

Specific evidence of low-dimensional continuous attractor dynamics in grid cells

KiJung Yoon¹, Michael A Buice¹, Caswell Barry²⁻⁴, Robin Hayman⁴, Neil Burgess^{2,3} & Ila R Fiete¹

We examined simultaneously recorded spikes from multiple rat grid cells, to explain mechanisms underlying their activity. Among grid cells with similar spatial periods, the population activity was confined to lie close to a two-dimensional (2D) manifold: grid cells differed only along two dimensions of their responses and otherwise were nearly identical. Relationships between cell pairs were conserved despite extensive deformations of single-neuron responses. Results from novel environments suggest such structure is not inherited from hippocampal or external sensory inputs. Across conditions, cell-cell relationships are better conserved than responses of single cells. Finally, the system is continually subject to perturbations that, were the 2D manifold not attractive, would drive the system to inhabit a different region of state space than observed. These findings have strong implications for theories of grid-cell activity and substantiate the general hypothesis that the brain computes using low-dimensional continuous attractors.

A set of N uncoupled spiking neurons, each with dynamic range Q , supply a vast representational space (volume $\sim Q^N$), (Fig. 1a). However, the representation has poor resistance to noise: each state is independent, and if it is changed to another, there is no restoring dynamics to correct the state. Even in the absence of noise, the states persist only for the time constant of single neurons.

Coupling between neurons generally disallows many states, shrinking the representational space. An advantage of coupling is that it can, in special cases, produce stable fixed points (attractors) of the network dynamics that allow the network to hold a state after inputs are removed, for far longer than the single-neuron time constant (Fig. 1a). Moreover, if noise is present in the system, it may perturb the system off the attractor, but the perturbations are transient and automatically corrected as the system rapidly flows back toward the attractor (Fig. 1a). Discrete or point attractors, as in Hopfield networks, may be used to represent discrete items¹. In many cases, the brain must represent continuous variables. In these cases, the value of the variable could be represented as a point on a continuous manifold of stable fixed points, of the same dimensionality D as the variable²⁻⁵. This manifold is called a low-dimensional continuous attractor if its dimensionality is much smaller than the number of neurons in the network ($D \ll N$). In these ways, attractors enable robust representation and memory, albeit at the cost of diminished representational space.

Low-dimensional continuous attractor dynamics have been widely hypothesized to underlie the stable tuning curves of population codes^{2,6}, motor control^{3,7}, neural integration^{3,4,8-15} and parametric working memory^{16,17}. The predicted signatures of low-dimensional continuous attractors in the neural context are systematic differences in neural responses along the attractor manifold but conformity and stability otherwise^{3,4,12,15}. For example, in a ring of neurons with local recurrent excitation and global inhibition along the ring, a local bump

of activity is a stable state (bump A in Fig. 1a) as well as all its translations (bump B and other translations, Fig. 1a). Thus, each activity bump is one point on the 1D continuous attractor of all possible bump locations in the network (Fig. 1a). If points on this attractor are identified with the values of some circular variable in the external world, then all neural tuning curves for that variable will be identical, except for systematic differences along one dimension, in the preferred angle (translation along the ring). Differences in preferred angle are determined by neural placement and connectivity in the network: neurons in the span of one activity bump will be co-active, and the preferred angles of neurons separated by 90° on the ring will be in quadrature. When recurrent connectivity is fixed, these rules of co-activation will remain absolutely stable, regardless of how the mapping from the attractor states to the external variable might vary due to changing associations between the network and the external world (for example, rotation of the mapping induced by rotation of the world⁴ or more elaborate changes involving stretching or squeezing of tuning curves¹⁵).

In other words, the responses of cells, when plotted against the external represented quantity, may change substantially, but pairwise relationships between cells should not. Despite these predictions, and beautiful empirical results^{7-11,18,19}, definitive validation of the low-dimensional continuous attractor hypothesis has been elusive. In many cases, partly because of the difficulty of inducing sufficient change or perturbation in the neural responses and partly because quantitative analyses on simultaneously recorded neural pairs have not been conducted, it has been unclear whether the dynamics are truly low-dimensional, what the dimension is or where the dynamics originates^{10,18,19}.

Mammalian grid cells²⁰, each of which fires at the vertices of a regular spatial grid as the animal moves through its environment,

¹Center for Learning and Memory, University of Texas at Austin, Austin, Texas, USA. ²Institute of Cognitive Neuroscience, University College London, London, UK. ³Institute of Neurology, University College London, London, UK. ⁴Institute of Behavioural Neuroscience, Division of Psychology and Language Sciences, University College London, London, UK. Correspondence should be addressed to I.R.F. (ila.fiete@mail.cln.utexas.edu).

Received 28 December 2012; accepted 23 May 2013; published online 14 July 2013; doi:10.1038/nn.3450

have been hypothesized^{12–15,20–23} to compute ongoing location estimates through integration of self-motion cues, based on the theoretical argument that their responses constitute a relatively context-independent code for spatial displacements. Across different familiar environments, the firing field locations in a grid cell change only through global phase shifts and rotations^{20,24}, in contrast to what occurs in place cells, which change in more elaborate ways, by gains or losses and shifts in subsets of their place fields^{25–29}. The notable regularity and stability of a grid cell's response hints that the population might be well described by relatively few parameters, a signature of an underlying low-dimensional dynamic system.

In contrast, spatially periodic firing in grid cells is neither necessary nor sufficient for inferring low-dimensional population dynamics: it is theoretically possible for the population dynamics to be low-dimensional and periodic without spatially regular firing in individual cells, because of poor velocity integration¹⁵. Conversely, if the N cells in a single population have periodic spatial responses but each exhibits independent shifts (relative to the other cells) of its spatial phase across environments, the dimensionality of the population response would be high, or $\sim N$. Finally, experiments that involve resizing of a familiar environment, or exploration in novel environments, reveal that spatial responses of grid cells stretch along one or both dimensions^{30–32}, a malleable response that is possibly consistent with high-dimensional dynamics.

Here we examined the spikes of simultaneously recorded pairs of grid cells, from several data sets across a range of experimental conditions, including cases where the single-cell responses changed considerably and where external inputs do not provide reliable spatial cues, to rigorously determine across conditions the dimensionality of the population response for each grid-cell 'network' or 'module' (that is, local groups of cells with a common spatial period and orientation; such discrete networks or modules have been predicted to exist through modeling^{12,15,22,33} and subsequently experimentally validated^{30,32}), and thus probe for evidence of low-dimensional

continuous attractor dynamics in the brain. The results reported below include all simultaneously recorded cell pairs from these data sets, in which both cells of the pair have a common spatial period and meet a modified gridness score that assigns similarly high scores to equilateral and nonequilateral triangular patterns of spatial activity (Supplementary Figs. 1 and 2 and Online Methods). We relate the empirical findings to dynamic models of grid cells, to generate constraints on the mechanisms that underlie grid-cell response.

RESULTS

Identical spatial responses up to 2D translation

We examined spikes from neurons recorded simultaneously from the same or nearby tetrodes. The activity peaks of a sample pair (Fig. 1b) were arranged in the spatially periodic firing patterns characteristic of grid cells. Our definition of the spatial responses of grid cells, here and in the rest of this work, is the set of locations of the firing peaks. Six parameters are sufficient to characterize any periodic tiling in two dimensions, regardless of the shape of the tiles³⁴. Thus, the spatial response of an individual cell in a particular environment is well described by four parameters for the angles and lengths of two primary lattice vectors (Fig. 1c), with two additional parameters that specify the 2D spatial phase of the lattice, relative to some reference phase or location.

