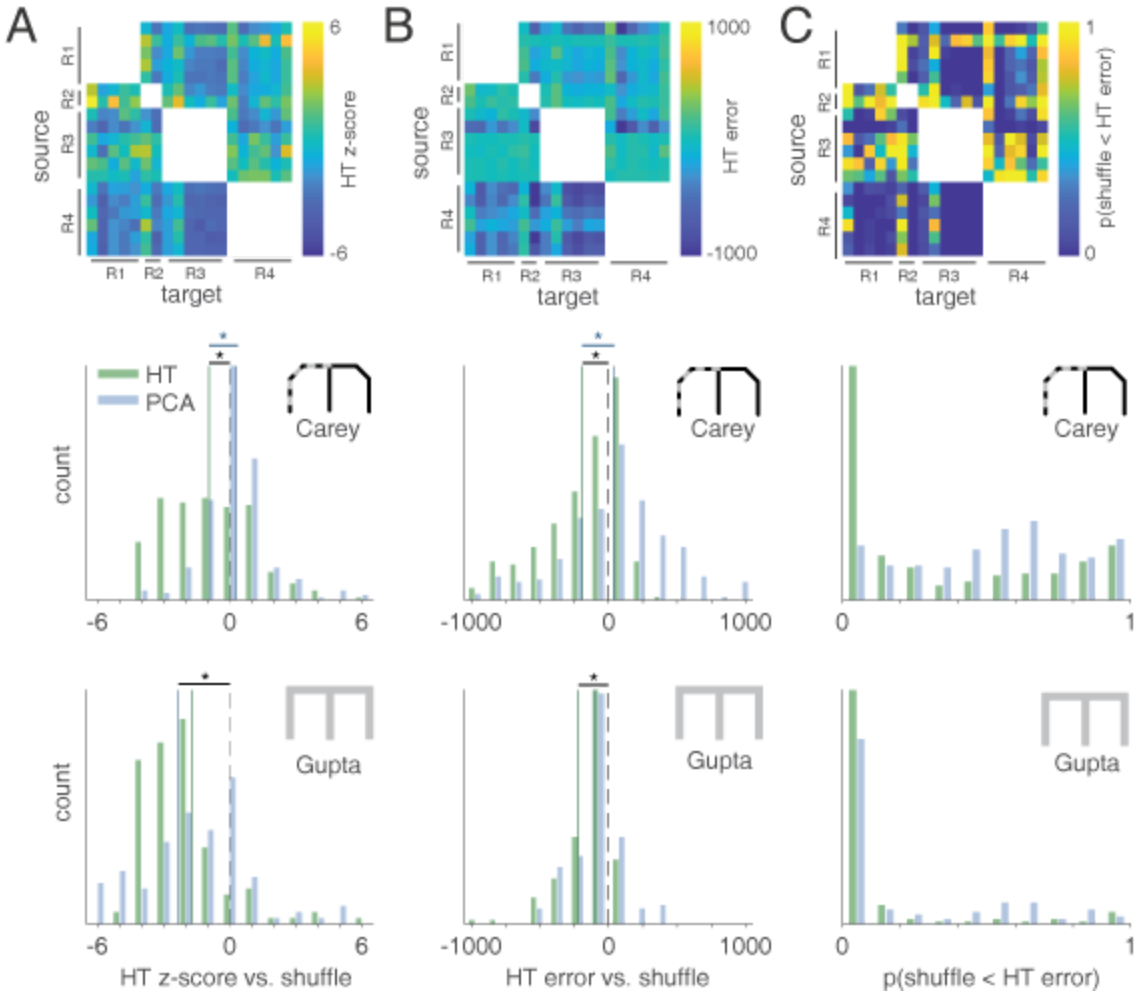

Figure 2: Cross-subject prediction of R trials of a "target" subject based on how a "source" subject encodes L and R trials outperforms prediction based on shuffled source data. For each source-target pair, we computed a z-score of the actually observed error between predicted and actual R trials (based on the hypertransform procedure, "HT") compared to a shuffled distribution in which the R rows of the source subject we randomly permuted. Thus, lower z-scores indicate lower error and therefore better prediction than chance. Across all source-target pairs, this z-scored error varied depending on the pair used (column **A**, top row), but was lower than chance overall, as indicated by a shift in the z-score histogram relative to 0 ("HT" green bars in column **A**, middle row;

–0.94 +/- 0.56, SEM across unique subject pairs, $p < 0.001$ for Wilcoxon signed rank test vs. 0). Cross-subject predictions based on the L-R transform in common space ("hypertransform") outperformed predictions based on the L-R transform in PCA space ("PCA" blue bars; see *Methods* for details; HT < PCA: 66.15% of subject pairs, $p < 0.001$ for binomial test). Next, we applied the same analysis to a different data set ("Gupta", bottom row, in which the L and R maze arms were more similar to each other than in the "Carey" data), and found that although the hypertransform prediction was again significantly better than chance ("HT"; –1.60 +/- 1.00, $p < 0.001$, bottom row), this prediction was not different from PCA-only (HT < PCA, 47.89% of pairs, $p = 0.67$; the reason for this will be revealed in Figure 3 and the main text). Column **B** and **C** use the same layout as column **A**, but using different metrics to describe prediction accuracy. **B** uses the raw error (between predicted and actual R neural activity; lower error/negative indicate better prediction) compared to the mean of the shuffle distribution, and **C** uses the proportion of the shuffle distribution with smaller error than the actually observed error (lower proportions indicate better prediction). For the raw error measure, the HT prediction was better than chance and better than the PCA prediction in the Carey data. (B: –195.23 +/- 88.45, $p < 0.001$ for HT vs. 0; HT < PCA: 66.15%, $p < 0.001$ for HT vs. PCA). In the Gupta data, HT prediction was again better than chance (B: –277.87 +/- 117.14, $p < 0.001$ for HT) but the HT prediction was not better than PCA (B: HT < PCA: 48.59% of pairs, $p = 0.80$).

A possible explanation for the better-than-chance cross-subject prediction is that rats represent experiences in L and R similarly, so that a duplicate of L activity already forms a reasonable prediction of R activity. Such a scenario would be a trivial use of cross-subject prediction. To test if L-R correlations underlie the cross-subject prediction results in Figure 2, we compared cross-subject predictions based on the hypertransform (L-R mapping in common space) with those based on the *identity transform*: a within-subject prediction that simply takes a duplicate of the L trajectory in common space and uses it as the prediction for R. For the Carey data set, the cross-subject HT prediction was significantly better than that based on the identity transform (ID; left panel in Figure 3a; $p < 0.001$, binomial test), demonstrating that the better-than-chance prediction of R is not due to linear correlations with L activity. In contrast, for the Gupta data, the same effect was not observed, suggesting that the better-than-chance cross-subject prediction for this data set (Figure 2, bottom row) may be due to correlations between L and R activity.

