

Segmentation of spatial experience by hippocampal theta sequences

Anoopum S Gupta¹⁻³, Matthijs A A van der Meer^{4,5}, David S Touretzky^{2,6} & A David Redish⁷

The encoding and storage of experience by the hippocampus is essential for the formation of episodic memories and the transformation of individual experiences into semantic structures such as maps and schemas. The rodent hippocampus compresses ongoing experience into repeating theta sequences, but the factors determining the content of theta sequences are not understood. Here we first show that the spatial paths represented by theta sequences in rats extend farther in front of the rat during acceleration and higher running speeds and begin farther behind the rat during deceleration. Second, the length of the path is directly related to the length of the theta cycle and the number of gamma cycles in it. Finally, theta sequences represent the environment in segments or 'chunks'. These results imply that information encoded in theta sequences is subject to powerful modulation by behavior and task variables. Furthermore, these findings suggest a potential mechanism for the cognitive 'chunking' of experience.

The hippocampus has a central role in episodic memory¹, navigation² and episodic future thinking^{3,4}. An essential aspect of hippocampal contributions to these behavioral phenomena is the notion that the hippocampus encodes and stores a record of experience⁵. If we are ultimately to uncover the role the hippocampus plays in navigation, imagination and planning processes, it is necessary to understand its information processing characteristics. How is experiential information encoded and subsequently recalled within the hippocampal system? Which features of the encoding process are invariant under, and which ones depend on, the behavioral state of the subject and the structure of the task?

The leading systems-level model for memory encoding in the hippocampus is provided by recordings from hippocampal place cells in freely moving rodents^{2,6}. These cells tend to fire when the animal is in specific locations in an environment, such that a particular trajectory taken by the animal corresponds to a unique sequence of active place cells—the experience to be stored. Notably, the hippocampus compresses sequences of active place cells, continually repeating them at a speed of 5–10× during each cycle of the theta oscillation⁷ as the animal runs^{8–10}. These theta sequences not only provide a repetition of ongoing experience but also bring the spike timing of successively activated place cells into the range conducive to spike timing-dependent plasticity^{8,11–14}. As such, the content of theta sequences provides a unique representational window into what the hippocampus encodes and recalls during active experience.

Previous studies have concentrated on this phenomenon from the perspective of phase precession, examining how the phase of spiking of individual cells and cell assemblies changes^{8,13–16}. Studies of single passes through a place field¹⁷ and of the timing intervals between pairs

of cells spiking within a single theta cycle^{9,18} indicate that the sequence within the cycle itself is more reliable than the phase of any single cell's spiking, suggesting that it would be fruitful to directly examine the relationship between sequences themselves and behavior.

In this study, instead of measuring the relationship of single cell firing to phase of theta, we used the multicell sequence represented within each individual theta cycle to determine the spatial path represented and examined the properties of that path in relation to behavior and task structure. We found that paths represented by theta sequences extended farther ahead of the animal as it left landmarks and accelerated, whereas sequences started farther behind as the animal approached landmarks and decelerated. Even after taking these effects into account, we found that theta sequences more frequently represented the space between adjacent landmarks. Thus, the environment was represented in chunks, implying a task-related segmentation of information processing in hippocampus.

RESULTS

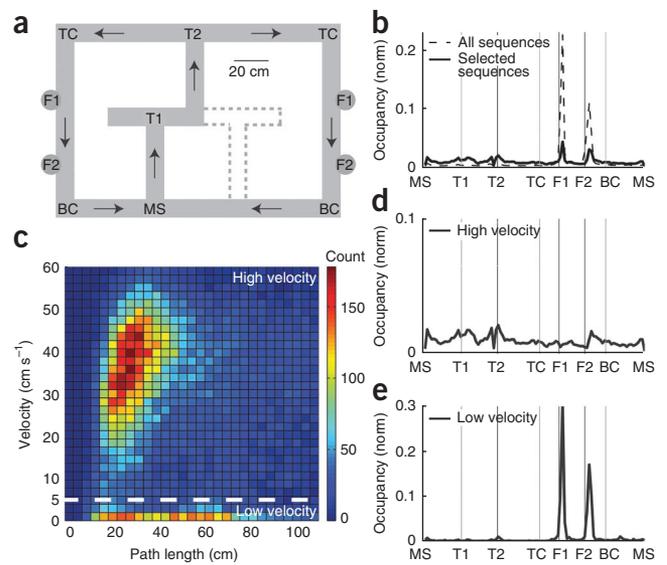
To investigate the relationship between behavior and the paths represented by theta sequences, we recorded neural ensembles from the dorsal CA1 region of hippocampus in six rats trained on a spatial decision task¹⁹ (Fig. 1a). On a given day the rats had to learn whether the rewarded policy was alternation, left turns, or right turns at turn 2 on the maze. The rewarded policy changed approximately halfway through the session. Rats determined both the initial policy and the switched policy quickly¹⁹.

To determine what spatial path was represented by each theta sequence, the beginning and end of each theta cycle were identified (see Online Methods) and the neural activity during each cycle was analyzed.

¹Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. ²Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. ³School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ⁴Department of Biology, University of Waterloo, Waterloo, Ontario, Canada. ⁵Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario, Canada. ⁶Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. ⁷Department of Neuroscience, University of Minnesota, Minneapolis, Minnesota, USA. Correspondence should be addressed to A.D.R. (redish@umn.edu).

Received 29 February; accepted 14 May; published online 17 June 2012; doi:10.1038/nn.3138

Figure 1 Distinct populations of theta sequences on the two-choice T maze. **(a)** The two-choice T maze. The maze had two possible physical configurations, the second indicated by dashed lines. Noteworthy landmarks on the maze include the maze start (MS), turn 1 (T1), turn 2 (T2), top corner (TC), feeder 1 (F1), feeder 2 (F2) and bottom corner (BC). Arrows indicate the direction of maze traversal. **(b)** Overall and selected sequence count over the linearized maze. Overall sequence count over the maze ($n = 618,408$ sequences) is a reflection of the amount of time rats spent at each location on the maze. Selected sequence count ($n = 33,397$ sequences) indicates the number of sequences at each location on the maze that passed the inclusion criteria described in the main text. **(c)** Theta sequence histogram of rat velocity versus path length, defined as the distance (in centimeters) from the decoded start of the sequence of firing within the theta cycle to the decoded end of the sequence (see Online Methods). The color of each pixel indicates the number of theta sequences with a specified path length (x axis) recorded while a rat was running at a particular velocity (y axis). The horizontal dashed line separates low- from high-velocity theta sequences. **(d)** High-velocity sequence count ($n = 29,351$ sequences) over the linearized maze. The number of high-velocity sequences (sequences above horizontal line in **c**) at each location on the maze. **(e)** Low-velocity sequence count ($n = 4,046$ sequences) over the linearized maze. Occupancy is defined as time spent at each location on the maze.



Theta cycles with three or more active cells, a mean sharp wave ripple (SWR) power <1 s.d. above the mean (to exclude awake replay sequences^{20–22}) and theta power >1 s.d. below the mean were included for analysis. Results were similar with thresholds of four or five active cells and with theta power threshold set to the mean theta power or to 2 s.d. below the mean (data not shown). Of 618,408 theta cycles, 168,770 met these criteria. These theta cycles were analyzed for sequential spiking structure, and only cycles with statistically significant sequential structure (defined in Online Methods) were used in this study (33,397 out of 168,770; 19.8%) (**Supplementary Fig. 1**). Using a Bayesian decoding approach, we identified the spatial path represented during each theta cycle (see Online Methods). An alternative, non-Bayesian approach for decoding spatial paths²³ showed qualitatively similar results to those presented here. Theta sequences occurred at all locations on the maze (**Fig. 1b**). As expected, rats spent the most time at feeder locations, and a smaller proportion of theta cycles were significant at feeder locations compared with other locations on the maze.

From the start and end locations of each represented path and the rat's location, we calculated three key distances: the distance from the rat's location to the end of the path (ahead length), the distance from the start of the path to the rat's location (behind length) and the total length of the path (path length). We examined the relationship between rat velocity and path length by creating a two-dimensional histogram of velocity as a function of path length. The theta sequences naturally grouped into high- and low-velocity families (**Fig. 1c**). The high-velocity sequences were evenly distributed across the maze (**Fig. 1d**), whereas the low-velocity sequences were preferentially observed when rats were at the feeder locations and the choice point (**Fig. 1e**). For the remainder of the analyses presented here, we focus on high-velocity theta sequences (29,351 sequences), although all results are qualitatively unchanged by the inclusion of low-velocity sequences.

Ahead and behind representations

We observed that some sequences represented a spatial path extending farther ahead of the rat, whereas other sequences represented paths beginning more behind the rat's location. Ahead sequences were observed at the choice point (**Fig. 2**), potentially analogous to the forward sweeps previously observed in CA3 while rats were paused and deliberating at a choice point²⁴. However, we also observed ahead-extending sequences at other locations on the

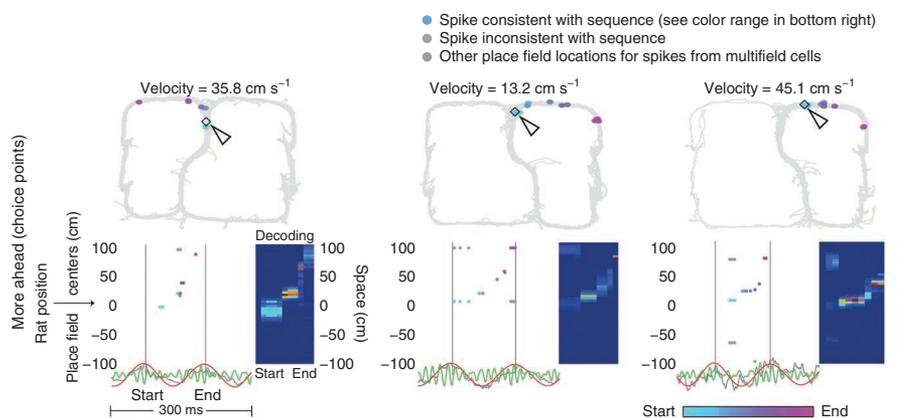
maze (**Fig. 3**). Examples of ahead and behind representations at various locations on the maze are shown in **Figures 2** and **3**. Ahead sequences tended to occur as rats left physical landmarks on the maze. Behind sequences tended to occur as rats approached landmarks. In between sequences tended to occur during journeys between those landmarks.

Ahead and behind length are inversely related

Ahead, behind and overall path length define the spatial information represented by each given theta cycle. Ahead length measures how far the multicell sequence extends into the space ahead of the rat, behind length measures how far behind the rat the multicell sequence begins and path length measures the distance covered by the entire sequence. (see Online Methods.) To determine the relationships among these three lengths, we plotted each mean length with respect to the others. As path length increased, both ahead and behind length increased (**Fig. 4**), as expected, because the path length of each theta sequence is the sum of its ahead and behind lengths. On the other hand, as ahead length increased, behind length decreased. Thus, ahead length and behind length were inversely related; spatial paths that extended farther forward did not begin as far back and vice versa.