We found that cell pairs from the same or nearby tetrodes had extremely similar values for the first four parameters (Fig. 1c; 223 cell pairs: 24 from ref. 20; 97 from ref. 35; 12 from ref. 30; 90 from ref. 31). This was the case even though the cells had very different spatial phases (Fig. 1d), that is, even when the cells were active in complementary parts of the environment. The relative phase between cell pairs, defined as the difference in their spatial phases, seems to be uniformly distributed ($n = 223$ cell pairs; uniform under the two-dimensional paired Kolmogorov-Smirnov test, $P = 0.2$) over the unit cell of the lattice (Fig. 1d; consistent with a previous result²⁰).

Figure 1 Spatial grid parameters other than phase are identical across cells recorded on the same or nearby tetrodes; cell-cell relationships are stable over time. (a) State space of N independent neurons (shown for three neurons; top), each described by a firing rate r_i in $[0, r_{\max}]$, may lie anywhere in an N -dimensional cube of side length r_{\max} . A low-dimensional attractor (dark blue) induced by appropriate neural coupling. Transient states (gray) are unstable and rarely seen. States very close to the attractor (light blue), through transient, may be observed if perturbations frequently drive the system into those states. An example ring network (bottom) of N neurons (small circles) with 1D continuous attractor dynamics. Blue bumps A and B, stable activity patterns. Gray, unstable population activity profile. (b) Column one: recorded spikes (red dots) of two simultaneously recorded cells as a function of space (rat trajectory: gray lines). Column two: autocorrelograms of the smoothed spatial response (peaks identified by black asterisks). Column three: template lattice (red circles) is fit to all the peaks of the autocorrelogram. Parameters of the template (c, inset) include the two primary axis lengths (λ_1, λ_2) and two angles (θ, ψ). Column four: cross-correlogram between the two cells (top), and the corresponding template fit (bottom). (c) Ratio of each lattice parameter across 223 cell pairs (for example, θ (cell i)/ θ (cell j), where $i > j$) (median ratio: center line in box plots; interquartile ranges: box; lowest and highest values within 1.5x of interquartile range: outer horizontal lines; 95% confidence interval based on 223 randomly chosen pairs not recorded simultaneously: dotted outer horizontal lines). (d) Distribution of relative phases ($\delta^{\alpha\beta}$, black circles; see definitions in Online Methods) between all cell pairs, plotted in a canonical unit cell of the grid lattice. (e) Discharge maps (as in b) of the same cell pair, recorded again after an interval of >60 min. (f) Parameter ratios (as in c) from this later trial, for the subset of cell pairs from c that were also recorded in this trial ($n = 84$ cell pairs).

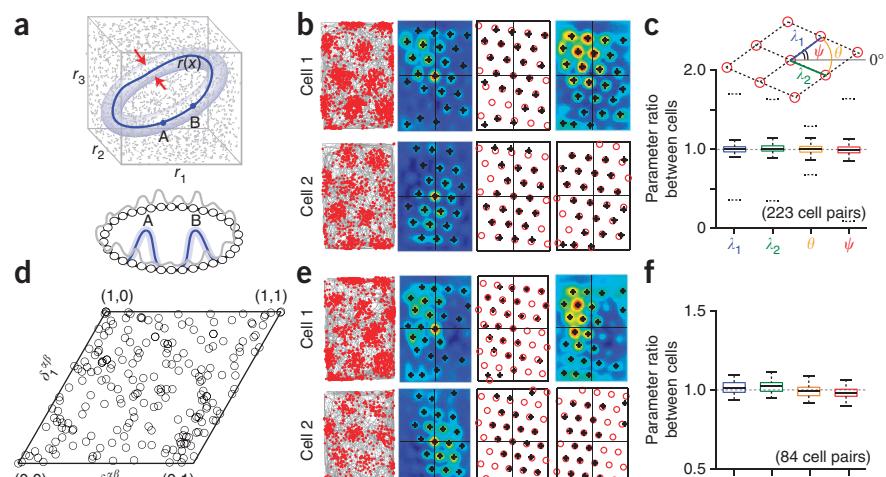
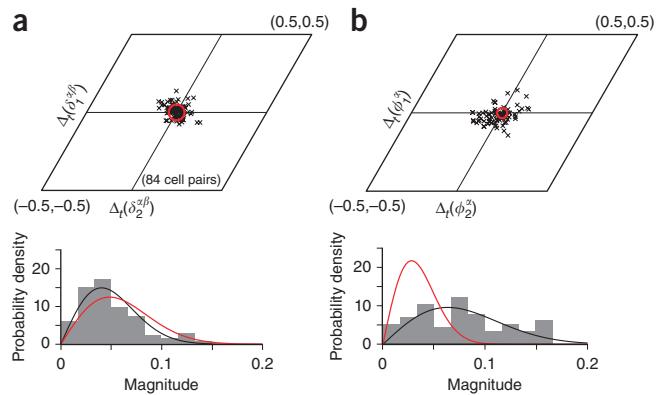


Figure 2 Across time in familiar environments, the relative phases between cells are more stable than the phases of single cells. (a) Difference across time (trials separated by >60 min) in the relative phase between cell pairs ($\Delta_t(\vec{\phi}^{\alpha\beta})$, black \times symbols; see definitions in Online Methods); red circle indicates uncertainty in estimating relative phase differences (error analysis in Online Methods). Histogram (normalized by histogram area) of the magnitudes of these relative phase differences, with the null distribution (red curve), in which phase differences are not significantly different from zero and drawn independently from a Gaussian with s.d. equal to the uncertainty in phase estimation (bottom). The null distribution of magnitudes is Rayleigh. Black curve is the best-fit Rayleigh distribution to the data. (b) Difference across time (trials) of the phase of single cells ($\Delta_t(\vec{\phi}^\alpha)$, black \times symbols; see definitions in Online Methods; $n = 75$ distinct cells of the cell pairs in a), and the normalized histogram of magnitudes (bottom), as in a. Data in a are not significantly different from the null hypothesis, but those in b are: $P = 0.58$ (a) and $P < 10^{-4}$ (b); F test for whether the data and the null distribution come from a distribution of the same variance. $P < 0.001$; F -test for whether the data in a,b (bottom) come from a distribution of the same variance.



Cell-cell relationships more stable than single cells

We next examined the stability over time of each cell's response and of cell-cell response relationships. Without any detailed analysis, the fact that a clear grid pattern was visible in the responses of individual neurons over a 20-min recording session means that the individual spatial phases, relative to some environmental landmark, remained essentially constant over the session; if the phase shifted over time, the cell would fire at these different phases, and its spatial activity pattern would be washed out. It follows directly that the relative phase between cells of the same spatial period and orientation will also be constant over that interval.

We analyzed responses of grid cells over longer time intervals: we recorded cells in an environment, then after an interval in which

we tested the rat under varying conditions and environments, we recorded the cells again in the original environment (Fig. 1e). The elapsed time between recordings in the original environment was >60 min. We found that cells that had essentially identical values of the first four grid parameters in the first measurement (Fig. 1b,c) had essentially identical parameter values in the subsequent measurement in the original environment (Fig. 1f; 84 cell pairs from refs. 30,31).