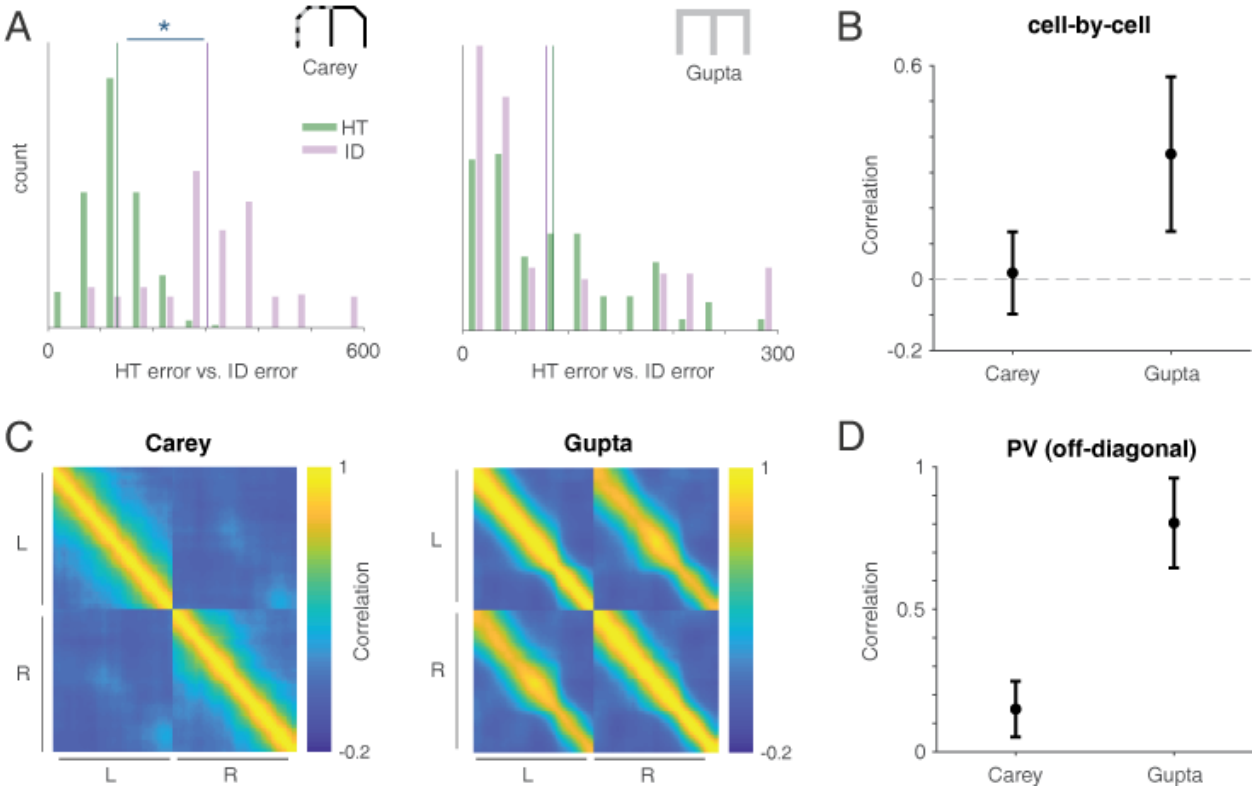

Figure 3: Cross-subject prediction outperforms within-subject prediction only in the absence of cell-by-cell correlations. A: Comparison of cross-subject prediction error ("hypertransform", green bars; HT) with within-subject prediction error ("identity transform"; blue bars, ID) for two different data sets. In the "Carey" data (left panel) the left and right arms of the maze had different texture and color patterns; in the "Gupta" data (right panel) the two maze arms were identical. For the Carey data, cross-subject prediction was significantly better than within-subject prediction (HT < ID: 96.92% of subject pairs, $p < 0.001$ for binomial test) whereas for the Gupta data, this effect was not observed. **B**: Cell-by-cell correlation of firing rates between L and R arms (i.e. row-wise correlations of the L and R matrices in Figure S1), averaged across all cells and subjects, for Carey and Gupta data. Cell activity was significantly more correlated in the Gupta data compared to Carey data (Gupta: r = 0.35 +/- 0.27, SEM across subjects, Carey: r = 0.02 +/- 0.12, Wilcoxon ranksum test, $p < 0.001$). Also, cell activity in Carey data was not significantly different from 0 ($p = 0.44$ for Wilcoxon signed rank test vs. 0). **C**: Population vector (PV) correlations between ensemble activity at each time point and every other time point, i.e column-wise correlations of the L and R activity matrices, averaged across sessions. Both Carey and Gupta data sets show high correlations around the diagonal, indicating an overall autocorrelation in time; however, the Gupta data additionally shows high off-diagonal correlations between L and R which are barely visible in the Carey data. **D**: Quantification of the mean PV correlation between L and R (i.e. the values along the diagonal of the first quadrant in **C**). For Gupta data, this correlation is remarkably high (r = 0.80 +/- 0.16) whereas for Carey data, it is significantly lower (r = 0.15 +/- 0.10, $p < 0.001$ for Wilcoxon rank sum test) and consistent with previous reports of global remapping[2,4]. The above results explain why the HT is not needed for Gupta data to achieve better than chance predictions (see bottom row in Figure 2): L and R activity is sufficiently similar such that the L trajectory alone in either PCA or common space can predict R activity.

To test this idea, we investigated the correlation structure between the L and R firing rate matrices using two different measures: the cell-by-cell (row-wise) correlation, averaged across all cells, and the column-wise population vector (PV) correlation (averaged across sessions, see Figure S1 for schematic; note that in order to compute these measures, putative interneurons were removed from the data; see Figure S3 and Methods). The cell-by-cell firing rate correlations between L and R arms were significantly more correlated in the Gupta data compared to Carey data (Gupta: r = 0.35 +/- 0.27, Carey: r = 0.02 +/- 0.12, SEM across subjects; $p < 0.001$ for Wilcoxon ranksum test; Figure 3b). Similarly, PV correlations in the Gupta data showed high off-diagonal values between L and R, which were barely visible in the Carey data (Figure 3c-d). The low correlation values observed in the Carey data are consistent with those previously reported and characterized as global remapping [2,4], whereas the Gupta correlations are strikingly high, indicating the presence of "symmetric" cells with similar firing patterns on the L and R arms. High L-R correlations in the Gupta data imply that the cross-subject (hypertransform) method cannot outperform the already very good prediction based on within-subject correlations, whereas for the nearly uncorrelated Carey data, there is room for cross-subject prediction to improve.

Importantly, the comparison between the two data sets suggests that the cross-subject prediction on Carey data is not the result of within-subject correlations -- because, if it were, then the HT prediction would be similar to the ID prediction. So, if cross-subject prediction for the Carey data is not simply a consequence of within-subject correlations between L and R, what *is* the prediction based on? In other words, can we identify what features of the L-R relationship are generalizable across subjects without appearing as within-subject correlations between L and R activity? To address this question, we generated synthetic neural activity matrices using 1-D Gaussians with three parameters: time, peak firing rate (FR) and width. Specifically, three simulated data sets captured different potential place cell properties: (1) each neuron has an independent probability of having a firing field on L and R, and all parameters are randomly and independently chosen for L and R (**ind-ind**, top row in Figure 4a), (2) if a neuron has a firing field on L, it does not have a field on R (and vice versa), and the parameters of the field are chosen randomly (**x-or**, second row) and (3) each neuron has an independent probability of having a field on L and R, but if a cell has a field in both, all three parameters are the same (**ind-same-all**, third row). For all these scenarios, we generated synthetic data matching the number of recording sessions and the number of neurons recorded in the Carey data, and applied exactly the same analysis procedure.

The independent (**ind-ind**) simulation serves as a sanity check to verify that our cross-subject prediction procedure cannot exploit shared structure where none exists; as expected, cross-subject prediction was not different from chance in this scenario (Figure 4a, right column; $p = 0.92$ for Wilcoxon signed rank test vs. 0). In contrast, both **x-or** and **ind-same-all** showed better-than-chance cross-subject prediction ($p < 0.001$ for **x-or**, $p < 0.001$ for **ind-same-all**). If the **x-or** or **ind-same-all** rules are potential explanations for better-than-chance predictions in Carey data, we should see uncorrelated cell-by-cell correlations and low PV correlations between L and R in these two data sets as observed in Carey (Figure 3b-d). However, the **x-or**

scenario shows negatively correlated PV correlations inconsistent with the Carey data (Figure 4b-c; r = -0.17). The **ind-same-all** rule shows both high cell-by-cell correlations (Figure 4d; r = 0.25) and high PV correlations (r = 0.41), which is again inconsistent with the Carey data, but more in line with the Gupta data (compare Figure 3b-d). Thus, although both simple rules can support cross-subject prediction, the resulting correlations are inconsistent with those observed in the Carey data.

In further simulations, we separately investigated the role of each parameter of the 1-D gaussian place fields (time, FR and width) as a potential explanation for cross-subject prediction. We created data sets in a manner similar to the **ind-same-all** scenario**,** i.e assigning each neuron an independent probability of having a field on L and R, but with the difference that if a cell has a field in both, only one parameter is the same, and the other two are randomly and independently chosen. For instance, in the **ind-same-time** case, cells with fields on both L and R have independent width and peak firing rates on L and R, but fire at the same (mean) time. Only when the same time is shared across L and R are the cross-subject predictions better than chance (Figure S4; $p < 0.001$). Importantly, this excludes the possibility that correlations in (peak) firing rates between L and R (as suggested by the data in Lee et al. (2019)[20], and by the absence of cross-subject prediction in z-scored data, Figure S3) is the cause for better-than-chance cross-subject prediction.