To determine how quickly prospective and retrospective representations changed on this task, we measured the changes in ahead and behind length as a function of theta-cycle separation. The ahead and behind lengths of individual theta cycles showed similarities with the preceding and following theta cycles; however, they could also shift substantially from one theta cycle to the next. For adjacent theta cycles, 63% had ahead-length differences <10 cm. The average change in the distance represented by a theta sequence ahead or behind the rat's location increased from successive theta cycles to a separation of about seven theta cycles and then leveled off (**Fig. 4d**). Because theta sequences generally represent positions near the rat, one would expect adjacent theta sequences to represent similar information; this analysis shows that the prospective or retrospective nature of adjacent theta cycles were more similar than those of nonadjacent theta cycles. However, examination of the distribution of changes in ahead and behind length between adjacent theta cycles showed that there were cases in which large differences occurred, which implies that the hippocampal representation can quickly shift from more behind to more ahead of the rat or from more ahead to more behind. These results may explain the improvement in observations looking at phase

Figure 2 Examples of ahead sequences while rats were located at the choice point. Top, the rat's average velocity over the theta cycle. Each place field center represented by a spike in the sequence (colored point in corresponding bottom panel) is plotted on the two-dimensional maze at the location of its two-dimensional place field center. The arrow shows the rat's location. Bottom left, spikes plotted by place field center location (space) relative to the rat's position (along either a left or right loop of the maze) over a single theta cycle (see Online Methods). LFPs filtered between 6 Hz and 12 Hz (red), 40 Hz and 100 Hz (green) and unfiltered (gray) are plotted below. Colored points indicate spikes that contribute positively to the sequence score according to the sequence detection algorithm (for visualization purposes only). The color of the spike indicates its relative time within the sequence (light blue, early; light purple, late). For cells with multiple place fields, gray points are plotted at every place field center belonging to the cell (colored points occupy the place field center that contributes maximally to the score). Spikes that do not contribute positively to the sequence score are also plotted in gray. Bottom right, the Bayesian decoded spatial probability distributions computed over the theta cycle (see Online Methods). Red indicates high probability, blue low probability. These examples likely correspond to the previously reported 'sweeps'²⁴.



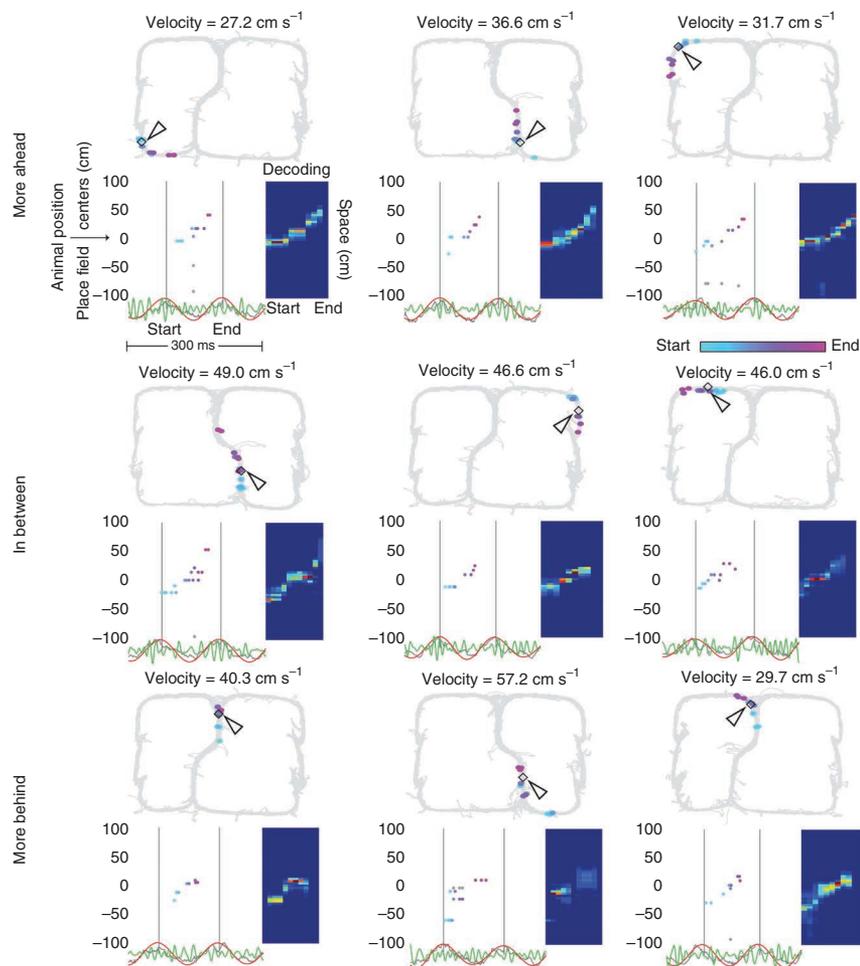
precession on a trial-by-trial basis¹⁷ and the more reliable timing of cell pairs and sequences within a theta cycle than what would be observed if cells phase precessed independently⁹.

Theta period and gamma cycles increase with path length

As path length increased, we observed that the theta period and the number of gamma cycles per theta period also increased (Fig. 5a,b).

Using the previously reported relationship between theta period and velocity (ref. 2 and Fig. 5c) and the observed relationship between velocity and path length (Fig. 1c), we determined the relationship between theta period and path length if the theta period were purely a function of velocity and compared it with the observed relationship (Fig. 5a). The relationship between velocity and theta period did not account for the increase in theta period with increasing path length (Fig. 5a). Similarly, the number of gamma cycles increased with increasing path length (Fig. 5b). Thus, longer theta periods and more gamma cycles per theta cycle were associated with longer paths represented by theta sequences. Longer theta periods were also associated with more active place cells, more spikes and more gamma cycles (Fig. 5d). Although it is possible that the increased cellular activity during longer theta cycles contributed to the representation of longer paths, it should be noted that there was only a slight increase in distance traveled by the rats with increasing theta period, and controlling for velocity does not affect the relationships shown in Figure 5a,b.

Figure 3 Examples of ahead and behind sequences. Top, sequences representing paths more ahead of the rat while rats were at various locations on the maze. Middle, sequences representing paths that are behind and ahead of the rat. Bottom, paths more behind the rat. See Figure 2 caption for details.



Average ahead and behind length over space

To determine whether theta sequences preferentially represented space ahead or behind the rat at certain locations on the maze, the average ahead length and average behind length were plotted as a function of linearized location on the maze (Fig. 6). We observed clear peaks and troughs in average ahead length over space and average behind length over space (Fig. 6a,c). Furthermore, average ahead

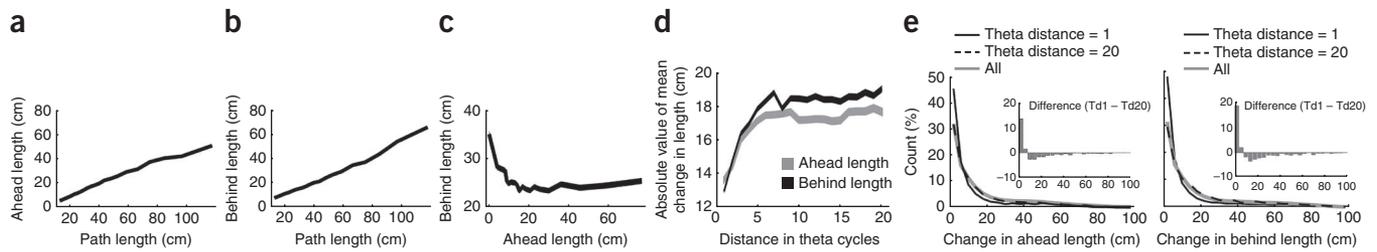


Figure 4 Characteristics of theta sequence ahead and behind length. **(a)** Mean ahead length as a function of path length. **(b)** Mean behind length as a function of path length. **(c)** Mean behind length as a function of ahead length. Note the inverse relationship between ahead length and behind length. **(d)** Ahead and behind length changes as a function of theta cycle separation. The absolute value of the average change in both ahead and behind length increased from successive theta cycles (beginning with 1 on the x axis) to a separation of about seven theta cycles and then leveled off. This indicates that the spaces mapped by neural activity in adjacent theta cycles are more similar than the spaces mapped in nonadjacent cycles. **(e)** Adjacent sequences tend to have related ahead and behind lengths but also occasionally include lengths that are substantially different. Sequences that are more separated (twenty theta cycles shown) tend to have larger changes in ahead and behind length. The subtraction between these two distributions is shown in an inset. Also shown is the distribution of all differences, ignoring the theta cycle distance between the two cycles. Sequences separated by 20 cycles approach random chance, whereas sequences separated by only a single theta cycle are more likely to be similar to each other. These analyses included 29,351 theta cycles. Line width in **a–d** indicate s.e.m. Note that some y axes have non-zero origins.

length and average behind length were anticorrelated over space ($r = -0.34$, $P < 0.05$), with ahead length peaks aligning with behind length troughs and vice versa. This observation is consistent with, but not a necessary result of, the inverse relationship between ahead and behind length presented above (Fig. 4). Thus, at certain locations on the maze, the hippocampus represented more space ahead of the rat and less space behind, and at other locations on the maze, the hippocampus represented more space behind the rat and less space ahead.

The peaks of ahead length and peaks of behind length had a relationship to physical landmarks on the maze (turns and feeders). Ahead length peaks were generally observed just after landmarks (for example, maze start, turn 2 (the choice point), feeder 1, bottom corner) and behind length peaks were observed just before landmarks. Thus, as rats entered a segment of the maze defined by two physical landmarks, the spatial representation was more ahead of the rat; as they exited the segment, the representation was more behind the rat. In effect, the path representations adjusted so as to focus on the segment of space between the two landmarks.

Spatial distribution of represented paths

The alternating pattern of ahead and behind representations resulted in a sequence of segment representations, as reflected in the spatial distribution of represented paths (Fig. 6b). The distribution demonstrated that paths represented by theta sequences preferentially occurred between landmarks on the maze, instead of across them. The distribution shown was normalized such that the paths represented

at each location on the maze contributed equally to the distribution (occupancy normalization). This ensured that factors such as the rat consistently spending more time at a particular location on the maze did not misleadingly influence the distribution around that point. This distribution was robust across rats (Supplementary Fig. 2).

The observed distribution was compared to a distribution created by a shuffling procedure. Keeping the rat location for each theta cycle constant, we shuffled the decoded paths and randomly assigned them to the rat locations. This was done 100 times, and the average spatial distribution was plotted along with the observed distribution (Fig. 6b). As would be expected from the occupancy normalization, the shuffled distribution was relatively flat, showing that the observed spatial distribution of paths was a product of specific sequences occurring at particular locations on the maze. The shuffling procedure was also done for the average ahead and behind length over space plot (Fig. 6a,c), and, as expected, the shuffled distribution was flat.