Moreover, the relative phases between cell pairs remained essentially constant over this interval (Fig. 2a, 84 cell pairs), consistent with continuous attractor dynamics that stabilizes cell-cell relationships. However, this stability of relative phase between cells may be attributable to the stability of the phases of individual cells across visits to the original environment. To differentiate between the two possibilities,

Figure 3 Grid parameter ratios and relative phases are stable even when grid parameters are rescaled as the environment is resized. (a) Firing fields of two simultaneously recorded cells in a familiar environment (trials 1 and 5) and resized versions of the familiar environments (trials 2–4). (b) Spatial cross-correlograms for the cell pair (top) and the best-fit template lattices (bottom). Black asterisks denote local peaks in the cross-correlogram. (c) Each grid parameter for cell 1 (top) or cell 2 (bottom) normalized by the value from trial 1. The parameters are substantially rescaled across trials 2–4. (d) The ratio, between cells 1 and 2, of each grid parameter, for each trial. The ratios are statistically very close to one, despite the significant rescaling in each cell, seen in c. Error bars (c,d), ± 1 s.d. ($n = 100$ bootstrap samples, Online Methods). (e) Histogram of all grid parameters for the 11 cells in the resizing experiments from trials 2–4 normalized to the corresponding value from trial 1 (top). Histogram of the ratios of all grid parameter values between cells 1 and 2 for all seven cell pairs from trials 1–4 (bottom). Kolmogorov-Smirnov test for whether the two data samples come from the same distribution, $P < 0.001$ (top, $n = 117$; bottom, $n = 121$). F test for whether the two data samples come from a distribution of the same variance, $P < 0.001$. (f) Relative phases for the 7 cell pairs span the unit cell (copies of the same symbol represent different trials for one cell pair; different symbols represent different cell pairs). Gray 'x's, relative phase differences, computed across all cell pairs and trials. Red circle, uncertainty in the relative phase difference magnitude (Online Methods). The relative phase differences are not significantly different from zero ($P = 0.6$ for the same null hypothesis as in Fig. 2).

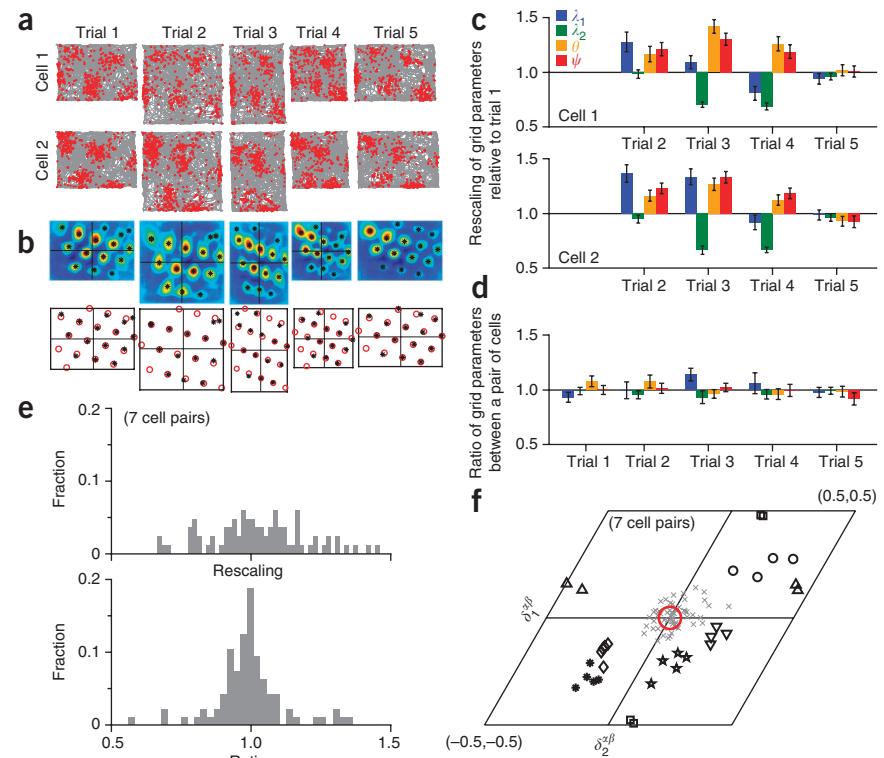


Figure 4 Spatial patterns of grid cells become less stable and expand in novel enclosures, but grid parameter ratios between cells remain stable. (a) Firing fields of pairs of simultaneously recorded cells in a familiar environment and novel ones across five consecutive trials on one day, and corresponding cross-correlograms and best-fit template lattices. On different days, the recordings involved different cells from the same tetrodes in the same area in the rat (Supplementary Fig. 4). (b) Development of average modified gridness in novel environments (24 cell pairs, from ref. 31; means \pm s.e.m.). (c) Change, across trials and days in the novel environment, of the average grid period (mean of the first two grid parameters across all cells in a trial: means \pm s.e.m.; 24 cell pairs total: 1, 6, 10, 3, 1 and 3 on days 1, 3, 4, 5, 6 and 7, respectively; no cells in day 2 passed the gridness criterion). (d) Grid parameters of one typical cell pair from each day (all cell pairs shown in Supplementary Fig. 4), normalized by the corresponding parameter values from the first trial (familiar environment) of the day. Clusters of four narrow bars represent the four parameters, in the same ordering and color scheme as in Figures 1 and 3. (e) Grid parameter ratios for the two cells, across trials and days. Almost all these ratios are statistically indistinguishable from 1 (for all cell pairs, see Supplementary Figs. 4 and 5). Error bars (d,e), ± 1 s.d. (Online Methods).

we compared the stability in the spatial phase of single cells with the stability of relative phase between cells. The relative phase between cells was more tightly preserved than the phase of each cell (Fig. 2b), strongly suggesting a low-dimensional internal dynamics that yokes together the responses of different cells in the network rather than spatially informative external cues.