If these predetermined rules cannot be the source of cross-subject prediction in the Carey data, what can? The hyperalignment procedure we used to derive the L-R transform ("hypertransform") was originally developed to capture so-called *representational geometry*: a shared rule that specifies how *differences* in neural responses to a set of stimuli may be preserved across subjects even though each subject may encode a given stimulus quite differently [11]. To test if such a geometry is consistent with the data, we created another synthetic data set in which the activity on L is simulated by assigning each neuron an independent probability of having a 1-D gaussian place field whose parameters are randomly chosen, and the activity on R is obtained by applying the L-R transform (hypertransform, HT) from the Carey to the simulated L activity (**sim. HT**, bottom row in Figure 4a). Not only did **sim. HT** show significant better-than-chance cross-subject predictions on the simulated data ($p < 0.001$) but the cell-by-cell (r = 0.05) and PV correlations (r = 0.17) were similar to the Carey data. Thus, unlike the simple x-or or same-parameter scenarios, a shared representational geometry is consistent with the correlations observed in the data.

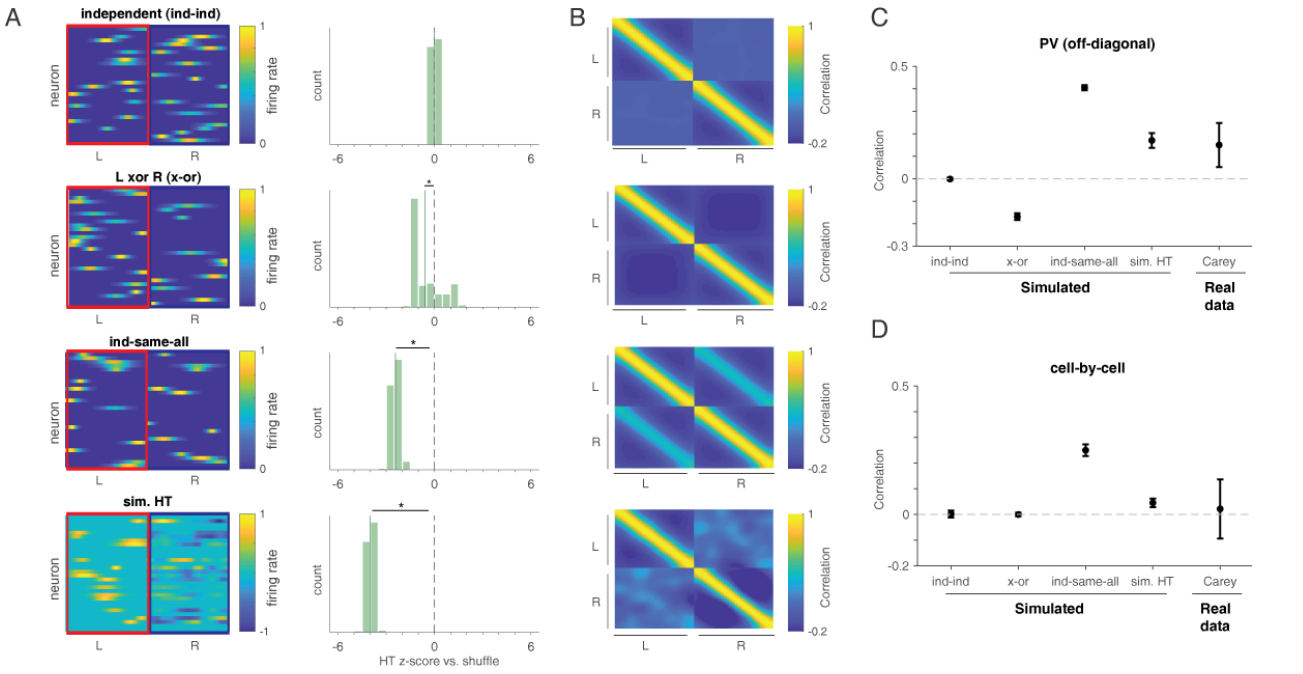

Figure 4: Representational geometry, but not simple rules such as exclusive-or and firing rate correlations, result in cross-subject prediction while being consistent with the data. A: Example L and R activity matrices (left column) and histogram of z-scores of cross-subject prediction compared to the distribution of shuffle predictions (right column; z-scores lower than zero indicate better-than-chance predictions) of four simulated data sets: (1) neurons have a fixed, independent probability (0.5) of having a 1-D Gaussian place field on L and/or R, with the three parameters of time, peak firing rate (FR) and width randomly and independently chosen for L and R (**ind-ind**, top row), (2) neurons only have a field on *either* L or R but not both, and parameters of the field are chosen randomly as in (1) (**x-or**, second row), (3) neurons have a fixed independent probability of having L and R fields as in (1) but with the additional constraint that neurons with both L and R fields must have the same three parameters (**ind-same-all**, third row) and (4) the activity on L is simulated by assigning each neuron an independent probability of having a field whose parameters are randomly chosen, then the activity on R is obtained by applying L-R transform (hypertransform, HT) from real data (Carey) to the simulated L activity (**sim. HT**, last row). As expected, in the **ind-ind** (independent) case, cross-subject prediction is not possible because firing field properties are chosen independently; this can be seen from the histogram of z-scores (prediction vs. shuffle) not being different from 0 (0.0015 +/- 0.03, SEM across unique subject pairs, p = 0.92 for Wilcoxon signed rank test vs. 0). In contrast, **X-or**, **ind-same-all** and **sim. HT** all show better-than-chance cross-subject predictions (-0.59 +/- 0.27 for **x-or**, -2.42 +/- 0.07 for **ind-same-all**, -3.98 +/- 0.07 for **sim. HT,** all $p < 0.001$ for Wilcoxon signed rank test vs. 0), indicating that if there is a non-random L-R relationship in the underlying data, the hypertransform procedure can discover and exploit it. **B**: Population vector (PV; column-wise) correlations between ensemble activity at each time point and every other time point of the L and R activity matrices. Only **ind-same-all** shows high off-diagonal correlations between L and R, resembling the Gupta data set in which L and R arms were identical and firing activity on both arms is highly correlated (compare with Figure 3C). In **x-or**, off-diagonal correlations are slightly negative. **C**: Quantification of the mean PV correlation between L and R (i.e. the values along the diagonal of the first quadrant in **B**). PV correlations between L and R in **ind-ind** are uncorrelated since no prediction of ensemble activity of one time point on L can be made based on the same time of R.

In contrast, L and R are positively correlated in **ind-same-all** (r = 0.41 +/- 0.01) since for every time point in L where there is a field, the same ensemble activity appears at the same time point in R with probability 0.5. **X-or** shows a negative correlation (r = -0.17 +/- 0.02) because for every time point in L where there is a field, the same ensemble activity would deterministically be absent in R, and vice versa. None of these simple rules are consistent with the PV correlation found in the Carey data; however, **sim. HT** does show similar correlations as the data (r = 0.17 +/- 0.03 for **sim. HT** and r = 0.15 +/- 0.10 for Carey). **D**: Cell-by-cell (row-wise) correlations of L and R show that **ind-same-all** is more correlated than the data, whereas **sim. HT** yields similar correlations (r = 0.05 +/- 0.02 for **sim. HT** and r = 0.02 +/- 0.12 for Carey). Thus, taken across panels C and D, simple rules (**x-or** and **ind-same-all**) are inconsistent with the data, but the representational geometry embodied in the **sim. HT** yields correlations that are similar to the data.

Discussion

We have shown that it is possible to predict across subjects, better than chance, how a given experience will be encoded in the hippocampus. In particular, we predict how a "target" subject will represent the right arm of a T-maze (R), given (1) how that same subject represents the left arm (L) of the same maze, and (2) the relationship between L and R activity in a different "source" subject. Control analyses based on within-subject prediction and a comparison of the properties of various simulated data sets with the real data suggests that this cross-subject prediction is unlikely to be the result of trivial relationships such as cells with symmetric firing fields, or simple rules such as exclusive-or. Thus, our results imply that the hippocampal encoding of different locations in space, commonly reported to be random *within* subjects[2–4,6,7], in fact has a shared structure *between* subjects.

An innovative aspect of this work that contrasts with the vast majority of neural recording studies in rodents is the use of cross-subject prediction. This approach is attractive because it provides rigorous, quantifiable tests of how generalizable a given model of neural activity is, and because it can provide insights into what is shared and what is unique between subjects. Using calcium imaging data from hippocampal ensembles, Rubin et al. (2019)[21] showed that the location of one animal could be decoded using a decoder trained on data from another animal. Our approach is similar in that it also uses cross-subject prediction, but addresses a different question in that we seek to predict not the location of the animal, but the tuning curves -- that is, the way different and potentially not-yet-experienced locations will be encoded.