Hippocampal cells show spatially localized firing (place fields)^{2,6}. Place fields were defined as contiguous linear firing of a given cell (Online Methods). We compared the spatial distribution of represented paths (and ahead and behind lengths over space) shown in Figure 6 with the place field density, linearized place field length and the rat's velocity. Neither place field density, place field length nor average velocity over space matched the shape of the path distribution or ahead and behind length (Supplementary Fig. 3). Place field length, plotted by location on the maze, showed some peaks in between landmarks; however, the overall shape and magnitude was not

Figure 5 Theta period and gamma cycles vary as functions of path length. **(a–c)** The thick black curves show mean theta period **(a)** or mean number of gamma cycles **(b)** (y axis) with increasing path length (x axis). The gray curves were generated by multiplying the average velocity versus path length relationship (Fig. 8b) by the theta period versus velocity relationship or gamma cycle versus velocity relationship **(c)**. Thus, the gray curve is the expected relationship if the theta period **(a)** or the number of gamma cycles **(b)** were merely a function of the relationship between theta period and velocity and path length and velocity. **(d)** Mean spike count, cell count, rat distance traveled and gamma cycles as functions of theta period. This analysis included 29,351 theta sequences. Line width indicates s.e.m. for **a–d**. Note that some y axes have non-zero origins.

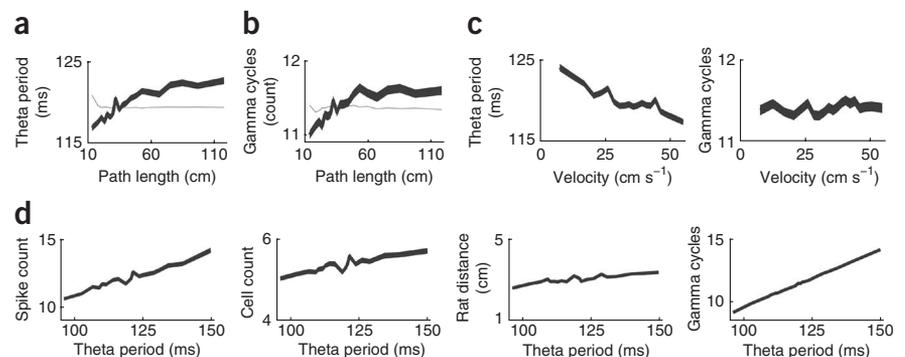
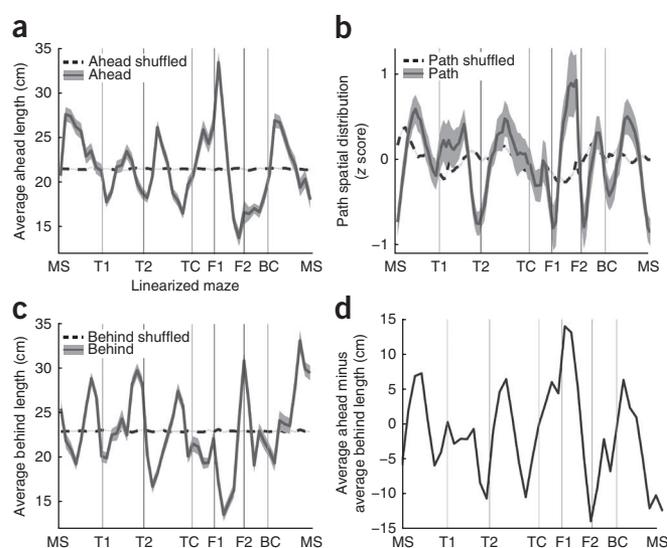


Figure 6 Ahead length, behind length and path distribution vary according to landmarks on the maze. (a) Ahead length as a function of location on maze. Peaks indicate locations at which the represented path extended farther ahead of the rat, on average. Troughs indicate locations where the represented path ended closer to the rat's location. (b) Density of represented paths as a function of location on maze. Peaks indicate regions of the maze that were overrepresented by theta sequences. The y axis is the average z score (z score computed over each session). (c) Behind length as a function of location on maze. Peaks indicate locations at which the represented path began farther behind the rat, on average. Troughs indicate locations where the represented path began closer to the rat's location. (d) Average ahead length minus average behind length. This plot demonstrates the alternating pattern of representations that are more ahead and more behind the rat, as a function of location on the maze. This analysis included 29,351 theta sequences. Line width indicates s.e.m. Controls in a–c include a curve produced by randomly reassigning theta sequences with different rat locations. This shuffling procedure was done one hundred times, and the average shuffled distribution is plotted alongside the real data. Note that some y axes have non-zero origins.

consistent with the ahead, behind and spatial path distributions. For example, peaks in the ahead and behind distributions between the maze start and turn 1 had peak-to-trough distances of ~11 cm and 10 cm, respectively (Fig. 6a,c), whereas the peak-to-trough distance of the place field length distribution was approximately half that distance (6 cm; Supplementary Fig. 3). Although some changes in ahead and behind length occurred at physical turns on the maze (such as the second choice point), other changes occurred on straight paths, without corresponding changes in orientation. For example, a large peak was seen in the ahead distribution just after the first feeder, and a large peak was seen in the behind distribution just at the second feeder, leading to an increased likelihood of representing the between-feeder zone, in the absence of a peak in the place field length distribution.

To further examine the representation of spatial paths as a function of location, we calculated the distribution of positions represented by all theta cycles occurring at that location. This produced a two-dimensional plot with each column representing the average region of space covered by theta sequences occurring at each maze location. Taking the circular mean of the locations covered by this plot produced small deviations above the $x = y$ line where rats left landmarks and below the $x = y$ line where they approached them (Fig. 7a). Deviations above the line indicate maze locations where representations were more ahead of the rat, and deviations below the line indicate locations where representations were more behind the rat.

These small deviations were amplified by creating the full joint probability distribution plot (Fig. 7b, which measures the probability that the rat is at location x and the decoded location is y given the sequence S , calculated by squaring the matrix in Fig. 7a)



and by creating the correlation plot (Fig. 7c, which correlates each column of Fig. 7a with each column in Fig. 7a). The correlation plot is analogous to one used previously²⁵, except that the correlation here was carried out on the decoding relationship rather than the raw population spiking vectors.

Given that theta sequences over-represented the space between landmarks and occur at a fast time scale, one would expect to see greater similarity in the population of cells firing in the space between landmarks, compared with the space across landmarks. This was tested by correlating the average population activity vectors at each location on the maze²⁵. This approach did in fact show increased similarity of the average spatial population vectors in the space between landmarks (Supplementary Fig. 4). However, the segmentation of space by theta sequences (Fig. 7) indicates that not only was segmentation occurring when considering the average cellular activity over many experiences, but it occurred during each theta cycle, at a time scale that allowed for spike-timing dependent plasticity.

In summary, the spatial distribution of paths represented by theta sequences reflected an over-representation of certain segments or 'chunks' of the environment that were bounded by physical landmarks on the maze.

Lengths showed a complex relationship with acceleration

Whereas the spatial distribution of paths was only weakly visible in measures of place field density, place field length and rat velocity (see Supplementary Fig. 3), it did show striking similarities with rat acceleration (compare Fig. 6b with Supplementary Fig. 5a). The peaks and troughs of acceleration were generally aligned with peaks in ahead length and peaks in behind length, respectively (Fig. 6a,c), and ahead length

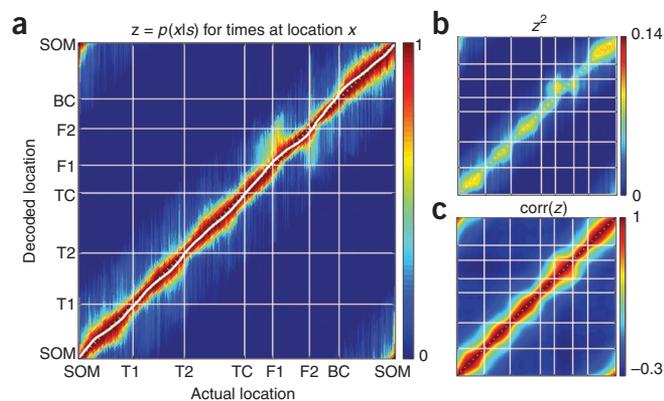


Figure 7 Segmented representation of space by theta sequences. (a) Average decoded spatial path during a theta sequence as a function of the rat's location on the maze. Each column corresponds to the average decoded path when the rat was at a particular location on the maze. Thus, each column corresponds to the mean $p(x|s)$, where x is the position along the maze and s is the spike sequence. The solid white line shows the weighted circular mean of each column. Dotted white line indicates the $x = y$ line. (b) The chunking process can be seen in the pinches in spatial correlation between chunks (at landmark locations marked by intersections of white lines) found when we calculated the full joint probability by squaring the decoding matrix or by calculating the cross-spatial correlations of the decoding matrix (c). This analysis included 29,351 theta sequences.

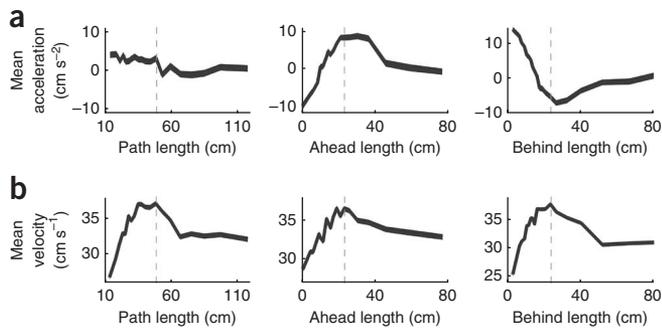


Figure 8 Relationships between rat behavior and the represented path. **(a)** Relationship between acceleration and path, ahead and behind lengths. **(b)** Relationship between velocity and path, ahead and behind lengths. Dashed lines in **a** and **b** indicate the peak of the velocity versus length relationships. As seen in **a**, acceleration is relatively constant with increasing path length. However, acceleration and ahead length are positively correlated, and acceleration and behind length are negatively correlated for shorter sequences (left of dashed line). This relationship changes for longer sequences (right of dashed line). Similarly, in **b**, velocity is positively correlated with path, ahead and behind length for shorter sequences, but the relationship changes for longer sequences. This analysis included 29,351 theta sequences. Line width indicates s.e.m. Note that some y axes have non-zero origins.

was positively correlated ($r = 0.41$, $P = 0.016$), and behind length negatively correlated ($r = -0.74$, $P < 0.01$), with acceleration over space.

When looking at acceleration as a function of ahead and behind length, for shorter length sequences, as ahead length increased, acceleration also increased; conversely, as behind length increased, acceleration decreased (Fig. 8a). In contrast, acceleration remained relatively constant with increasing path length. Although these relationships were strong for shorter ahead and behind lengths, they leveled off to zero acceleration for longer lengths. This peak and the subsequent drop-off were also seen in the relationship between velocity and path, ahead and behind lengths (Fig. 8b); therefore, the peak of the velocity versus length distributions were identified and marked on both the acceleration and velocity plots. Ahead length and behind length for shorter sequences (70% of all sequences) showed a linear relationship with acceleration, but they were more variable as a function of acceleration for longer sequences.

To determine whether the relationship between ahead and behind length and acceleration could account for the spatial distribution of represented paths, we built a model data set using acceleration and deceleration within each theta cycle to create ahead and behind representations. Using the actual behavioral data and the observed relationship between acceleration and ahead and behind length, we assigned each theta cycle an ahead and behind length on the basis of the rat's acceleration during the cycle. We then plotted the spatial distribution of represented paths on the basis of this model and found that it did not match the segmented nature of the spatial distribution from the actual data (compare Fig. 6b with Supplementary Fig. 5b). This difference was not unexpected given the complex relationship between ahead and behind length and acceleration shown in Figure 8a.