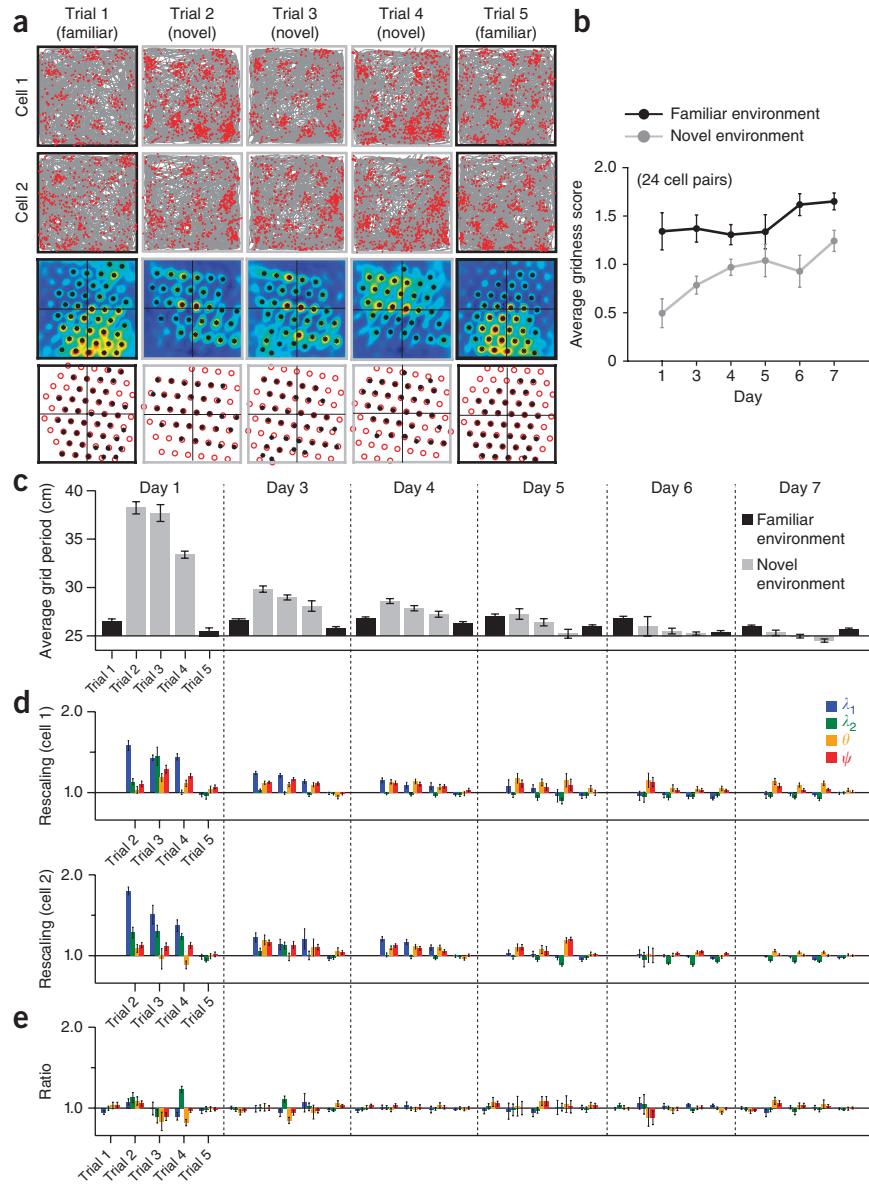
Cell-cell relationships stable despite grid distortions

Next we tested pairwise relationships between grid cells that undergo a nonuniform rescaling of their individual spatial response patterns when a familiar enclosure is suddenly resized³⁰ (Fig. 3a,b). This rescaling constituted a major change in the responses and grid parameters of individual cells (Fig. 3c). But despite rescaling, the ratios of the first four grid parameters between cell pairs remained fixed very close to 1, indicating that these parameters changed in tandem across the population (Fig. 3d). This result held for all cell pairs (Fig. 3e, 7 cell pairs from ref. 30).

The relative phases between cells remained constant (Fig. 3f). Thus relative phase was strongly conserved even when the responses of single cells, influenced by changing external cues, had changed considerably, again strongly suggesting a 2D state space and simultaneously suggesting that internal dynamics rather than external cues are responsible for the low-dimensional response.

Cell-cell relationships stable without place cell stability

The preceding results are consistent with a population response that is confined to a 2D manifold: given the spatial response of one cell,



the responses of the others are always predictable, differing from the single cell only by a fixed 2D phase shift, which remains invariant across conditions. However, it remains possible that, rather than arising from attractor dynamics generated in the grid-cell network, the low-dimensional response of grid cells is externally imposed, by sensory cues from the familiar room, that somehow are flexible enough to permit rescaling of individual responses yet rigidly force fixed relative phases, or more plausibly, by inputs from the hippocampus.

To address these possibilities, we analyzed responses of grid cells from the rats' first exposures to novel environments. Grid cells had been recorded first in a familiar environment, then in a novel environment and then again in several subsequent sessions in the novel environment as it became gradually more familiar³¹. In the first exposure to a novel environment, the spatial periods of grid cells expanded suddenly and the responses became less regular (Fig. 4a-c; 24 cell pairs from ref. 31). As the novel environment became more familiar, notable changes occurred: the responses became more grid-like (Fig. 4b) and the grid periods contracted steadily (Fig. 4c). Indeed, the four parameters of the spatial grid patterns all changed in the novel environment then relaxed back over ~ 7 d to values observed

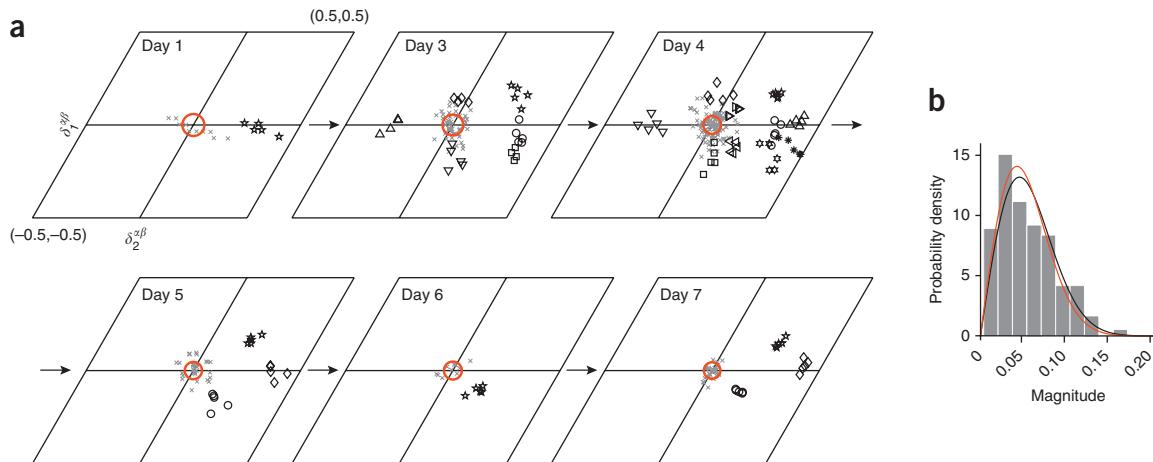


Figure 5 Relative phase remains stable in novel enclosures. **(a)** Relative phases of cell pairs (each day, multiples of one symbol represent a single pair across trials), across different trials and days ($n = 24$ cell pairs from ref. 31, across all 7 d). Gray 'x's, relative phase difference for every trial and all pairs. Red circle, uncertainty in the magnitude of relative phase differences (Online Methods). **(b)** The relative phase differences are not significantly different from zero for the same null hypothesis as in **Figure 2** ($P = 0.38$ under the *F*-test for whether the data and the null distribution come from a distribution of the same variance).

in familiar environments (**Fig. 4d**), while the environmental sensory input remained unchanged. This suggests that the response of grid cells and their relationships are not determined, and by extension, not stabilized, by the environmental sensory input during this contraction period. With this in mind, we next examined the relationships of grid parameters between cells. The ratios of each grid parameter between cell pairs remained close to unity in the novel environment, as in familiar environments, starting with the very first exposure in the novel environment (**Fig. 4e**) and continuing throughout the period of contraction of the grid period. The relative phase between cells again remained essentially unchanged between the familiar environment and the very first exposure to a novel environment (**Fig. 5**; 24 cell pairs from ref. 31) despite the large changes in the spatial activity patterns of neurons and thus in their absolute phases. The relative phase between cells continued to remain fixed as their spatial activity patterns shrank over repeated exposures to the novel environment (**Fig. 5**).

Because the rat in a novel space has not yet learned to associate external sensory cues with location^{36,37} and because the external cues remained fixed as the spatial activity patterns of grid cells shrank over several days, it follows that the relationship between grid cells and the external world is less stable than is the relationship between grids cells. It is therefore unlikely that external sensory cues are stabilizing cell-cell relationships across novel and familiar environments.