Unlike in the rodent literature, a substantial number of human fMRI and ECoG studies has used cross-subject prediction[11–13,22,23]. Particularly effective are procedures that do not only align across subjects anatomically (e.g. by mapping each subject to a reference brain) but functionally, i.e. by finding structure in how related experiences are represented, even though across subjects the same experience may be represented very differently. Haxby et al. (2011)[11] refer to such shared structure as "representational geometry", an idea consistent with the results reported here.

So what is the shared similarity we have uncovered, and why does it matter? One possibility is suggested by the relationship between grid cells and place cells: Dordek et al. (2016)[24] showed that applying nonnegative principal component analysis to ensemble activity of place cells yields grid cell-like activity patterns. In addition, grid cell firing patterns were known to remain intact but realign linearly (and differently for different subjects) during global remapping[25]. This suggests that what our hypertransform procedure found was a common coding strategy that transforms grid cell activity from one environment to the other, a possibility supported by recent work[26,27]. We provide a novel perspective on this issue by considering shared structure across subjects. More generally, uncovering the ways in which hippocampal activity is non-random can ultimately inform how processes such as generalization and structure learning are realized in neural circuits.

Acknowledgements

We thank Jim Haxby and Feilong Ma for technical advice on the hyperalignment procedure; Youki Tanaka, Caleb Kemere, and the participants of the 2018 Methods in Neuroscience at Dartmouth summer school for discussion; Neil Burgess and Sam Mckenzie for valuable comments and suggestions, and A. David Redish for the use of one of the data sets in this study. This work was supported by HFSP Young Investigator Award RGY0088/2014, NWO VENI grant 863.10.013 and NSF CAREER award IOS-1844935.

Author contributions

HTC wrote computer code and performed data analysis, in part using tools developed by JRM and MvdM. HTC and MvdM jointly conceived the analyses and discussed the results, with suggestions from JRM. HTC and MvdM wrote the paper with comments from JRM.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Figure 1: **Workflow for cross-subject prediction of place cell data using the hypertransform**. The objective of the main procedure in this study is to predict place cell activity on the right arm (R) of a T-maze in a "target" subject, based on (a) place cell activity in the left arm (L) in the target subject, and (b) L and R place cell activity in a different, "source" subject. These input data are shown schematically in the leftmost column: both the source and target rats have two matrices each that describe, for each recorded neuron, how its activity varies during left and right trials. Note that although the number of time bins is the same across subjects, the number of recorded cells may be different. Therefore, the first step of the analysis workflow (step 1 in the panel above) is to apply principal component analysis (PCA), resulting in neural activity trajectories for left and right trials (red and blue, respectively) in each subject's own PCA space. 3 principal components are shown here for display purposes, but in the main analysis 10 PCs were used. Next, these neural activity trajectories are mapped into a common space using a "hyperalignment" procedure that minimizes the Euclidean distance between the trajectories across subjects (step 2, see *Methods* for details). In this common space, a procrustes transformation[15] (M_{12} in step 3) is derived that maps L to R trajectories for the "source" subject (step 3), which can then be applied to the L trajectory of the "target" subject (step 4) to obtain its predicted R trajectory in the common space (step 5). This predicted R trajectory is then projected back to the "target" PCA space using the inverse of the matrix used in step 2 (step 6) and expanded back into the target's original neuron space (step 7). Finally, the predicted R neural activity is compared to the actual R activity to yield an error measure (step 8).

Figure 2: Cross-subject prediction of R trials of a "target" subject based on how a "source" subject encodes L and R trials outperforms prediction based on shuffled source data. For each source-target pair, we computed a z-score of the actually observed error between predicted and actual R trials (based on the hypertransform procedure, "HT") compared to a shuffled distribution in which the R rows of the source subject we randomly permuted. Thus, lower z-scores indicate lower error and therefore better prediction than chance. Across all source-target pairs, this z-scored error varied depending on the pair used (column **A**, top row), but was lower than chance overall, as indicated by a shift in the z-score histogram relative to 0 ("HT" green bars in column **A**, middle row; –0.94 +/- 0.56, SEM across unique subject pairs, $p < 0.001$ for Wilcoxon signed rank test vs. 0). Cross-subject predictions based on the L-R transform in common space ("hypertransform") outperformed predictions based on the L-R transform in PCA space ("PCA" blue bars; see *Methods* for details; HT < PCA: 66.15% of subject pairs, $p < 0.001$ for binomial test). Next, we applied the same analysis to a different data set ("Gupta", bottom row, in which the L and R maze arms were more similar to each other than in the "Carey" data), and found that although the hypertransform prediction was again significantly better than chance ("HT"; –1.60 +/- 1.00, $p < 0.001$, bottom row), this prediction was not different from PCA-only (HT < PCA, 47.89% of pairs, $p = 0.67$; the reason for this will be revealed in Figure 3 and the main text). Column **B** and **C** use the same layout as column **A**, but using different metrics to describe prediction accuracy. **B** uses the raw error (between predicted

and actual R neural activity; lower error/negative indicate better prediction) compared to the mean of the shuffle distribution, and **C** uses the proportion of the shuffle distribution with smaller error than the actually observed error (lower proportions indicate better prediction). For the raw error measure, the HT prediction was better than chance and better than the PCA prediction in the Carey data. (B: –195.23 +/- 88.45, $p < 0.001$ for HT vs. 0; HT < PCA: 66.15%, $p < 0.001$ for HT vs. PCA). In the Gupta data, HT prediction was again better than chance (B: –277.87 +/- 117.14, $p < 0.001$ for HT) but the HT prediction was not better than PCA (B: HT < PCA: 48.59% of pairs, $p = 0.80$).

Figure 3: Cross-subject prediction outperforms within-subject prediction only in the absence of cell-by-cell correlations. A: Comparison of cross-subject prediction error ("hypertransform", green bars; HT) with within-subject prediction error ("identity transform"; blue bars, ID) for two different data sets. In the "Carey" data (left panel) the left and right arms of the maze had different texture and color patterns; in the "Gupta" data (right panel) the two maze arms were identical. For the Carey data, cross-subject prediction was significantly better than within-subject prediction (HT < ID: 96.92% of subject pairs, $p < 0.001$ for binomial test) whereas for the Gupta data, this effect was not observed. **B**: Cell-by-cell correlation of firing rates between L and R arms (i.e. row-wise correlations of the L and R matrices in Figure S1), averaged across all cells and subjects, for Carey and Gupta data. Cell activity was significantly more correlated in the Gupta data compared to Carey data (Gupta: r = 0.35 +/- 0.27, SEM across subjects, Carey: r = 0.02 +/- 0.12, Wilcoxon ranksum test, $p < 0.001$). Also, cell activity in Carey data was not significantly different from 0 ($p = 0.44$ for Wilcoxon signed rank test vs. 0). **C**: Population vector (PV) correlations between ensemble activity at each time point and every other time point, i.e column-wise correlations of the L and R activity matrices, averaged across sessions. Both Carey and Gupta data sets show high correlations around the diagonal, indicating an overall autocorrelation in time; however, the Gupta data additionally shows high off-diagonal correlations between L and R which are barely visible in the Carey data. **D**: Quantification of the mean PV correlation between L and R (i.e. the values along the diagonal of the first quadrant in **C**). For Gupta data, this correlation is remarkably high (r = 0.80 +/- 0.16) whereas for Carey data, it is significantly lower (r = 0.15 +/- 0.10, $p < 0.001$ for Wilcoxon rank sum test) and consistent with previous reports of global remapping[2,4]. The above results explain why the HT is not needed for Gupta data to achieve better than chance predictions (see bottom row in Figure 2): L and R activity is sufficiently similar such that the L trajectory alone in either PCA or common space can predict R activity.