Lengths showed a complex relationship with velocity

To determine the relationship between rat velocity and theta sequence lengths, we plotted velocity as a function of path, ahead and behind lengths (Fig. 8b). As each of the three lengths increased, there was initially an increase in velocity and subsequently a drop off. For shorter theta sequences, path, ahead and behind length all increased with increasing velocity; however, for longer sequences, velocity decreased

or was relatively constant with changes in path, ahead and behind length. Thus, the relationship between velocity and path length cannot be completely explained by a greater sampling of place fields at higher speeds. Changes in cell firing with velocity also cannot explain the ahead-shifted and behind-shifted representations.

This observation and the presence of low- and high-velocity theta sequences presented earlier (Fig. 1c) indicate that there was not a simple relationship between velocity and the paths represented during theta sequences. Instances of both short and long paths occurred while rats were stationary or moving at low velocities in the environment, as previously observed²⁴. Sequences also occurred as rats were moving at higher velocities. These sequences could be divided into a set of short-path sequences that increased in length with increasing rat velocity similar to those seen by others^{16,26} and a set of long-path sequences that decreased or remained constant in length with increasing velocity. It should be noted that theta, beta, gamma and SWR power were relatively constant with increasing path, ahead and behind length (Supplementary Fig. 6); thus, the local field potential (LFP) profile did not change as the represented path increased in length.

DISCUSSION

Hippocampal neural activity during each theta cycle sequentially represents a short path that begins just behind the animal and ends just in front of the animal^{8,10,13}; however, we found considerable variability in the actual paths represented on each theta cycle, comparable to the differences seen on each phase-precession pass¹⁷. By examining each theta cycle independently, we found interesting relationships between the spatial representations and velocity, acceleration and physical landmarks on the maze. Most notably, we found that theta sequences represented the maze in segments. Consistent with this, we found that more space was represented ahead of the rat as it left physical landmarks on the maze, and more space behind the rat as it approached those same physical landmarks, leading to a segmentation of the spatial information. These results have implications for how the rat might organize its spatial and task knowledge.

Phase precession and neural sequences

As an animal passes through a cell's place field toward a goal, the cell initially fires late in the theta cycle and then gradually fires earlier and earlier in the theta cycle as the animal traverses the place field^{13,15}. At the population level, phase precession implies that compressed sequences of place cell activity emerge during each theta cycle^{8-14,16,26}.

As noted in the introduction, evidence suggests that the compressed sequences are actually occurring independently on each theta cycle^{9,27}. Our analyses exploit this and differ from previous studies by taking a dynamic view of the theta sequences; rather than averaging across passes, trials or cells, we calculated the path represented by each theta cycle sequence individually and then examined the distribution of properties of the individual sequences.

These sequences included both a set of sequences that were observed to represent longer spatial paths with increasing velocity (consistent with recent reports^{16,26}) and a set of sequences that represented a variety of path lengths independent of velocity. This second set tended to occur while rats were paused in the environment, consistent with previously observed sweep sequences in CA3 (ref. 24), but this second set of sequences also occurred during movement. In general, we found a more complex relationship between the represented path and the rat's behavior than have been observed in previous studies^{9,10,14,16,24,26}. For example, long ahead representations were previously observed while animals were paused at a choice point²⁴.

In contrast, we observed that these longer sequences occurred at other locations on the maze as well, and they did not occur only when rats were paused. These longer sequences sometimes represented more space ahead of the rat and at other times represented more space behind the rat. Our analyses found that representations can start farther behind the rat or end farther ahead of the rat on particular theta cycles. Whereas adjacent theta cycle sequences are often very similar, they can on occasion differ substantially from theta cycle to theta cycle. These results indicated that theta sequences are not always a simple function of animal behavior and can flexibly represent different information on different theta cycles.

The sequence data presented in this paper suggested that there may be subtle changes in single-cell phase precession as animals approach or leave landmarks. As rats approached landmarks, sequences tended to include more space behind the rat (Fig. 6). Thus, we predict that single cells participating in these sequences would be more likely to be firing in the late aspects of their place fields and early phases of theta as animals approach landmarks. Similarly, as rats accelerated away from landmarks, sequences tended to include more space ahead of the rat, suggesting that cells participating in these sequences would be more likely to be firing during late phases of theta. Hints of such single cell effects may be seen in **Supplementary Figure 7**; however, they should be clearest in phase-precessing cells whose firing field covers a sufficiently large space to be bisected by a landmark, such as those in the ventral hippocampus or ventral striatum^{28,29}. Indeed, single phase-precessing cells in the ventral striatum show transient phase changes that could be indicative of sequence effects²⁹. However, a systematic test of this idea should establish if these cells participate in theta sequences, but this requires ensemble sizes beyond those currently available for ventral regions.

Representations of past and future

A number of studies have investigated whether hippocampal activity is reflective of the past and/or the future. Some of these studies have focused on the spatial path represented during neural activity observed in each theta cycle^{9,10,24,26}. Other studies have analyzed hippocampal activity with respect to the animal's past and future behaviors and have shown that the activity of individual place cells or place cell ensembles (at relatively slow time scales) is sometimes more reflective of past behaviors and is sometimes more reflective of future behaviors^{30–33}. Thus, there is evidence that the hippocampus represents more than just the current state of the animal during behavior and does so at multiple timescales³⁴.

We observed a diversity of sequence representations during theta, sometimes reflective of a 'look ahead' function and at other times reflective of 'look behind'. We found that when an animal accelerates, the paths represented are shifted forward in space, possibly in preparation or anticipation of reaching a desired location, whereas when the animal decelerates, the paths represented shift backward in space, potentially in review of the experience. It is possible that the ahead representations observed here reflect a predictive recall process to aid in navigation^{26,27} whereas the behind representations reflect an encoding process to store what was just experienced^{35,36}. This type of recall process could be cued by sensory input from upcoming landmarks, which were previously associated with locations in space³⁷. However, it should be noted that the link between multicell sequences in hippocampus and memory processes such as encoding, retrieval, planning and anticipation is an active area of research and has not been proven.

Additionally, place fields are smaller when spikes are shifted to anticipate the animal's location by 120 ms³⁸, and approaches in

two dimensions overlap only at the late phases of theta^{39,40}, which suggests that portions of the place field may be thought of as representing approach to a future position. We add to this body of work by showing that at the theta time scale, the representation of future space and past space is dissociable and dependent on the animal's acceleration and landmarks in the environment.

Information processing: theta and gamma

It has been proposed that the information represented during each theta cycle is discretized into parts by each cycle of gamma^{41,42}. In this view, the information capacity of a given cycle of theta would be determined by the number of gamma cycles that occur over the course of the theta cycle. In support of these proposals, we observed a linear increase in theta period and the number of gamma cycles per theta cycle as the length of the represented path increased. It is possible that the increase in gamma cycles per theta cycle allows for an increase in the number of information processing steps, thus contributing to longer represented paths.

Behavioral and neural chunking

Chunking refers to a phenomenon in which information is grouped into intuitive, information-rich units. Such recoding is thought to increase information-handling abilities⁴³ and to enable planning through subgoals^{44,45}. It has been observed that individual ventral striatal neurons, which receive hippocampal projections, have receptive fields that represent navigationally relevant segments in an environment⁴⁶. We did find some evidence that the hippocampal spatial place field density and place field length distributions reflected discrete segments (**Supplementary Fig. 3**), but the effect was small compared to the changes in ahead and behind length (Fig. 6). Instead, we found that sequences of place cell activity preferentially represented particular, navigationally relevant chunks of the environment. Thus, chunking was evident from the neural representation of space in the hippocampus.

Several studies have suggested that these neural sequences could be related to navigation^{8,26,47}, planning^{24,42} or memory^{35–37}. For example, evidence exists that systemic cannabinoid agonists disrupt both phase precession and behavioral performance while leaving average place fields unchanged⁴⁷. However, it is unknown whether phase precession and behavioral performance are causally related and whether the disruption of phase precession extends to theta sequences. Nevertheless, our data suggest that the information being processed during each neural sequence expressed in each cycle of theta reflects cognitive segments of the task at hand.

From a behavioral perspective, chunking was also evident on the 2T task from the rat's acceleration-deceleration profile (**Supplementary Fig. 5**). This was not surprising because landmarks on the maze were either turns or feeders, and it would be expected for the rats to slow down as they approached a turn and speed up afterwards. However, we also observed an increase in acceleration before arrival at the top corner of the maze, which implies that the behavioral chunking did not always perfectly align with the physical landmarks.

Although theta sequences preferentially represented segments of the environment, such segmentation was not seen in replay sequences during sharp wave ripples, occurring while animals were awake and paused at feeder locations on the 2T maze¹⁹. It was recently found that replay sequences can represent long paths in space and can occur across multiple ripple events, suggesting that replay events consist of chains of subsequences⁴⁸. Thus, it is possible that theta sequences preferentially encode particular segments of experience, which are then concatenated during extended, multiripple replay events.

Segmentation of the environment by hippocampal theta sequences has implications for the encoding of experience and for how the hippocampus represents space. Because neural activity during theta represents experience as it occurs and operates at the timescale required for spike timing-dependent plasticity, it is thought that synaptic learning during each theta cycle is the basis for storing experience. Thus, theta sequences representing the environment in segments may result in the encoding of experience in chunks. Storing experience this way could be a coding scheme that the brain uses to organize vast amounts of information in a way that is efficient and behaviorally useful to animals.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Boldt, K.D. Seeland and A.P. Steiner for technical assistance and A. Johnson and the members of the Redish lab for discussion. This work was supported by US National Institutes of Health grants (R01 MH-080318 and F30 MH-091821), the Pennsylvania Department of Health and a US National Science Foundation Integrative Graduate Education and Research Traineeship grant (DGE-0549352).