In contrast to responses of grid cells in the novel environment, simultaneously recorded place cells underwent complete remapping³¹, defined as the loss of some of their firing fields and gain of others, with little preservation of spatial correlations between place-cell firing fields. Similar to the grid cells, place cell firing fields also expanded, then shrank, but to a much lesser extent than the grid cells, and for a shorter time. The remapped responses were not immediately stable, taking hours to stabilize³¹ (see also refs. 25,38–40 for similar results on place-cell remapping). Thus, hippocampal representations, and by extension place cell-grid cell relationships, are in flux whereas grid cell-grid cell relationships remain stable, suggesting that hippocampal input is not stabilizing relative phases between grid cells in the transition from familiar to novel environments and in the novel environment. Taken together, these findings suggest that the hippocampus cannot be generating and feeding forward the 2D stable responses observed in grid cells.

The above result is not inconsistent with the finding that hippocampal inputs might be required for activity of grid cells⁴¹. A decrease in excitatory drive from the hippocampus and the resulting diminished activation of grid cells can, if the recurrent connections between grid cells are dominantly inhibitory^{42,43}, result in a failure of the recurrent connections to induce population patterning and low-dimensional dynamics, as shown in the models of refs. 15,41. Hippocampal inputs might also correct errors in path integration by selecting the appropriate grid-cell population state for a given location from a set of stable population patterns^{23,44,45}, thus enabling accurately patterned spatial responses over a trajectory¹⁵. In either of these cases, the two-dimensionality of the grid-cell population response is intrinsic to the entorhinal cortex, but abolishing hippocampal drive may abolish spatial patterning.

We conclude, first, that the response of the grid-cell population is restricted to the same 2D manifold at the first exposure to novel environments as in familiar environments and second, that this restriction to the 2D manifold cannot easily be ascribed to external sensory cues or hippocampal inputs because relative phases and parameter ratios are stable even when these inputs are not.

Smoothness (continuity) of 2D manifold

We next more closely examined the 2D manifold of stable grid parameters and relative phases to determine whether it exhibits a granular or 'lumpy' structure, in which cell-cell stability is dependent on cell-cell similarity. A scenario in which cells with similar spatial phases conserved their relationships with each other more strongly, would be consistent with distinct subnetworks of cells with similar spatial-response patterns stabilizing each other and not others with more dissimilar responses. Thus, we reexamined the results from familiar, rescaled and novel environments, plotting parameter ratios as a function of relative phase between cell pairs. Parameter ratios were consistently close to 1, independent of the distance in phase between cells (**Fig. 6a**). The stability in cell-cell relationships across rescaled and novel enclosure trials, as measured by stability of parameter ratios (**Fig. 6b**) and stability of relative phases (**Fig. 6c**), did not vary with relative phase magnitude between cell pairs. Therefore, up to experimental resolution, the 2D manifold of states was continuous.

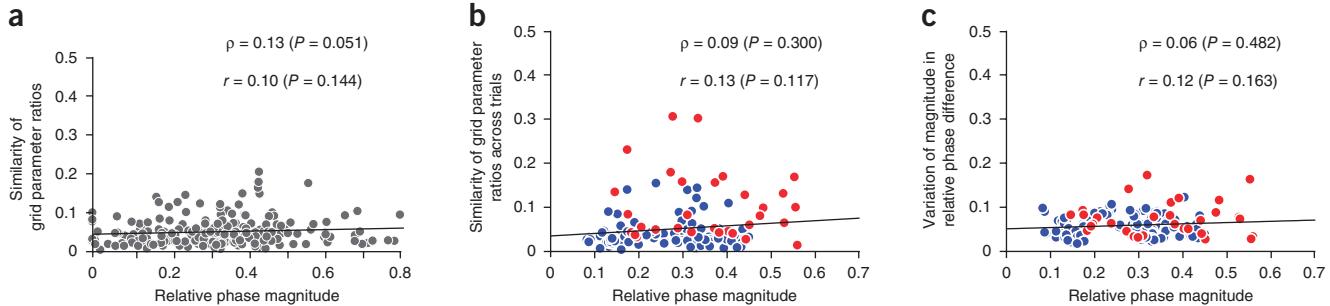


Figure 6 Stability of cell-cell relationships is independent of distance in spatial phase. (a) Parameter similarity between cell pairs (defined as the square-root of the squared deviation of parameter ratios around the value 1, averaged over all parameters for the cell pair), plotted as a function of the pair's relative phase (223 cell pairs from Fig. 1c,d). Each dot represents one trial from one cell pair. Black, linear regression; ρ , Spearman's rank correlation; r , Pearson's product-moment correlation. (b) Stability of parameter ratios between cell pairs across rescaling trials (red dots, 7 pairs from Fig. 3f) and novel enclosure trials (blue dots, 24 pairs from Figs. 4 and 5) is independent of the pair's relative phase. (c) The stability of relative phase (mean of magnitude of relative phase differences) across rescaling and novel enclosure trials is independent of relative phase between cells in a pair (same data set and color-coding as in b).

Stability (attractiveness) of the 2D manifold of states

We established that the response of the grid-cell population is localized to the neighborhood of a 2D manifold, over extended periods of time and across varying external conditions that induce considerable changes in responses of grid cells. This localization is very likely due to internal recurrent dynamics, not a low-dimensional external input. What can we say about the dynamical stability of the 2D manifold?

Real-world dynamic systems are localized to their stable or attractive states. They are seldom (with vanishing probability) found at or near unstable fixed points, precisely because such points are unstable

and the smallest perturbation will drive the system away. Conversely, the dimensionality of state space occupied by the system is indicative of the dimensionality of the attractive states of the system. If a low-dimensional manifold is stable but is part of a higher-dimensional manifold of stable fixed points (Fig. 7a), then high-dimensional noise, however small, will cause the system to random-walk through the larger manifold⁴⁶. The system will consequently be found to inhabit any of the states across all dimensions of the stable higher-dimensional manifold. Grid cells are likely subject to high-dimensional internal noise: for example, stochastic vesicle release