Figure 4: Representational geometry, but not simple rules such as exclusive-or and firing rate correlations, result in cross-subject prediction while being consistent with the data. A: Example L and R activity matrices (left column) and histogram of z-scores of cross-subject prediction compared to the distribution of shuffle predictions (right column; z-scores lower than zero indicate better-than-chance predictions) of four simulated data sets: (1) neurons have a fixed, independent probability (0.5) of having a 1-D Gaussian place field on L and/or R, with the three parameters of time, peak firing rate (FR) and width randomly and independently chosen for L and R (**ind-ind**, top row), (2) neurons only have a field on *either* L or R but not both, and

parameters of the field are chosen randomly as in (1) (**x-or**, second row), (3) neurons have a fixed independent probability of having L and R fields as in (1) but with the additional constraint that neurons with both L and R fields must have the same three parameters (**ind-same-all**, third row) and (4) the activity on L is simulated by assigning each neuron an independent probability of having a field whose parameters are randomly chosen, then the activity on R is obtained by applying L-R transform (hypertransform, HT) from real data (Carey) to the simulated L activity (**sim. HT**, last row). As expected, in the **ind-ind** (independent) case, cross-subject prediction is not possible because firing field properties are chosen independently; this can be seen from the histogram of z-scores (prediction vs. shuffle) not being different from 0 (0.0015 +/- 0.03, SEM across unique subject pairs, $p = 0.92$ for Wilcoxon signed rank test vs. 0). In contrast, **X-or**, **ind-same-all** and **sim. HT** all show better-than-chance cross-subject predictions (-0.59 +/- 0.27 for **x-or**, -2.42 +/- 0.07 for **ind-same-all**, -3.98 +/- 0.07 for **sim. HT,** all $p < 0.001$ for Wilcoxon signed rank test vs. 0), indicating that if there is a non-random L-R relationship in the underlying data, the hypertransform procedure can discover and exploit it. **B**: Population vector (PV; column-wise) correlations between ensemble activity at each time point and every other time point of the L and R activity matrices. Only **ind-same-all** shows high off-diagonal correlations between L and R, resembling the Gupta data set in which L and R arms were identical and firing activity on both arms is highly correlated (compare with Figure 3C). In **x-or**, off-diagonal correlations are slightly negative. **C**: Quantification of the mean PV correlation between L and R (i.e. the values along the diagonal of the first quadrant in **B**). PV correlations between L and R in **ind-ind** are uncorrelated since no prediction of ensemble activity of one time point on L can be made based on the same time of R. In contrast, L and R are positively correlated in **ind-same-all** ($r = 0.41$ +/- 0.01) since for every time point in L where there is a field, the same ensemble activity appears at the same time point in R with probability 0.5. **X-or** shows a negative correlation ($r = -0.17$ +/- 0.02) because for every time point in L where there is a field, the same ensemble activity would deterministically be absent in R, and vice versa. None of these simple rules are consistent with the PV correlation found in the Carey data; however, **sim. HT** does show similar correlations as the data ($r = 0.17$ +/- 0.03 for **sim. HT** and $r = 0.15$ +/- 0.10 for Carey). **D**: Cell-by-cell (row-wise) correlations of L and R show that **ind-same-all** is more correlated than the data, whereas **sim. HT** yields similar correlations ($r = 0.05$ +/- 0.02 for **sim. HT** and $r = 0.02$ +/- 0.12 for Carey). Thus, taken across panels C and D, simple rules (**x-or** and **ind-same-all**) are inconsistent with the data, but the representational geometry embodied in the **sim. HT** yields correlations that are similar to the data.

STAR Methods

Raw data

We used two different data sets containing ensemble recordings of hippocampal CA1 neurons in rats performing T-maze tasks.

The first data set ("Carey") is as described in van der Meer et al. (2017)[16] and Carey et al. (2019)[16,17]. Briefly, male rats (n = 4) performed daily sessions on a T-maze where they had free choice between left and right arms. Rats were alternately food- and water-restricted across days; the left arm resulted in food reward (five 45mg pellets), the right arm resulted in water reward (~0.2 ml sucrose solution). Rats ran 15-20 discrete trials per recording session, with no less than 5 trials for the least preferred choice (left or right). Only sessions with at least 40 simultaneously recorded neurons were included, this left 19 of 24 total sessions for analysis (range: 50 - 178 neurons per session).

The second data set ("Gupta") is as described in Gupta et al. (2010)[18] and Gupta et al. (2012)[18,19]. Briefly, male rats (n = 4) performed daily sessions on a continuous Multiple-T maze with free choice between left and right arms. Food pellet reward (four 45 mg pellets) was available either by choosing left only, right only, or alternating between left and right; which reward schedule was in effect was determined pseudorandomly at the start of daily recording sessions. In addition, the reward schedule switched approximately halfway throughout the session. Only sessions with at least 40 simultaneously recorded neurons were included, this left 14 of 42 total sessions for analysis (range: 41-101 neurons per session).

Both data sets consist of both left (L) and right (R) trials; the analyses in this study are concerned with the relationship between how L and R trials are encoded in hippocampal ensemble activity. To avoid the possibility that neural activity on a common trajectory shared between L and R is the main predictor of L and R relationship, data from the central stem of the maze was excluded.

Data preprocessing

Preparation of input data. Both data sets were preprocessed to obtain two types of neural activity matrices that form the starting point for all analyses (Figure S1). The first and main data type is the Q-matrix, which describes binned firing rate over time for simultaneously recorded neurons *[nNeurons x nTimeBins]* and is used in all main analyses. The second data type is the TC-matrix (place turning curves) matrices of dimension *[nNeurons x nSpaceBins]* for Figure S2 and S3).

Since both data sets contain different numbers of L and R trials within a session, trials were first subsampled so that equal number of L and R trials were used. Next, because trials differed in

length because of variations in running speed, all trials were truncated to the last 2.4 seconds (the median time between passing the choice point and reaching the reward site) of all included sessions.

To obtain Q-matrices for L and R trials, binned firing rate matrices (time bin width: 50 ms) were created for individual trials, smoothed with a window size = 1 s, σ = 50 ms Gaussian kernel, and then averaged across within session.

To obtain TC-matrices for L and R trials, spike firing data with only running speed > 5 cm/s rata were averaged for each place bin (~3 cm per bin) across within session, then smoothed with a window size = 11 bins, σ = 1 bin Gaussian kernel. Only data from the last 41 place bins were included so that only after-choice-point data was used.

Criterion for exclusion of interneurons (some analyses only). Neurons with mean firing rate > 10 Hz across the entire recording session were classified as putative interneurons. These were excluded for the correlation analysis in Figure 3, because otherwise variations in firing rates between putative interneurons and projection neurons would dominate the population vector correlations (described below). We verify that inclusion of interneurons was not required for the main results in Figure S3.

Normalization (some analyses only). Normalization of the input data was conducted by dividing the l^2-norm of each row (neuron) or z-scoring each row of data matrices, *independently* for the L and R parts of the input data matrices. Although it may seem intuitive to normalize each entire row of the input data (i.e. L and R data together), this actually introduces an artificial anticorrelation between the L and R parts of the matrix, such that even on data where no relationship exists between L and R, a relationship is introduced by normalization. Thus, we avoided normalization across L and R when testing cross-subject prediction. Normalization was only used on Q-matrices and TC-matrices with interneurons removed in Figure S3.

Hypertransform analysis procedure

Overview. The overall objective of the hypertransform procedure is to predict R data in a "target" subject based on (1) the target rat L data, and (2) the L and R data of a different "source" rat (see Figure 1 for a complete description and schematic). Each step of the procedure is described in detail below.

Each recording session was used as source and paired with all sessions from all other subjects to form cross-subject source-target pairs (260 unique source-target session pairs for Carey data and 142 pairs for Gupta data).

PCA. After preprocessing the data as described above, principal component analysis (PCA; *svd* function in MATLAB R2018b) was applied to concatenated L and R neural data matrices to reduce to the same dimension because (1) there were unequal numbers of neurons recorded

across sessions, and (2) we want to keep the principal components that capture the most information (variance). Ten principal components (PCs), accounting for approximately 95% of the variance (Figure S2d) were then used to project L and R matrices into neural activity trajectories in each subject's own PCA space.

Hyperalignment. Hyperalignment is procedure that applies a Procrustes transformation: the linear transformation that minimizes the Euclidean distance between input trajectories based on rotation and translation (and scaling, in some implementations; we did not use scaling in this study). This procedure is commonly used to align fMRI activity trajectories in subject-specific spaces into a common representational space (Haxby et al. 2011)[11]. In similar fashion, we aligned L and R neural trajectories in the PCA spaces from a source rat and a target rat into a common space by using the *hyperalign* function in *hypertools*[14] *(matlab version)*. We refer to the transformation from the PCA space of single subject S into the common space as H_S.