AUTHOR CONTRIBUTIONS

A.S.G., M.A.A.v.d.M. and A.D.R. designed the experiment; D.S.T. and A.D.R. supervised the project; A.S.G. and M.A.A.v.d.M. carried out the experiments; A.S.G. analyzed the data; and A.S.G., M.A.A.v.d.M., D.S.T. and A.D.R. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nn.3138>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Cohen, N.J. & Eichenbaum, H. *Memory, Amnesia, and the Hippocampal System* (MIT Press, Cambridge, Massachusetts, USA, 1993).
- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon Press, Oxford, UK, 1978).
- Hassabis, D., Kumaran, D., Vann, S.D. & Maguire, E.A. Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. USA* **104**, 1726–1731 (2007).
- Schacter, D.L., Addis, D.R. & Buckner, R.L. Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* **8**, 657–661 (2007).
- Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* **9**, 1057–1063 (2006).
- Redish, A.D. *Beyond the Cognitive Map: From Place Cells to Episodic Memory* (MIT Press, Cambridge, Massachusetts, USA, 1999).
- Vanderwolf, C.H. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* **26**, 407–418 (1969).
- Skaggs, W.E., McNaughton, B.L., Wilson, M.A. & Barnes, C.A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Dragoi, G. & Buzsáki, G. Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron* **50**, 145–157 (2006).
- Foster, D.J. & Wilson, M.A. Hippocampal theta sequences. *Hippocampus* **17**, 1093–1099 (2007).
- Tsodyks, M.V., Skaggs, W.E., Sejnowski, T.J. & McNaughton, B.L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6**, 271–280 (1996).
- Jensen, O. & Lisman, J.E. Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn. Mem.* **3**, 279–287 (1996).
- Maurer, A.P. & McNaughton, B.L. Network and intrinsic cellular mechanisms underlying theta phase precession of hippocampal neurons. *Trends Neurosci.* **30**, 325–333 (2007).
- Malhotra, S., Cross, R.W.A. & van der Meer, M.A.A. Theta phase precession beyond the hippocampus. *Rev. Neurosci.* **23**, 39–65 (2012).
- O'Keefe, J. & Recce, M. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- Geisler, C., Robbe, D., Zugaro, M., Sirota, A. & Buzsáki, G. Hippocampal place cell assemblies are speed-controlled oscillators. *Proc. Natl. Acad. Sci. USA* **104**, 8149–8154 (2007).
- Schmidt, R. *et al.* Single-trial phase precession in the hippocampus. *J. Neurosci.* **29**, 13232–13241 (2009).
- Geisler, C. *et al.* Temporal delays among place cells determine the frequency of population theta oscillations in the hippocampus. *Proc. Natl. Acad. Sci. USA* **107**, 7957–7962 (2010).
- Gupta, A.S., Van der Meer, M.A.A., Touretzky, D.S. & Redish, A.D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010).
- Foster, D.J. & Wilson, M.A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
- Diba, K. & Buzsáki, G. Forward and reverse hippocampal place-cell sequences during ripples. *Nat. Neurosci.* **10**, 1241–1242 (2007).
- Carr, M.F., Jadhav, S.P. & Frank, L.M. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* **14**, 147–153 (2011).
- Gupta, A.S. *Behavioral Correlates of Hippocampal Neural Sequences* PhD thesis, (Carnegie Mellon Univ, 2011).
- Johnson, A. & Redish, A.D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007).
- Gothard, K.M., Hoffman, K.L., Battaglia, F.P. & McNaughton, B.L. Dentate gyrus and CA1 ensemble activity during spatial reference frame shifts in the presence and absence of visual input. *J. Neurosci.* **21**, 7284–7292 (2001).
- Maurer, A.P., Burke, S.N., Lipa, P., Skaggs, W.E. & Barnes, C.A. Greater running speeds result in altered hippocampal phase sequence dynamics. *Hippocampus* **22**, 737–747 (2012).
- Lisman, J. & Redish, A.D. Prediction, sequences and the hippocampus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 1193–1201 (2009).
- Royer, S., Sirota, A., Patel, J. & Buzsáki, G. Distinct representations and theta dynamics in dorsal and ventral hippocampus. *J. Neurosci.* **30**, 1777–1787 (2010).
- van der Meer, M.A.A. & Redish, A.D. Theta phase precession in rat ventral striatum links place and reward information. *J. Neurosci.* **31**, 2843–2854 (2011).
- Wood, E.R., Dudchenko, P.A., Robitsek, R.J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
- Frank, L.M., Brown, E.N. & Wilson, M.A. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27**, 169–178 (2000).
- Ferbinteanu, J. & Shapiro, M.L. Prospective and retrospective memory coding in the hippocampus. *Neuron* **40**, 1227–1239 (2003).
- Ji, D. & Wilson, M.A. Firing rate dynamics in the hippocampus induced by trajectory learning. *J. Neurosci.* **28**, 4679–4689 (2008).
- Yartsev, M.M. Dissociating the effects of past and future on neural encoding of sequences in the hippocampus. *J. Neurosci.* **28**, 8383–8384 (2008).
- Buzsáki, G. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus* **15**, 827–840 (2005).
- Hasselmo, M.E. The role of hippocampal regions CA3 and CA1 in matching entorhinal input with retrieval of associations between objects and context: Theoretical comment on Lee *et al.* (2005). *Behav. Neurosci.* **119**, 342–345 (2005).
- Hasselmo, M.E. A model of episodic memory: mental time travel along encoded trajectories using grid cells. *Neurobiol. Learn. Mem.* **92**, 559–573 (2009).
- Muller, R.U. & Kubie, J.L. The firing of hippocampal place cells predicts the future position of freely moving rats. *J. Neurosci.* **9**, 4101–4110 (1989).
- Battaglia, F.P., Sutherland, G.R. & McNaughton, B.L. Local sensory cues and place cell directionality: additional evidence of prospective coding in the hippocampus. *J. Neurosci.* **24**, 4541–4550 (2004).
- Huxter, J.R., Senior, T.J., Allen, K. & Csicsvari, J. Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. *Nat. Neurosci.* **11**, 587–594 (2008).
- Lisman, J.E. & Idiart, M.A. Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* **267**, 1512–1515 (1995).
- Lisman, J. The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* **15**, 913–922 (2005).
- Miller, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956).
- Newell, A. *Unified Theories of Cognition* (Harvard University Press, 1990).
- Newell, A. & Simon, H. *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1972).
- Mulder, A.B., Tabuchi, E. & Wiener, S.I. Neurons in hippocampal afferent zones of rat striatum parse routes into multi-pace segments during maze navigation. *Eur. J. Neurosci.* **19**, 1923–1932 (2004).
- Robbe, D. & Buzsáki, G. Alteration of theta timescale dynamics of hippocampal place cells by a cannabinoid is associated with memory impairment. *J. Neurosci.* **29**, 12597–12605 (2009).
- Davidson, T.J., Kloosterman, F. & Wilson, M.A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507 (2009).

ONLINE METHODS

Methods including the details about the subjects, surgery, recording, histology and the 2T task have been described previously¹⁹.

Subjects. Six male Fisher-Brown-Norway hybrid rats, aged 7–10 months at the start of recording sessions. All procedures were in accordance with the US National Institutes of Health guidelines for animal care and were approved by the Institutional Animal Care and Use Committee at the University of Minnesota. Rats were housed individually and maintained on a 12-h light-dark cycle, with the lights on during the day, when training and recording sessions took place.

Surgery, recording and histology. Three rats were each implanted with a single bundle 12-tetrode, 2-reference microdrive (Neuro-Hyperdrive; Kopf) directed toward the CA1 hippocampal (HC) subfield (3.8 mm posterior and 2.5 mm right-lateral from bregma). Another three rats were each implanted with a double bundle 12-tetrode, 2-reference microdrive directed toward CA1 and ventral striatum (only HC data analyzed here). One reference was placed in the corpus callosum for common-mode noise rejection. The other reference was placed in the hippocampal fissure. All theta LFP signals were analyzed from the second reference electrode (a tetrode with all four channels shorted together) in the hippocampal fissure. The results presented here are qualitatively similar when sequences included for analysis were only allowed one cell per tetrode.

Additional details of the surgery were as previously presented⁴⁹. Neural activity was recorded, and spikes were sorted into putative cells⁴⁹. All units were examined to have a reasonable isolation distance and *L* ratio⁵⁰ and no spikes during the 2-ms refractory period. The median isolation distance and *L* ratio were 21 and 0.09, respectively. Only well-isolated units were included in the analyses. Recordings from an example tetrode are shown in **Supplementary Figure 8**.

After the recording phase of the experiment, rats were overdosed with sodium pentobarbital (150 mg/kg, Nembutal) and perfused intracardially with formalin. Their brains were then transferred to a 30% (wt/vol) sucrose-formalin solution, sectioned, and stained with cresyl violet with standard procedures. All HC recording locations were verified to lie in the CA1 region of the dorsal HC (see **Supplementary Fig. 9**).

The 2T task. The task was a continuous T maze with two choice points arranged in sequence (**Fig. 1a**). Food delivery occurred at two sites on each return rail, contingent on the rat's choice at the second turn. Training and recording sessions lasted 40 min. Training on the task was carried out in three phases. In phase 1, rats were trained to run laps on one side of the maze, while the other side was blocked. Phase 1 continued until rats ran at least 40 laps on two consecutive days. In phase 2, the blocks were removed and on any given day, rats had to run all left laps (L), all right laps (R) or alternating left and right laps (A) in order to receive reward. After consistently getting 80% of the laps correct on all three task contingencies, rats were implanted with hyperdrives. After recovery from surgery, phase two of training resumed until the rats regained proficiency and tetrodes were in the cell layer. At this point, a 6-d sequence of recording sessions began (phase 3). During recording sessions, the task contingency changed approximately halfway through the session (mean, 18.07 min \pm 1.13 min s.d.). There were six recording sessions to allow for all possible pairings of the three tasks (L-R, R-L, L-A, A-L, R-A, A-R).

Data analysis: place fields. Cells firing more than 15,000 spikes over the 40-min session (6.25 Hz) were excluded to filter putative inhibitory interneurons. Additionally, cells with <100 spikes (~0.04 Hz) were excluded. Sessions with >40 cells were considered for analysis (31 out of the 36 recording sessions). A total of 2,046 cells were isolated from 31 recording sessions; 1,881 cells satisfied the number-of-spikes criterion, and 1,470 cells contained at least one place field on the maze. Position along the maze was linearized separately for left and right laps, such that the rat's position along a lap could be described by a scalar value⁵¹. Place fields were then identified as contiguous linear pixels (one linear pixel is ~3.5 cm along the linearized maze) with average activity >5% of the maximum rate observed over the session for any cell at any pixel (cells could have more than one place field, although place fields separated by only a single pixel were merged). Additionally, cells were required to have at least one place field including three adjacent pixels of >2 Hz to be included. Activity during low animal velocity was

not excluded in place field identification. From the 1,470 place cells, 2,729 place fields (and place field centers) were identified. Centers were ordered from maze start to maze start in the direction the rats traveled around the maze.

Data analysis: detecting theta cycles with significant sequences. All theta signals were measured from the electrode placed in the hippocampal fissure taken relative to the electrode placed in the callosum. Individual theta cycles were separated peak to peak, which is the point of minimal cell firing. Peaks of the theta oscillation were identified by first filtering the recording from the reference electrode placed in the hippocampal fissure with a fourth-order Butterworth filter (pass-band between 6–12 Hz) to identify the theta oscillation. The peaks of the filtered signal were then identified by finding the maxima in the filtered signal. The point of minimal cell firing was consistent for the significant theta sequences used in analysis (see below) and was consistent with changes in ahead, behind, and path length (**Supplementary Fig. 7**). A sequence detection algorithm was then used to identify theta cycles whose spike sequence showed significant spatial and temporal structure. The algorithm was nearly identical to the one used previously to analyze replay sequences¹⁹, except that sequences were restricted to begin and end within the same theta cycle. The full algorithmic details are provided below. The algorithm used here and previously¹⁹ detected spatial and temporal structure in the pattern of place cell activity by comparing the times and place field centers of spike pairs. The algorithm was run twice for each session, once using left lap place field centers and once using right lap centers¹⁹. The resulting scores for each spike sequence were then analyzed using two independent bootstrapping procedures to identify sequences with significant spatial and temporal structure occurring during theta cycles with low sharp-wave ripple power. The sequence score procedure was only used to identify the set of candidate theta cycles to study.