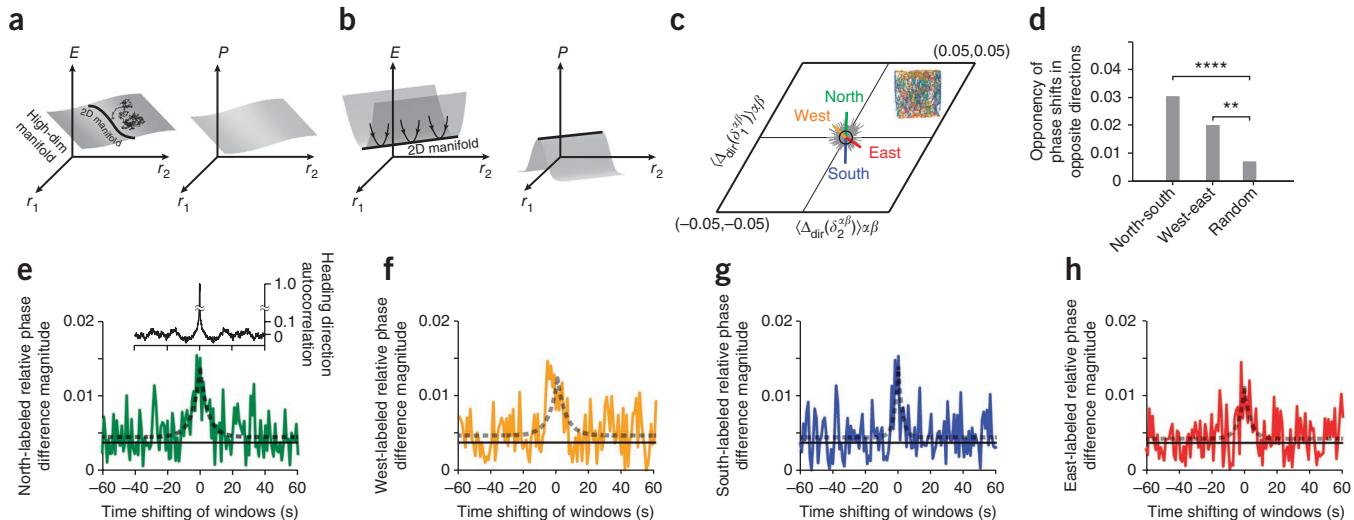


Figure 7 Evidence of external perturbation and attractor dynamics in grid-cell activity. (a,b) Schematic plots of the energy landscape (E ; left) and occupation probability density (P ; right) of a dynamic system, as a function of population activity (firing rates r_1, r_2 of neurons in the network). In a, the energy landscape has a flat plateau of dimension $D > 2$ in which the 2D manifold (depicted as a line) is embedded, and the system is equally likely to be found anywhere on the plateau. In b, there is a 2D energy valley (depicted as a line) or attractor, and the system will be localized to the valley bottom. (c) Green vector, relative phase between cell pairs for spikes from the northbound trajectory fragments (green parts of trajectory in inset), minus the relative phase obtained from the full trajectory, averaged across all (223) cell pairs. Blue, yellow and red vectors: corresponding differences in relative phase for south-, west- and east-bound trajectory fragments. Gray vectors: relative phases obtained by randomly segmenting the full trajectory into four sets of fragments of the same average lengths as the directional fragments but without direction specificity (this is the null hypothesis; Online Methods). Black circle, 1 s.d. of the null hypothesis distribution. P values for directional shifts in relative phase under null hypothesis: north ($P = 0.0004$), west ($P = 0.0444$), south ($P = 0.0003$) and east ($P = 0.0201$). (d) Relative phase shifts in opposite directions for opposite trajectory directions: north-south ($****P < 10^{-4}$), west-east ($**P < 0.0014$) (Online Methods). (e) Deviations in relative phase during northward motion as a spike selection window centered on the fragments (time-shift of 0) is slid, 1 s at a time, away from the fragment centers, to include spikes emitted just before or after northward movements. Inset, autocorrelation of movement direction in trajectories. (f-h) As in e, for south, east and west motion. Black solid line, radius of black circle from c. Gray dashed curve, best-fit Laplace distribution with zero mean.

in synapses⁴⁷ causes independent perturbations in every postsynaptic neuron (noise dimension $\sim N$, where N is the number of neurons). We analyzed spiking variability as a source of noise in grid cells (**Supplementary Fig. 3**). The fact that the grid-cell population nevertheless primarily occupied the neighborhood of a 2D manifold of states, in which cell-cell relationships are tightly conserved, suggests that the 2D manifold is attractive (**Fig. 7b**) and that locally, the attractive manifold has a dimension neither greater than nor smaller than 2.

In addition to the deduction above, that the grid-cell system is subject to perturbations off the attractor through internal stochastic dynamics, we found evidence that external perturbations, in the form of velocity inputs, drive the grid-cell system away from the 2D manifold. To see this, we selected all the spikes emitted by a cell during parts of the trajectory when the rat was headed ‘northward’ (north $\pm 45^\circ$), to form a north spatial activity map. We then computed the relative phase between a cell pair only for the northward maps of each pair. The difference between this relative phase and the relative phase computed over the full trajectory was the perturbation off the attractor that northward motion induced in the state. All grid cells in our analysis had negligible directional tuning (**Supplementary Fig. 2**); thus changes in relative phase cannot be attributed to a direct velocity modulation. We performed the same analysis for the south, east and west conditions. The resulting directional shifts in relative phase were significant (north, $P = 0.0004$; west, $P = 0.0444$; south, $P = 0.0003$; and east, $P = 0.0201$ under the null hypothesis of no shift in relative phase; see Online Methods for bootstrapped P values), compared to controls computed from similarly sized trajectory fragments unsorted by heading direction (**Fig. 7c,d**: data from the same cells in **Fig. 1c,d**). The shifts in relative phase between cells were consistent with the respective heading directions (**Fig. 7c**), suggesting that these shifts were indeed due to directional velocity input to the system. Thus, ongoing movements push the network away from the attractor by causing small deformations of the population pattern, in the form of a slight stretching of the population pattern along the direction of motion.

Finally, we examined the dynamics of perturbation by sliding the spike-selection windows in time relative to the centers of, for example, the northward trajectory fragments (Online Methods). The shift in relative phase decayed as the window was slid by a few seconds, and the decay time constant was very similar to the autocorrelation time for directional motion in the rat trajectories (**Fig. 7e**). Thus, the off-manifold (perturbed) components of the network state relaxed back to the 2D manifold on a time scale similar to or faster than the few-second time scale on which the perturbing input changed (**Fig. 7e–h**). The quick return of the system back to the 2D manifold of states after perturbation from the external velocity input is direct evidence that the 2D manifold is attractive.

DISCUSSION

We have shown that, over short durations and in familiar enclosures, the spatial responses of individual grid cells are well-characterized by a low-dimensional set of six parameters, with essentially the only difference between cells in the same network given by a 2D phase representing a rigid translation of the same basic response pattern. Over time and across environmental manipulations, the responses of individual grid cells change, and thus the parameters that describe their responses vary. Therefore, the responses of an individual grid cell are not described by a single set of six parameters.

Crucially, however, the dimensionality of the population response remains invariant. The responses of different grid cells are yoked

tightly together: over time or with experimental manipulation, when the spatial response patterns of the individual neurons change considerably, not only through rigid rotations but anisotropic and isotropic deformations of the grid pattern through stretching, the grid parameter ratios and relative phases between simultaneously recorded neurons in each network or module remained essentially constant (cells with distinct grid periods, that is, cells from different networks, could never have in common a stable spatial phase relationship even if all single-neuron grid parameters were perfectly stable, because the relative spatial phase between two perfectly periodic patterns of different frequency will necessarily precess relative to each other across cycles). In familiar enclosures, where stability may be attributed to external cues, we found that the cell-cell relationships are more stable than single-neuron responses, which argues against the external-cue hypothesis. Cell-cell relationships persist with the same fidelity immediately upon the rat entering a novel environment, even though landmark cues suddenly change and remain unassociated with specific locations, but responses of place cells continue to change, arguing strongly for stabilizing constraints in the grid-cell system.

Thus, population activity is confined to the immediate vicinity of a 2D manifold, across time and across conditions in different environments. Confinement of the system to a 2D manifold despite stochastic internal dynamics and external velocity-driven perturbations off the manifold suggest that the 2D manifold of states is an attractor.

Relationship to past work and implications for models

Data in ref. 24 (not included in our present analysis) afforded a glimpse that the responses of different grid cells are yoked together: the spatial phases of ~ 5 –9 simultaneously recorded grid cells shifted when the rat had been moved from one environment to another, and the shifts appeared to be of a similar magnitude and direction across cells. However, both environments were familiar, so that hippocampal representations were stable (albeit different), and external cues could be used to provide locational information. Thus, without a comparative analysis of variability of phase within and between cells in a single environment, or an experiment involving destabilized hippocampal representations and external cues, the qualitatively different possibility of feed-forward stabilization by hippocampal inputs or external cues cannot be ruled out.