Hypertransform. In this common space, a linear (Procrustes) transformation was derived between L and R neural trajectories of the source rat. This transformation (hypertransform, HT) was then applied to the target rat's L trajectory to yield a *predicted* R trajectory. Note that this transformation is not only subject-specific but subject-target-pair-specific since the common space is unique for each pair used.

Projection back into neural space. Next, the inverse of the subject-specific hyperalignment matrix (projecting from PCA to common space) was applied to project the predicted R trajectory back to the target rat's PCA space. Principal components obtained earlier were used to reconstruct the predicted R trajectory in the PCA space into the predicted R data matrix (Q or TC) for the target rat.

Shuffles and associated metrics. The prediction error for a specific source-target pair is calculated by summing squared errors between predicted and actual R data matrices of the target rat. This error is compared against a baseline control by permuting (shuffling) the rows of R data matrix of the source rat (but keeping the L matrix intact; **row shuffles**) and repeating the hypertransform procedure. The rationale for this is that if there is no shared relationship between L and R across subjects, then shuffled predicted errors should not be different from the actual observed predicted error. 1000 shuffles, hence 1000 shuffled predicted error were conducted for a source-target pair, and three metrics were used to compare the actual predicted error against shuffled predicted error: (1) z-scores of actual observed error compared to the distribution of shuffle predictions (2) raw prediction error compared to the mean of the shuffle distribution (middle), and (3) proportion of the shuffle distribution whose error was smaller than actual observed error. For all metrics, lower numbers indicate a non-random relationship between L and R share across subjects, i.e. a better-than-chance cross-subject prediction.

Identity transform. To test if a better-than-chance cross-subject prediction is simply due to the similarity of L and R neural activity within the target rat, a within-subject prediction is obtained by using a duplicate of target rat's L trajectory in the common space as the predicted R trajectory,

i.e. making the L-R transformation equal to the identity transformation. This predicted trajectory was then used to reconstruct the predicted target rat's R data matrix as in the hypertransform procedure.

Variation: withheld data. To exclude the possibility that including data to-be-predicted is the reason why better-than-chance predictions can be achieved, the R data matrix of target rat was withheld, i.e. padded with zeros before the rest of the hypertransform procedure was conducted.

Variation: PCA only. To test whether a relationship between L and R already exists in subject-specific PCA spaces, we modified the hypertransform procedure as follows: instead of aligning neural activity trajectories in the common space through hyperalignment, a L to R linear (Procrustes) transformation was derived from the source rat's PCA space and directly applied to the L trajectory in target rat's PCA space to obtain the predicted R trajectory. This predicted trajectory was then used to reconstruct the predicted target rat's R data matrix as in hypertransform procedure.

Simulations

As a first step towards understanding the possible explanation for cross-subject prediction, simulated neural activity matrices were generated using 1-D Gaussians with three parameters: time, peak firing rate (FR) and width. Several scenarios were created to test the possibility that different potential place cell properties may account for the observed cross-subject prediction:

- **Ind-ind**: Neurons have a fixed, independent probability (0.5) of having a 1-D Gaussian place field on L and/or R, with the three parameters of time, peak firing rate (FR) and width randomly and independently chosen for L and R.
- **X-or**: Neurons only have a field on *either* L or R but not both, and parameters of the field are chosen randomly as ind-ind
- **Ind-same-all**: Neurons have a fixed independent probability of having L and R fields as in **ind-ind** but with the additional constraint that neurons with both L and R fields must have the same three parameters.
- **Sim. HT**: The activity on L is simulated by assigning each neuron an independent probability of having a field whose parameters are randomly chosen, then the activity on R is obtained by applying L-R transform (hypertransform, HT) from real data (Carey) to the simulated L activity.
- **Ind-same-time**: Similar to **ind-same-all**, neurons have a fixed independent probability of having L and R fields as in **ind-ind,** but if when a cell has fields in both, one of the three parameters: **time** is the for L and R, and FR and width are independently chosen.
- **Ind-same-FR**: Similar to **ind-same-all**, neurons have a fixed independent probability of having L and R fields as in **ind-ind,** but if when a cell has fields in both, one of the three parameters: **FR** is the for L and R, and time and width are independently chosen.
- **Ind-same-width**: Similar to **ind-same-all**, neurons have a fixed independent probability of having L and R fields as in **ind-ind,** but if when a cell has fields in both, one of the

three parameters: **width** is the for L and R, and time and width are independently chosen.

For each simulation scenario, we generated a matching number of sessions (19) as in Carey data and a matching number of neurons within each session as a dataset. To avoid the possibility that one particular randomly generated dataset biases the results, 100 datasets were created and all statistics were averaged across datasets to match the real data.

For each of the 100 datasets in **sim. HT**, a randomly chosen session from the Carey data was used to hyperalign with simulated L activity (R activity was padded with zeros first). The hypertransform obtained from the real session was then applied to the simulated L trajectory in the common space to obtain a simulated R trajectory. This simulated trajectory was then used to reconstruct the simulated R activity as in hypertransform procedure.

Correlation analysis of neural and simulated data.

Cell-by-cell (row-wise) correlations. For each cell (a row in the neural data matrix), a correlation coefficient between L and R was computed (see schematic in Figure S1). Whitening noise (matrices of same size as L and R, in which each element is a number sampled from *Uniform*(0, 10^{-5})) was added so that a coefficient can be computed when there is no activity on either L and R trials, which would otherwise result in zero variance.

Population vector (PV; column-wise) correlations. L and R data matrices were (horizontally) concatenated, and a correlation coefficient was computed between each time point (a column) and all other columns of the L and R concatenated matrix (see schematic in Figure S1). Higher off-diagonal values (the values along the diagonal of the first quadrant) of this concatenated matrix indicate higher correlated ensemble neural activity between L and R.

All correlation coefficients were computed using the MATLAB *corrcoef* function.

Supplementary Information

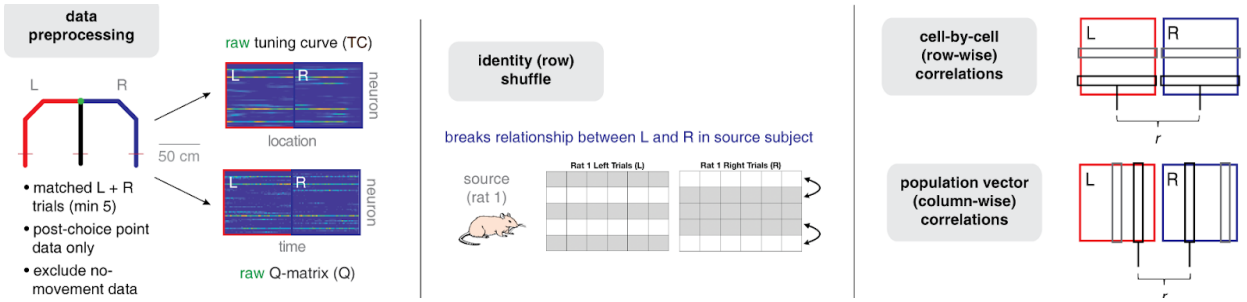

Figure S1: Data preprocessing and analysis schematics. Left: Data preprocessing. Neural ensemble activity for left (L) and right trials (R) in each recording session was shaped into two types of input data matrices: the *Q-matrix*, which describes binned firing rate over time for simultaneously recorded neurons (dimension [nNeurons x nTimeBins]), used in the main analyses, and the *TC-matrix* (spatial tuning curves, dimension [nNeurons x nSpaceBins]). Trials were subsampled to obtain an equal number of L and R trials were used, and truncated to the last 2.4 seconds (the median time taken from the choice point to the end of a trial). Times when the animal was deemed to be stationary were excluded from analysis. **Middle**: Illustration of the shuffling procedure used in the main analysis (Figure 2). To obtain a distribution of chance cross-subject predictions, the analysis steps illustrated in Figure 1 were applied, except that for the "source" subject, the relationship between L and R activity was disrupted by randomly permuting the rows of the R matrix. **Right**: Illustration of the correlation analyses used in Figures 3-4. *Cell-by-cell correlations* are obtained by row-wise correlating L and R activity for each cell, and then averaging across all cells. *Population vector correlations* are obtained by column-wise correlating activity at each time or location with activity at every other time or location. This yields a correlation matrix for each session, which are then averaged across sessions.