Data analysis: sequence identification. The identification of theta sequences with significant structure was carried out in two steps. The first step was to compute the sequence score based on the neural activity in each theta cycle. The second step was to compare the score of each sequence against two distributions created using two separate bootstrapping procedures. These steps are described below.

Step 1: Sequence scoring. The algorithm described here is nearly identical to the algorithm used previously¹⁹. In this study, the algorithm was used to detect coherent sequences of place cell activity during each cycle of the theta oscillation, rather than replay events during the awake state. The sequence score of place cell activity within each theta cycle was calculated as follows. Using the place field centers and spike times, each spike in the theta cycle was pairwise compared with other spikes occurring in the same theta cycle. If the place field center corresponding to spike A was traversed before the place field center for spike B and spike A occurred before spike B, the sequence score was +1, otherwise the score was -1. For all spikes in a given theta cycle, all pairwise comparisons were summed to determine the cumulative score of the sequence.

The basic sequence scoring algorithm was as follows.

Given:

1. Spike times: *T* vector: [nSpikes \times 1]
2. Spike place field centers: *C* vector: [nSpikes \times 1]

Compute:

1. Time difference for each spike pair: ΔT matrix: [nSpikes \times nSpikes]
2. Field center distance for each spike pair: ΔC matrix: [nSpikes \times nSpikes]
3. Element-wise multiply ΔT and ΔC to create ΔTC matrix: [nSpikes \times nSpikes]
4. Binarize ΔTC . For each element in ΔTC : if element is positive, replace with '+1'; if element is negative, replace with '-1'.
5. Sum over all elements in ΔTC to get the sequence score.

A positive sequence score indicates that the sequence is 'forward' ordered.

The scoring presented above requires that each spike be associated with a single place field. For cells with multiple place fields, it is unclear which place field is being represented by a spike. To address this issue, the algorithm assigned each spike from a multifield cell to the place field which best fit with the rest of the firing sequence. This was done by maximizing the score over each place field available to a spike. Always picking the place field that optimizes the sequence score could result in artificially high sequence scores. To provide a fair and

statistically justified comparison, this same score maximization procedure was conducted for the control sequences produced in the bootstrapping procedure described below.

Step 2: Sequence significance testing. Next, each theta sequence was individually analyzed to identify significant structure using two independent bootstrapping procedures. The first method preserved the identity of the cell to which each spike belonged and randomly shuffled the spike times for each event 300 times. Each shuffled event had its sequence score recomputed, once again with spikes from cells with multiple place fields being assigned to the place field center that maximized the sequence score. This ensured that there was no bias resulting from assigning spikes to fields that maximized the sequence score. The second bootstrapping procedure preserved the spike trains, but randomly reassigned each spike train to a different cell's firing field(s)⁴⁸. Each event was shuffled 300 times, once again with spikes from cells with multiple place fields being assigned to the place field center that maximized the sequence score. If the unshuffled sequence score was >95% of both independent sets of shuffled sequences, the sequence was deemed significant and included for analysis. These methods ensured that only significant ($P < 0.05$) theta sequences were included in the analysis.

Data analysis: Bayesian decoding. For theta cycles with significant sequential structure, the represented path in space was determined using a Bayesian decoding approach⁵². Decoding was carried out on each theta cycle using a 40-ms sliding time window shifted by 10 ms at each step. Decoding was restricted to an ~190-cm region (half a lap) centered at the rat's location. Thus, decoding over each theta cycle resulted in an $nTimeBins \times nSpaceBins$ matrix of probabilities, with each row containing the probability distribution over space for a particular time bin. All spikes were included in the Bayesian analyses, whether they contributed to the sequence score or not.

Data analysis: ahead, behind and path lengths. To detect the beginning of the spatial path represented, the probability distributions over space for the first half of the theta cycle (matrix size: $nTimeBins/2 \times nSpaceBins$) were summed to create a single probability distribution over space (an $nSpaceBins$ -element vector). The 5% tail of this distribution (tail on the end of the distribution that is behind the rat) was located, providing the beginning location of the spatial path.

The distance between this point and the location of the rat was defined as the behind length, as it measures the distance the sequence travels from behind the rat to the rat's location. Similarly, the end of the represented path was determined by summing the probability distributions over space for the second half of the theta cycle and locating the 5% tail of that distribution (tail on the end of the distribution that is ahead of the rat). The distance between the location of the rat and this point was defined as the ahead length, as it measures the distance ahead of the rat's location that the sequence progresses. Finally, the total distance from the start of the sequence to the end of the sequence was defined as the path length.

The Bayesian decoding results were robust to the use of other time windows (20 and 30 ms) or other tail sizes (10%, 15% and 20%) and were similar when the first and last time bin were used for determining the ends of the spatial path or when the spatial distributions were summed over the entire theta cycle rather than over the first and second half separately. Furthermore, the results obtained using this Bayesian decoding approach were qualitatively replicated using a separate, non-Bayesian approach for detecting the represented path²³.

Data analysis: location, velocity and acceleration. The rat's location was defined as the location at the middle of the theta cycle (at which cell firing was maximal, **Supplementary Fig. 7**); however, all results are qualitatively unchanged if the rat's location was taken as at another point (beginning or end) of the cycle. Velocity was measured as the derivative of the sequence of locations⁵³, and acceleration was measured as the derivative of the velocity sequences⁵³.

49. Jackson, J.C., Johnson, A. & Redish, A.D. Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience. *J. Neurosci.* **26**, 12415–12426 (2006).
50. Schmitzer-Torbert, N.C., Jackson, J., Henze, D., Harris, K. & Redish, A.D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131**, 1–11 (2005).
51. Schmitzer-Torbert, N. & Redish, A.D. Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. *J. Neurophysiol.* **91**, 2259–2272 (2004).
52. Zhang, K., Ginzburg, I., McNaughton, B.L. & Sejnowski, T.J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79**, 1017–1044 (1998).
53. Janabi-Sharifi, F., Hayward, V. & Chen, C.S.J. Discrete-time adaptive windowing for velocity estimation. *IEEE Trans. Control Syst. Technol.* **8**, 1003–1009 (2000).

**Segmentation of spatial experience by hippocampal theta
sequences**

Anoopum S. Gupta

Matthijs A. A. van der Meer

David S. Touretzky

A. David Redish

Supplemental Information.

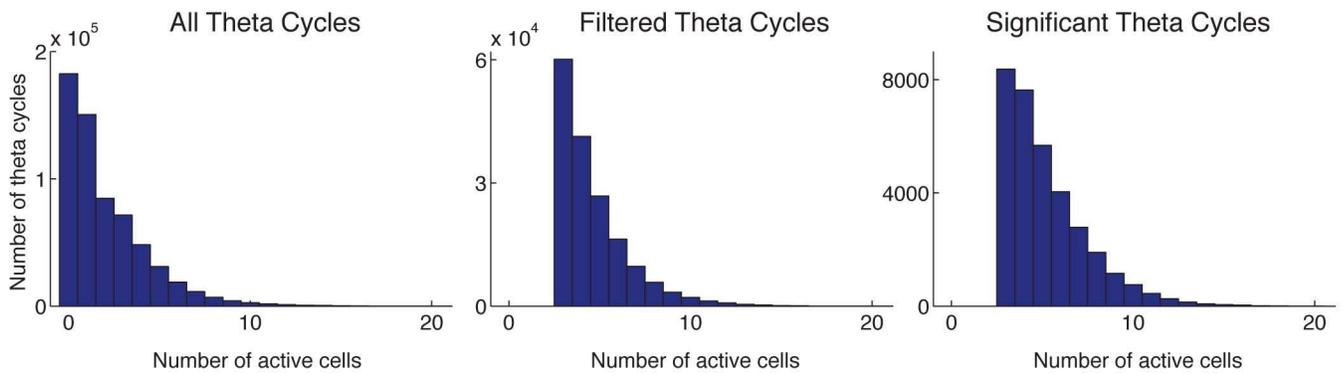


Figure S1: Histograms of the number of theta cycles versus the number of active cells per theta cycle. **Left** panel includes all theta cycles over 31 recording sessions (618,408 cycles). **Middle** panel includes theta cycles with three or more active cells, mean sharp wave ripple power less than one standard deviation above the mean, and theta power greater than one standard deviation below the mean (168,770 cycles). **Right** panel includes theta cycles whose neural activity contains significant sequential structure as described in the Experimental Procedures (33,397 cycles).

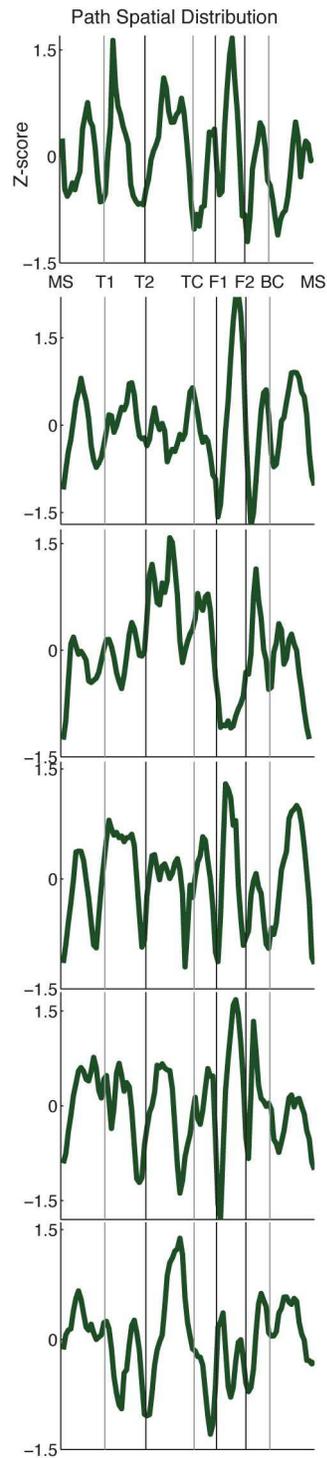


Figure S2: The distribution of space represented by theta sequences for each rat. The main results are robust across individual rats; all six rats chunked the behavior in similar ways.