Intracellular recordings in head-fixed mice navigating through virtual environments^{48,49} show that membrane potentials of grid cells exhibit a substantial slow depolarizing ramp at the firing field locations. This is consistent with excitatory (or disinhibitory) network drive that depolarizes neurons for the duration of a field. However, given that the intracellular recordings cannot distinguish between feed-forward and feed-back network inputs, and do not examine the structure and dimensionality of population activity, they do not provide specific evidence for continuous attractor models.

We have shown through population analysis that each grid-cell network is localized to a 2D manifold and that the manifold is attractive, which constitutes specific and direct evidence in support of continuous attractor network models of grid-cell activity^{12–15,22}.

To be consistent with our findings, models of populations of grid cells need to include recurrent interactions that constrain the system to lie on a 2D continuous attractor. The translation-invariant recurrent connectivity patterns described in refs. 2,4,12,14,15,22 are examples of such an interaction. Mechanisms for generating stretched or squeezed grid-cell tuning curves, as in rescaled familiar environments and novel environments, while maintaining low-dimensional continuous attractor dynamics in the population are described in

ref. 15. Within the constraint of 2D continuous attractor dynamics, however, models of grid-cell networks may be quite different: some consist of a single recurrent network with a 2D attractor, in which the grid cells integrate velocity inputs^{12,14,15}. In others, the 2D dynamics of grid cells arises from the feed-forward summation of inputs, from two 1D ring attractors, each of which integrates one component of rat velocity⁵⁰. In future work, it will be interesting to distinguish experimentally between such alternatives.

To conclude, our analysis contributes strong new evidence (see also refs. 10,11,19) supporting the idea that the brain uses low-dimensional continuous attractor dynamics in its integration and memory functions.

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 are grateful to D. Tank for insightful comments. We thank members of the Moser laboratory for public release of some of their data sets, which we used in this study. I.R.F. is funded as a Sloan Foundation Fellow, a Searle Scholar and a McKnight Scholar. We acknowledge funding from the Wellcome Trust, UK and the Office of Naval Research through ONR MURI N00014-10-1-0936 and the ONR-Young Investigator Program to I.R.F.

AUTHOR CONTRIBUTIONS

K.Y. performed the analysis. K.Y., M.A.B., C.B., N.B. and I.R.F. contributed ideas and plans for the analysis. C.B. and R.H. collected some of the data presented. K.Y. and I.R.F. wrote the paper with input from M.A.B., C.B. and N.B.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

- Hopfield, J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
- Ben-Yishai, R., Bar-Or, R. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA* **92**, 3844–3848 (1995).
- Seung, H.S. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA* **93**, 13339–13344 (1996).
- Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
- Seung, H.S. & Lee, D. The manifold ways of perception. *Science* **290**, 2268–2269 (2000).
- Latham, P.E., Deneve, S. & Pouget, A. Optimal computation with attractor networks. *J. Physiol. Paris* **97**, 683–694 (2003).
- Cannon, S.C. & Robinson, D. Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *J. Neurophysiol.* **57**, 1383–1409 (1987).
- Blair, H.T. & Sharp, P.E. Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. *J. Neurosci.* **15**, 6260–6270 (1995).
- Samsonovich, A. & McNaughton, B.L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17**, 5900–5920 (1997).
- Taube, J.S., Müller, R.U. & Ranck, J.B. Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990).
- Aksay, E., Gamkrelidze, G., Seung, H.S., Baker, R. & Tank, D.W. *In vivo* intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci.* **4**, 184–193 (2001).
- Fuhs, M.C. & Touretzky, D.S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
- Burak, Y. & Fiete, I. Do we understand the emergent dynamics of grid cell activity? *J. Neurosci.* **26**, 9352–9354 (2006).
- Guenella, A., Kiper, D. & Verschure, P. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.* **17**, 231–240 (2007).
- Burak, Y. & Fiete, I.R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).
- Romo, R., Brody, C.D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
- Miller, P., Brody, C., Romo, R. & Wang, X.-J. A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb. Cortex* **13**, 1208–1218 (2003).
- Ringach, D.L., Hawken, M.J. & Shapley, R. Dynamics of orientation tuning in macaque primary visual cortex. *Nature* **387**, 281–284 (1997).
- Aksay, E. *et al.* History dependence of rate covariation between neurons during persistent activity in an oculomotor integrator. *Cereb. Cortex* **13**, 1173–1184 (2003).
- Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
- O'Keefe, J. & Burgess, N. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* **15**, 853–866 (2005).
- McNaughton, B.L., Battaglia, F.P., Jensen, O., Moser, E.I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nat. Rev. Neurosci.* **7**, 663–678 (2006).
- Burgess, N., Barry, C. & O'Keefe, J. An oscillatory interference model of grid cell firing. *Hippocampus* **17**, 801–812 (2007).
- 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).
- Muller, R. & Kubie, J. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *J. Neurosci.* **7**, 1951–1968 (1987).
- Quirk, G.J., Muller, R.U. & Kubie, J.L. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J. Neurosci.* **10**, 2008–2017 (1990).
- Bostock, E., Muller, R. & Kubie, J. Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* **1**, 193–205 (1991).
- Markus, E.J. *et al.* Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J. Neurosci.* **15**, 7079–7094 (1995).
- Wood, E., Dudchenko, P., Robitsek, R. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
- Barry, C., Hayman, R., Burgess, N. & Jeffery, K.J. Experience-dependent rescaling of entorhinal grids. *Nat. Neurosci.* **10**, 682–684 (2007).
- Barry, C., Ginzberg, L., O'Keefe, J. & Burgess, N. Grid cell firing patterns signal environmental novelty by expansion. *Proc. Natl. Acad. Sci. USA* **109**, 17687–17692 (2012).
- Stensola, H. *et al.* The entorhinal grid map is discretized. *Nature* **492**, 72–78 (2012).
- Fiete, I.R., Burak, Y. & Brookings, T. What grid cells convey about rat location. *J. Neurosci.* **28**, 6858–6871 (2008).
- Artin, M. *Algebra* (Prentice Hall, 1991).
- Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312**, 758–762 (2006).
- Morris, R. Spatial localization does not require the presence of local cues. *Learn. Motiv.* **12**, 239–260 (1981).
- Steele, R. & Morris, R. Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus* **9**, 118–136 (1999).
- Frank, L., Stanley, G. & Brown, E. Hippocampal plasticity across multiple days of exposure to novel environments. *J. Neurosci.* **24**, 7681–7689 (2004).
- Wilson, M.A. & McNaughton, B.L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
- Karlsson, M.P. & Frank, L.M. Network dynamics underlying the formation of sparse, informative representations in the hippocampus. *J. Neurosci.* **28**, 14271–14281 (2008).
- Bonnevie, T. *et al.* Grid cells require excitatory drive from the hippocampus. *Nat. Neurosci.* **16**, 309–317 (2013).
- Pastoll, H., Solanka, L., van Rossum, M.C. & Nolan, M.F. Feedback inhibition enables theta-nested gamma oscillations and grid firing fields. *Neuron* **77**, 141–154 (2013).
- Couey, J.J. *et al.* Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat. Neurosci.* **16**, 318–324 (2013).
- Welinder, P.E., Burak, Y. & Fiete, I.R. Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus* **18**, 1283–1300 (2008).
- Sreenivasan, S. & Fiete, I.R. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci.* **14**, 1330–1337 (2011).
- Burak, Y. & Fiete, I.R. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl. Acad. Sci. USA* **109**, 17645–17650 (2012).
- Stevens, C.F. & Wang, Y. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature* **371**, 704–707 (1994).
- Domnisoru, C., Kinkhabwala, A.A. & Tank, D.W. Membrane potential dynamics of grid cells. *Nature* **495**, 199–204 (2013).
- Schmidt-Hieber, C. & Häusser, M. Cellular mechanisms of spatial navigation in the medial entorhinal cortex. *Nat. Neurosci.* **16**, 325–331 (2013).
- Welday, A., Shlifer, I., Bloom, M., Zhang, K. & Blair, H. Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *J. Neurosci.* **31**, 16157–16176 (2011).