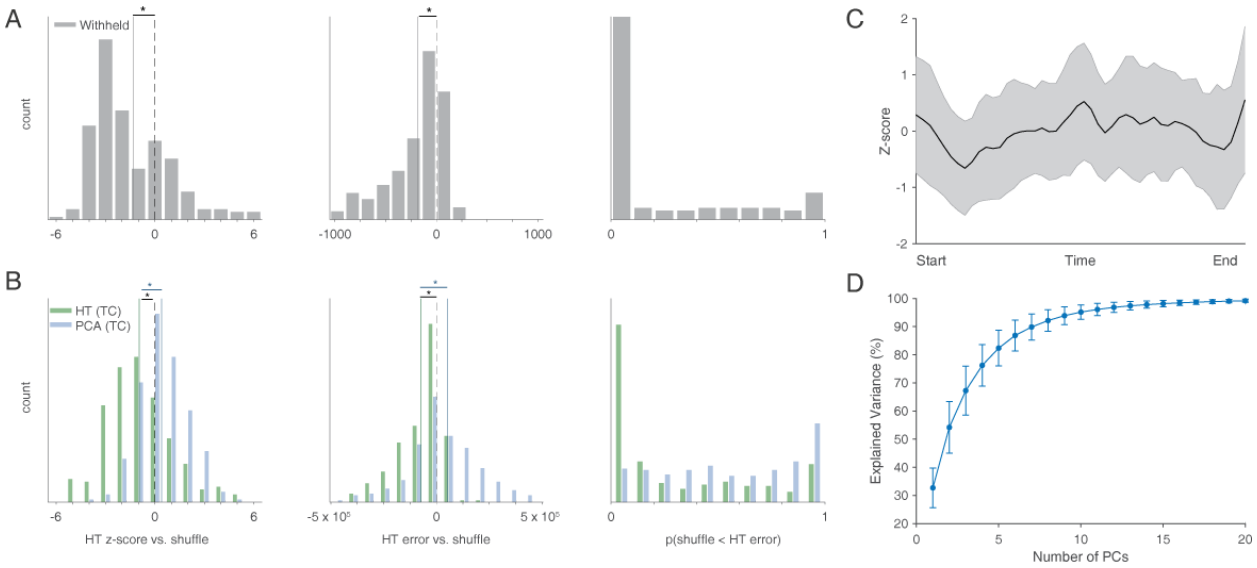

Figure S2: Better-than-chance cross-subject predictions can be observed even when the to-be-predicted target data was withheld, and spatial turning curves were used. A: Histogram of three cross-subject prediction metrics: z-scores of actual observed error compared to the distribution of shuffle predictions (left column), raw prediction error compared to the mean of the shuffle distribution (middle), proportion of the shuffle distribution whose error was smaller than actual observed error (right). For all metrics, lower numbers indicate better cross-subject predictions. The better than chance cross-subject predictions can be observed even when the R activity of the target subject, which is the activity to be predicted, is withheld from the hyperalignment procedure (-1.33 +/- 0.76, SEM across unique subject pairs, $p < 0.001$ for Wilcoxon signed rank test vs. 0). **B**: Histogram of three cross-subject prediction metrics as in **A** for neural activity matrices calculated as a function of locations (turning curves; TC) instead of time (See Figure S1 and *Methods* for details) were used. The cross-subject predictions are significantly better compared to shuffles (-1.07 +/- 0.64, $p < 0.001$ for HT vs. 0) and significantly better than PCA-only (HT < PCA: 69.23% of subject pairs, $p < 0.001$ for binomial test), suggesting time and location yield similar results in our data (see Figure 2). **C**: Z-scores of cross-subject prediction errors (normalized within each session) as a function of time, averaged across sessions. Errors varied across time but did not highlight particular time points, indicating that cross-subject prediction is not disproportionately due to certain time points such as the end of the trial. **D**: Explained variance as a function of the number of principal components (PCs). In our hypertransform procedure, 10 PCs were used, which accounts for ~95% of the variance of data.

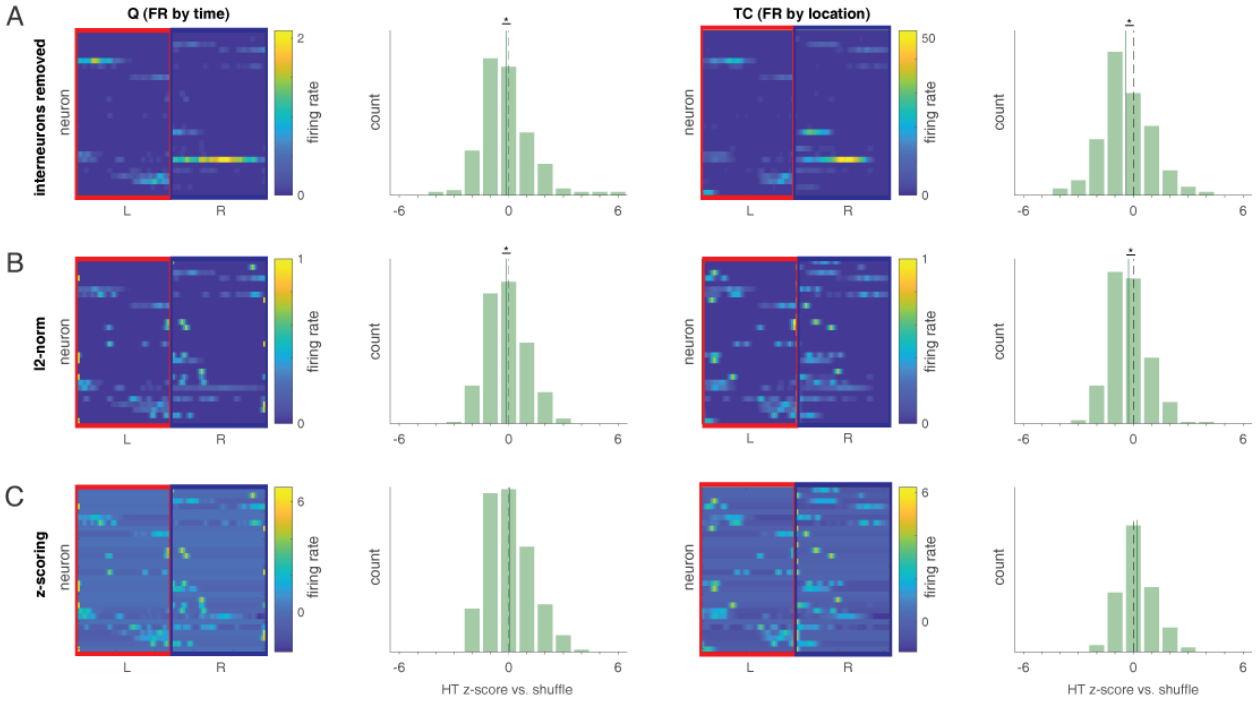

Figure S3: Cross-subject prediction is preserved when removing putative interneurons and normalizing L and R independently by l^2-norm, but abolished when independently z-scoring L and R firing rates. A: Example L and R activity matrices with interneurons (mean firing rate > 10 Hz) removed and corresponding histogram of z-scores of cross-subject prediction compared to the distribution of shuffled predictions (z-scores lower than zero indicate better-than-chance prediction) for both temporal (**Q**) and spatial (**TC**) tuning curves. Both Q and TC show significantly better-than-chance cross-subject predictions (Q: -0.14 +/- 0.42, SEM across unique subject pairs, $p < 0.01$ for Wilcoxon signed rank test vs. 0; TC: -0.40 +/- 0.39, $p < 0.001$ for HT vs. 0), indicating high-firing rate interneurons are not required for cross-subject prediction. **B**: Same layout as A, but L and R activity matrices were obtained by removing interneurons and dividing separately each L and R row by its l^2-norm (vector length). Both Q and TC again show significantly better-than-chance prediction (Q: mean = -0.12 +/- 0.31, $p < 0.05$ for HT vs. 0; TC: -0.26 +/- 0.31, $p < 0.001$ for HT vs. 0). **C**: Same layout as A, but L and R activity matrices were obtained by removing interneurons and z-scoring separately each L and R row. Unlike the l^2-norm case, normalization by z-scoring does not show better-than-chance prediction. Note that for both l^2-norm and z-scoring cases, independently normalizing L and R activity (L separately from R), rather than normalizing L and R together (normalizing the entire row), avoids introducing artifactual anti-correlations between L and R which would be exploited by the cross-subject prediction algorithm even for independent data.