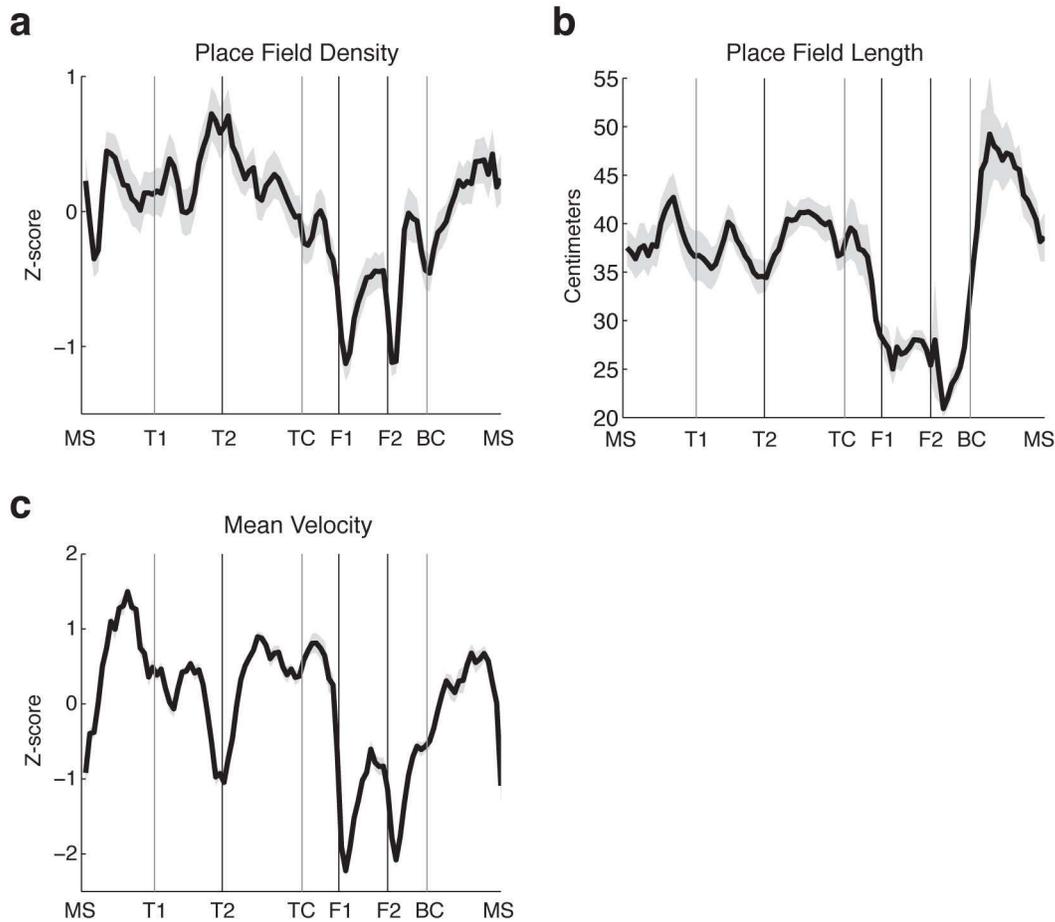


Figure S3: Relationship between place field density, place field length, and animal velocity, and location on maze. **(a)** Place field density as a function of location on linearized maze ($n = 2,729$ place fields). The height of the curve reflects the number of place fields overlapping a particular location in the environment. Y-axis is z-scored place field distribution by session, x-axis is linearized maze. **(b)** Mean place field length as a function of location. Place field length is defined as the length of the place field measured along the linearized maze. Y-axis is centimeters averaged by session, x-axis is linearized maze. **(c)** Animal velocity as a function of location on linearized maze. Y-axis is z-scored velocity by session, x-axis is linearized maze. Gray line width is s.e.m.

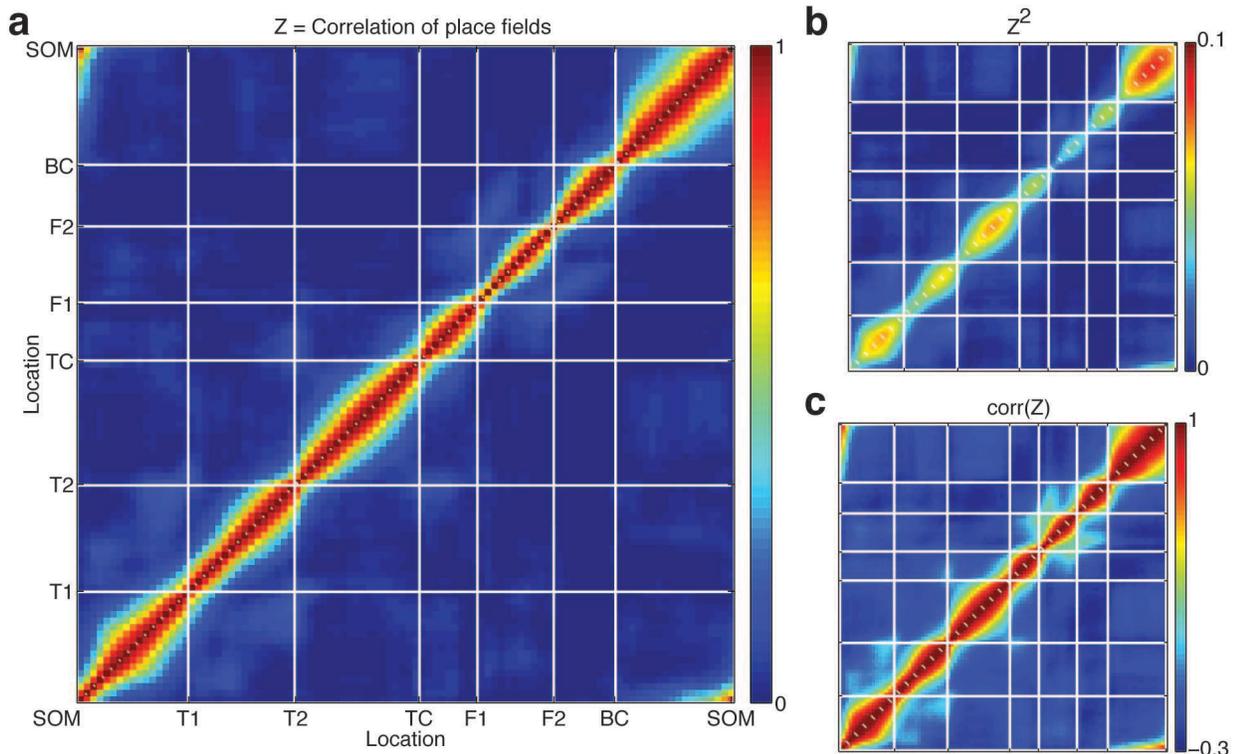


Figure S4: (a) Spatial correlation matrix of place fields. To calculate this matrix, the spatial firing tuning curve (place field) of each cell was calculated, creating an $n_{\text{cell}} \times n_{\text{spatial-bin}}$ matrix, with each element in the matrix the average firing rate of the cell at that particular location in the environment. This matrix was then correlated to create an $n_{\text{spatial-bin}} \times n_{\text{spatial-bin}}$ matrix. (b) The square of the matrix shown in panel a. (c) The cross-correlation of the matrix generated in panel a. 2,729 place fields were used in this analysis.

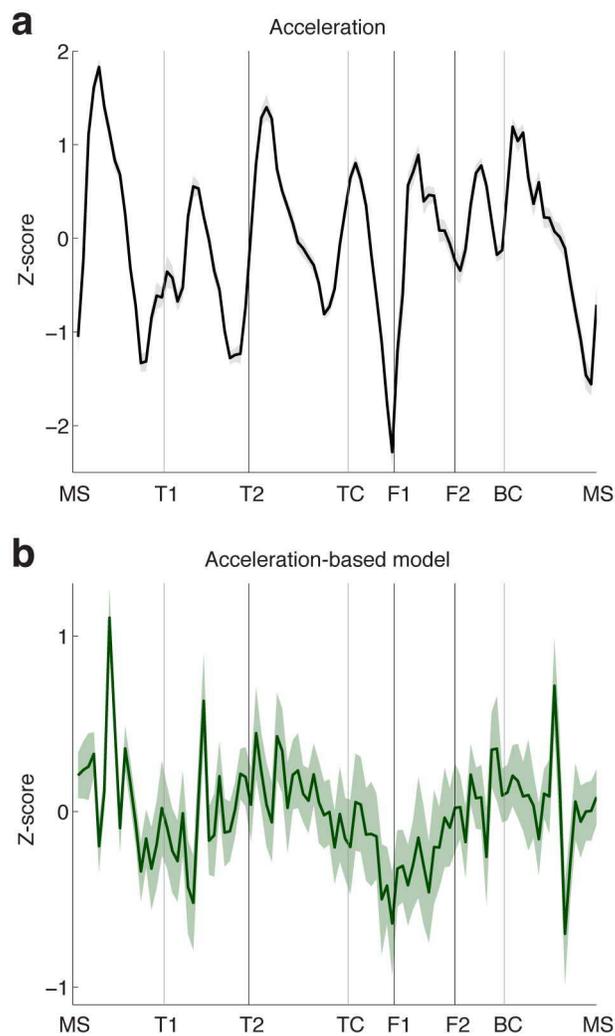


Figure S5: Spatial distribution of paths constructed from an acceleration-based model of represented spatial paths. **(a)** Mean acceleration (z-scored by session) as a function of location on linearized maze. **(b)** Spatial distribution of paths constructed based on the relationship between acceleration and ahead and behind length (see main text for details). Comparing this to the spatial distribution shown in **Figure 6** of the main text clearly shows that acceleration is not the sole driving force for spatial chunking. 29,351 theta sequences were used in this analysis. Line width is s.e.m.

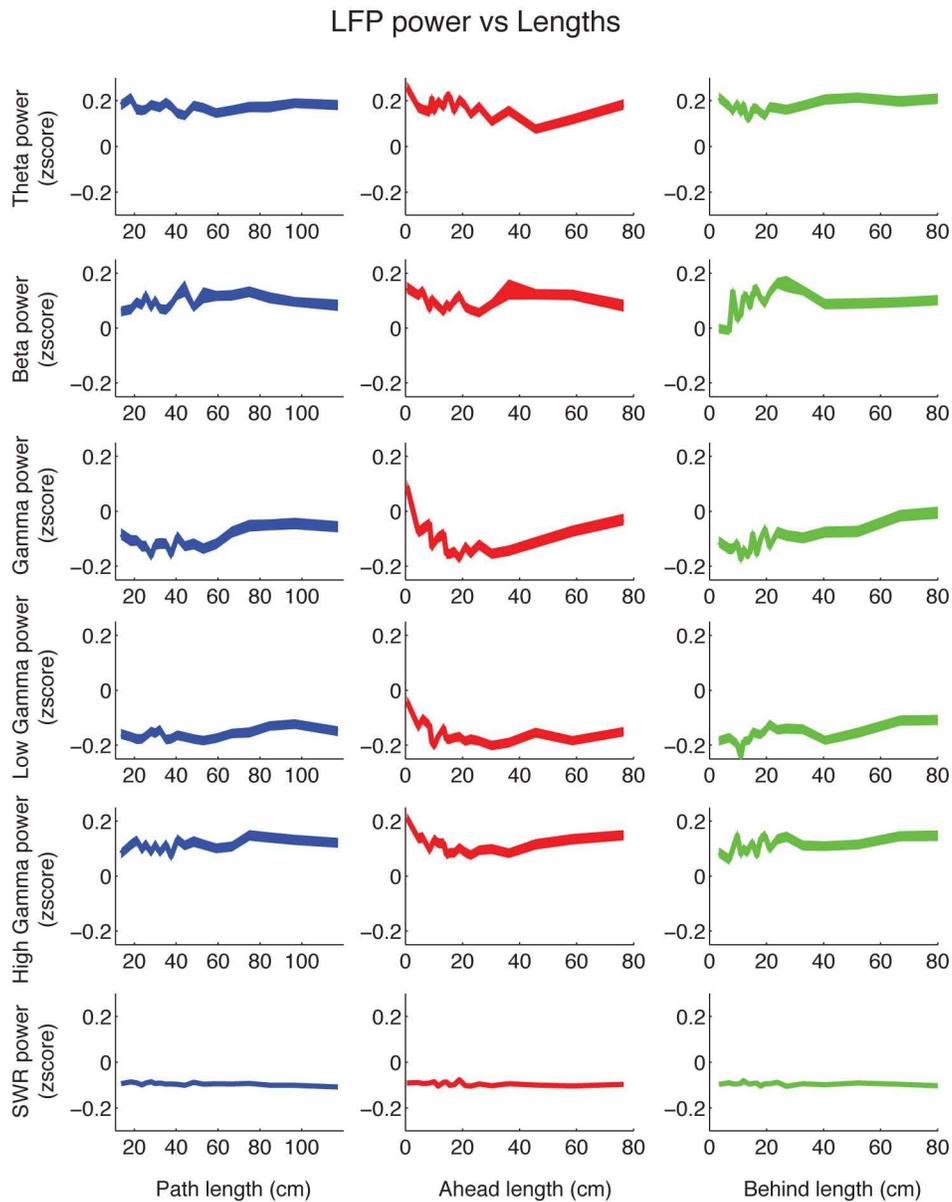


Figure S6: Mean LFP power as a function of path, ahead, and behind length. 29,351 theta cycles were used in this analysis. Mean LFP power was z-scored by session. Line width is s.e.m.

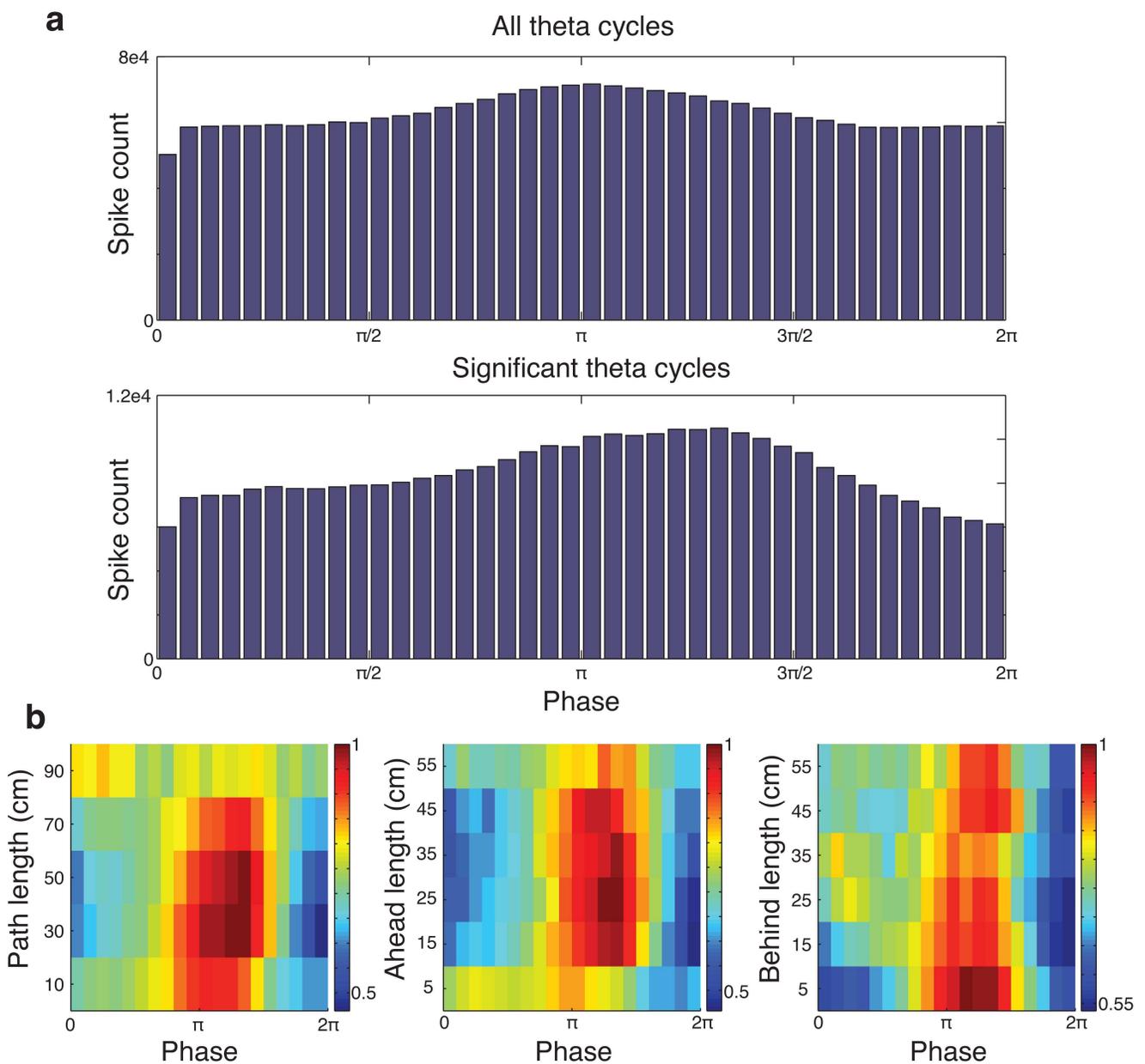


Figure S7: Theta sequence versus theta phase relationships. **(a)** Histogram of putative pyramidal cell spiking as a function of theta phase for all theta cycles (**top**) and for those theta cycles included in the analysis (**bottom**). Minimal cell firing was observed at the peak of theta (0 and 2π). **(b)** 2D histograms showing the distribution of spikes over theta phase as path length (**left**), ahead length (**middle**), and behind length (**right**) increase. Red indicates high spike counts and blue is low spike counts. As seen in this figure, the distribution of spikes over the theta cycle is similar as the three lengths increase, and are similar to the distribution of spikes for all significant theta cycles (see panel

a). Note that the distribution of spikes over the theta cycle for the longest path lengths is based on relatively few theta cycles, which is the likely cause of the differing distribution. Phases of zero and 2π correspond to peaks of the theta oscillation recorded from the hippocampal fissure.

Tetrode 4 taken from session R153-2008-09-15

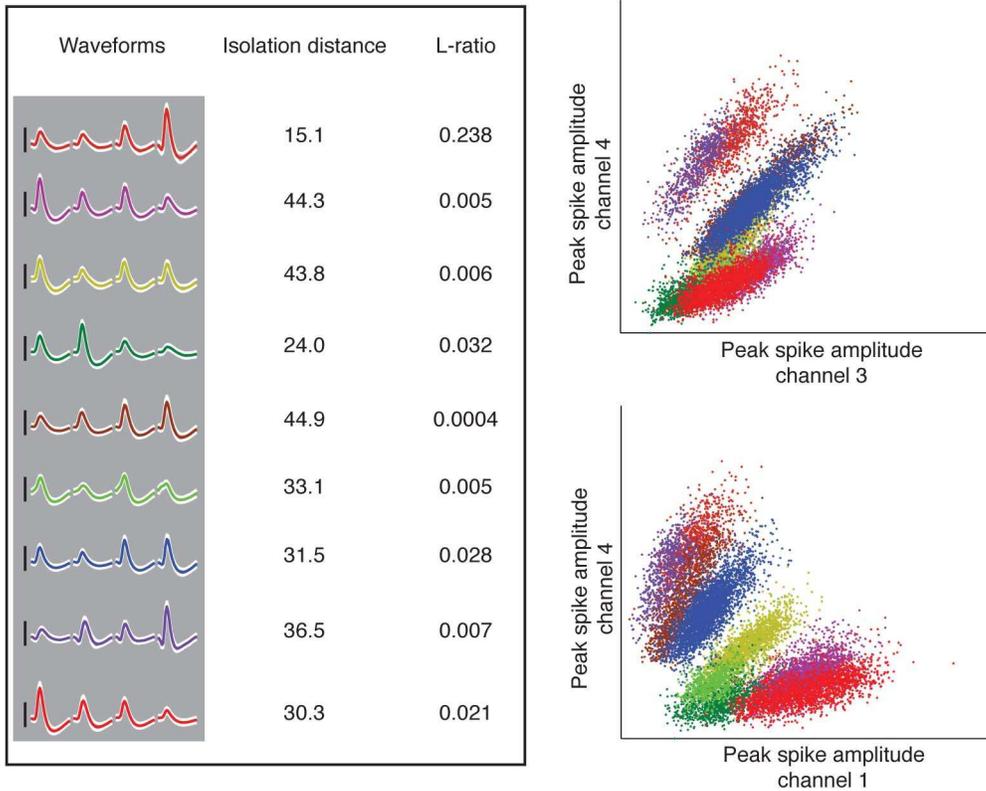


Figure S8: Example tetrode recordings from one recording session. Each spike cluster is shown with different colors on two projections (right). One millisecond waveforms for each of the four tetrode channels are plotted for each separable cluster along with the cluster's Isolation distance and L-ratio (left). Scale bar next to each cluster's waveforms is 100 μ V.

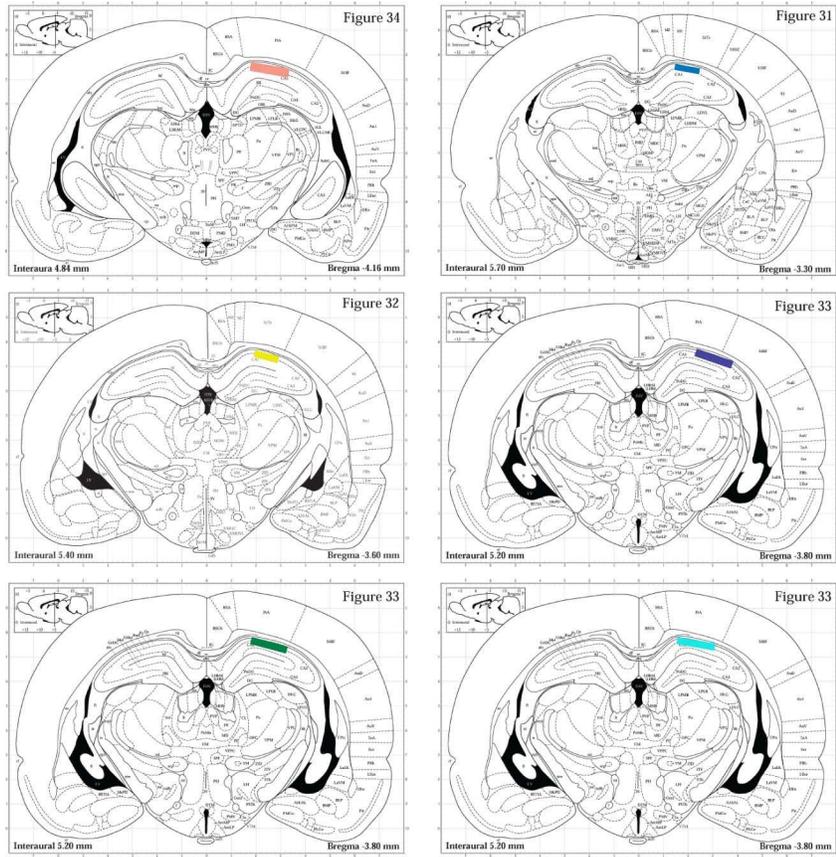


Figure S9: Schematics showing the position of implanted electrodes. Each colored box indicates the location of the electrodes within hippocampal region CA1 for each of the six rats in this study.