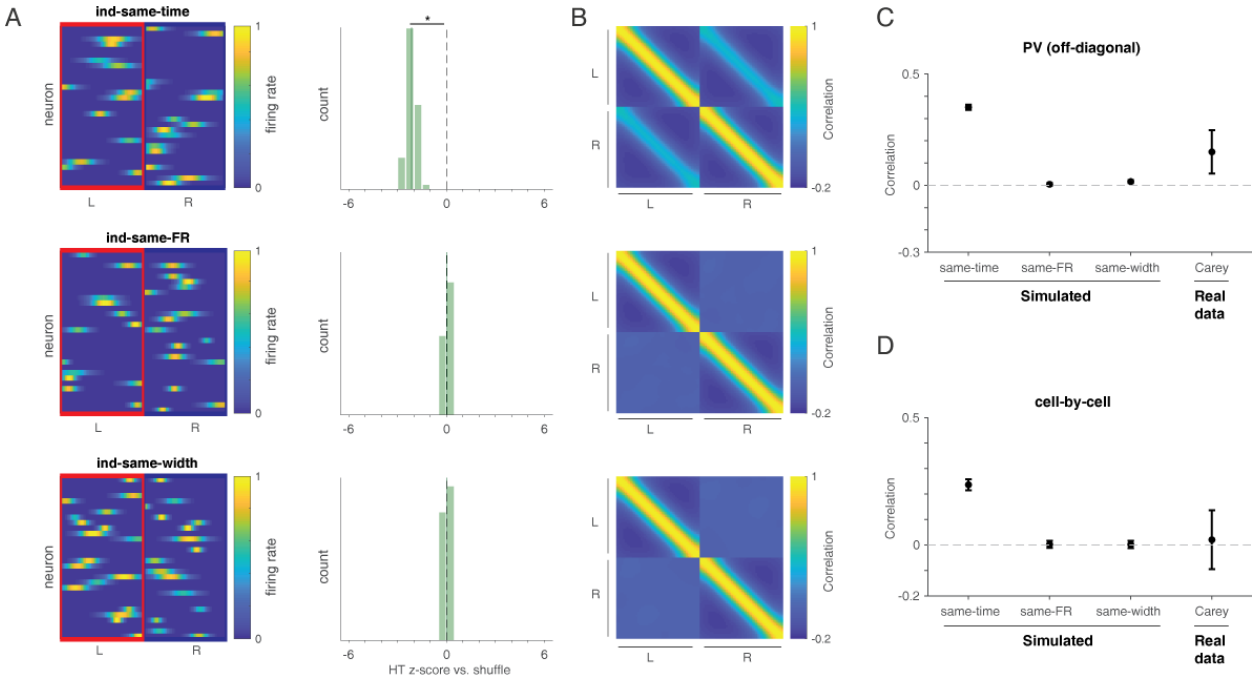

Figure S4: Simulations show that either the same place field firing time, peak firing rate, or place field width alone cannot account for cross-subject prediction in the Carey data. A: Example L and R activity matrices (left column) and histogram of z-scores of cross-subject prediction compared to the distribution of shuffle predictions (z-scores lower than zero indicate better-than-chance predictions; right column) of three simulated data sets. All three data sets are generated by assigning each neuron a fixed, independent probability (0.5) of having a field on L and/or R, but when a cell has fields in both, one of the three parameters of its 1-D Gaussian place field (**ind-same-time**, **ind-same-FR (firing rate)** or **ind-same-width**) is the same for L and R; the other two are randomly and independently chosen. **Ind-same-time** shows better-than-chance cross-subject predictions (-2.14 +/- 0.08, SEM across unique subject pairs, $p < 0.001$ for Wilcoxon signed rank test vs. zero) but **ind-same-FR** and **ind-same-width** do not, indicating cross-subject predictions cannot be better than chance when time points of fields are unrelated between L and R. **B**: Population vector (PV; column-wise) correlations between ensemble activity at each time point and every other time point of the L and R activity matrices. **C**: Quantification of the mean PV correlation between L and R (i.e. the values along the diagonal of the first quadrant in **B**). **Ind-same-time** shows highly positive correlation ($r = 0.35$ +/- 0.01) since for every location of L where there are fields, there would some chance that some (although random) amount of ensemble activity would appear on the same location of R. In contrast, L and R in **ind-same-FR** and **ind-same-width** are nearly uncorrelated since no prediction of ensemble activity of one location on L can be made based on the same location of R (see also **ind-ind i**n Figure 3). The **ind-same-time** case is potentially consistent with the Carey data as measured by PV correlations. **D**: Cell-by-cell (row-wise) correlations of L and R. Here, the **ind-same-time** correlations are higher than in the Carey data ($r = 0.24$ +/- 0.02 vs. $r =$ 0.02 +/- 0.12 for Carey), indicating that this scenario does not accurately capture the source of cross-subject prediction in the data.

References

1. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
2. Leutgeb, S. *et al.* Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science* **309**, 619–623 (2005).
3. Alme, C. B. *et al.* Place cells in the hippocampus: eleven maps for eleven rooms. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18428–18435 (2014).
4. Bostock, E., Muller, R. U. & Kubie, J. L. Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* **1**, 193–205 (1991).
5. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* **7**, 1951–1968 (1987).
6. Redish, A. D. *et al.* Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *J. Neurosci.* **21**, RC134 (2001).
7. Rich, P. D., Liaw, H.-P. & Lee, A. K. Place cells. Large environments reveal the statistical structure governing hippocampal representations. *Science* **345**, 814–817 (2014).
8. Josselyn, S. A. & Frankland, P. W. Memory Allocation: Mechanisms and Function. *Annu. Rev. Neurosci.* **41**, 389–413 (2018).
9. Gauthier, J. L. & Tank, D. W. A Dedicated Population for Reward Coding in the Hippocampus. *Neuron* **99**, 179–193.e7 (2018).
10. Kubie, J. L., Levy, E. R. J. & Fenton, A. A. Is hippocampal remapping the physiological basis for context? *Hippocampus* (2019) doi:10.1002/hipo.23160.
11. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in

human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).

12. Nastase, S. A. *et al.* Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. *Cereb. Cortex* **27**, 4277–4291 (2017).
13. Chen, P.-H. (cameron) *et al.* A Reduced-Dimension fMRI Shared Response Model. in *Advances in Neural Information Processing Systems 28* (eds. Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 460–468 (Curran Associates, Inc., 2015).
14. Heusser, A. C., Ziman, K., Owen, L. L. W. & Manning, J. R. HyperTools: a Python toolbox for gaining geometric insights into high-dimensional data. *J. Mach. Learn. Res.* **18**, 5589–5594 (2017).
15. Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. *Psychometrika* **31**, 1–10 (1966).
16. van der Meer, M. A. A., Carey, A. A. & Tanaka, Y. Optimizing for generalization in the decoding of internally generated activity in the hippocampus. *Hippocampus* **27**, 580–595 (2017).
17. Carey, A. A., Tanaka, Y. & van der Meer, M. A. A. Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nat. Neurosci.* **22**, 1450–1459 (2019).
18. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & David Redish, A. Hippocampal Replay Is Not a Simple Function of Experience. *Neuron* vol. 65 695–705 (2010).
19. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & David Redish, A. Segmentation of spatial experience by hippocampal theta sequences. *Nature Neuroscience* vol. 15 1032–1039 (2012).
20. Lee, J. S., Briguglio, J., Romani, S. & Lee, A. K. The statistical structure of the hippocampal code for space as a function of time, context, and value. *bioRxiv* 615203 (2019) doi:10.1101/615203.

21. Rubin, A. *et al.* Revealing neural correlates of behavior without behavioral measurements. *Nat. Commun.* **10**, 4745 (2019).
22. Simony, E. *et al.* Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* **7**, 12141 (2016).
23. Owen, L. L. W., Heusser, A. C. & Manning, J. R. A Gaussian process model of human electrocorticographic data. doi:10.1101/121020.
24. Dordek, Y., Soudry, D., Meir, R. & Derdikman, D. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *Elife* **5**, e10094 (2016).
25. Fyhn, M., Hafting, T., Treves, A., Moser, M.-B. & Moser, E. I. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* **446**, 190–194 (2007).
26. Mark, S., Moran, R., Parr, T., Kennerley, S. & Behrens, T. Transferring structural knowledge across cognitive maps in humans and models. *bioRxiv* 19 (2019).
27. Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G. & Barry, C. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *bioRxiv* (2019).