ONLINE METHODS

Binning and rate maps. Cell-sorted spikes of putative grid cells from foraging rats were assigned to $1\text{ cm} \times 1\text{ cm}$ spatial bins derived from position samples taken at 50 Hz. The number of spikes assigned to each bin was divided by the rat's total dwell time in that bin, to remove the effects of inhomogeneous spatial exploration on estimating the probability of spiking at each location. This defined the rate map. Smoothed rate maps were generated by convolving the rate maps with a 2D Gaussian kernel ($\sigma = 4$ bins).

Autocorrelations and cross-correlations. To characterize the spatial response patterns of grid cells, we computed spatial autocorrelations from the smoothed rate maps of individual cells. To compare pairs of cells, we computed spatial cross-correlations from the smoothed rate maps of simultaneously recorded cell pairs. If the smoothed rate maps are R_1 and R_2 , both spatial correlations are generated as follows:

$$\gamma(u, v) = \frac{\sum_{x, y \in \Gamma} (R_1(x, y) - \bar{R}_{1,\Gamma})(R_2(x-u, y-v) - \bar{R}_{2,\Gamma})}{\sqrt{\sum_{x, y \in \Gamma} (R_1(x, y) - \bar{R}_{1,\Gamma})^2 \sum_{x, y \in \Gamma} (R_2(x-u, y-v) - \bar{R}_{2,\Gamma})^2}}$$

where $\gamma(u, v)$ is the correlation coefficient at the bin (u, v) , Γ is region of spatial overlap between R_1 and R_2 , and $\bar{R}_{i,\Gamma}$ is the mean of $R_i(x, y)$ within the region Γ . For autocorrelations, R_1 replaces R_2 .

Modified gridness score. The standard gridness score^{35,51} penalizes any regular grid pattern if it is not of an equilateral triangular pattern (Supplementary Fig. 1a,b). A novel scoring procedure was proposed⁵², with the aim of assigning high scores despite elliptical distortions in a hexagonal grid. However, it was limited by the fact that either the major or minor axis of the circumscribed ellipse was always assumed to pass through one of six nearest peaks. We used a modified gridness score that more generally allows both isotropic (equilateral triangle) and anisotropic (squeezed or stretched) grids to get a high score. For maximal sample size, and to test in principle the dimensionality of the grid-cell response, it is important to include cells with anisotropic triangular grids: it is critical to determine whether cells with deformed spatial activity patterns still lie on a low-dimensional manifold.

The modified gridness score is defined on the autocorrelogram of a cell. We first apply a transform on the correlogram that maps the central lattice cell (given by the six peaks nearest the center) into a regular hexagon (Supplementary Fig. 1c). This transform is determined by mapping the ellipse that circumscribes the central cell into a circle via a combined rotation, rescaling, and rotation transformation. If the number of nearest peaks (defined as the six or fewer peaks whose distance from the origin is less than two times the shortest distance from the origin to a neighboring peak) is less than or equal to four, the transformation is not applied.

Given the transformed autocorrelogram, we defined an annular region with inner radius R_i and outer radius R_o . We rotated the autocorrelation map in steps of 6° and computed the Pearson correlation between the rotated map and the original map with each confined to the annular region. The gridness for a given annular region is defined by the minimum difference between crests and troughs in rotated correlations:

$$\text{Gridness}(R_i, R_o) = \min\{\rho_{i,o}(60^\circ), \rho_{i,o}(120^\circ)\} - \max\{\rho_{i,o}(30^\circ), \rho_{i,o}(90^\circ), \rho_{i,o}(150^\circ)\}$$

where $\rho_{i,o}(\phi)$ is the correlation value when one map is rotated by angle ϕ relative to another over the annular region defined by R_i, R_o .

We did this for various values of R_i and R_o , letting R_i change from $0.5r$ to r and R_o change from $R_i + 1\text{ cm}$ to $1.5r$ (or to the maximum allowed value based on the autocorrelogram), each independently and in steps of 1 cm . r is the mean distance to the nearest six peaks from the center in the transformed autocorrelogram. The modified gridness of the cell is then defined as the maximum gridness score over these various annular regions.

Our results were not qualitatively changed if we used the former gridness scoring technique⁵²; they also did not qualitatively change if we used our technique with a higher threshold (threshold of 0.5).

Cell selection. In this paper, we analyzed data sets from four different sources^{20,30,31,35} (data^{20,35} available at <http://www.ntnu.no/cbm/gridcell>). A modified gridness score was computed for each grid cell, and cells with a gridness score less than zero were rejected (Supplementary Fig. 2). (When we restricted our cell sample based on the gridness score more commonly used in the past, the sample became smaller but the qualitative results remained unchanged. Indeed, because the uncertainty in grid parameter estimation drops for cleaner spatial activity patterns, the cell-cell relationships and ratios become slightly tighter.)

For pairwise analyses, we used all possible cell pairs that were simultaneously recorded in the same individual rat and had a common grid period whose maximum difference between cells was smaller than 30% in the familiar enclosure (provided both members were above threshold on the modified gridness score). Cells above threshold that had no simultaneously recorded cells to pair with that also met the threshold for cell-cell comparisons, were also rejected from further analysis. Data from 223 cell pairs are shown in Figure 1c,d, from many experiments (24 from ref. 20; 97 from ref. 35; 12 from ref. 30; 90 from ref. 31). Data from 84 cell pairs (75 cells) are shown in Figures 1f and 2a,b, from both resizing³⁰ and novel enclosure experiments³¹, recorded in both trials 1 and 5. Data from the same cells in Figure 1c are shown in Figure 7c-h. Note that all grid cells from all data sets that met our gridness criterion happened to have extremely weak directional sensitivity (Watson U2 score < 10, by the scoring method of refs. 35,52; Supplementary Fig. 2). In Figures 3–5, cell pairs that have lower gridness than the fixed threshold in both familiar enclosures or in more than two intermediate trials were discarded, which results in 7 cell pairs in Figure 3 and 24 cell pairs in Figures 4 and 5.

Template matching algorithm to estimate grid parameters. We identified local maxima in the autocorrelogram and noted the coordinates and heights of the peaks. The local maxima whose heights were lower than 1% of the height of the global maximum were not considered as 'local peaks'. We then generated a 2D template lattice. Any 2D periodic lattice centered at the origin is fully specified by the magnitudes (λ_1, λ_2) and orientations (ψ_1, ψ_2) = ($\psi, \theta - \psi$) of two basis vectors (Fig. 1c). The angles are measured from the x axis. The template lattice was generated by populating the explored spatial environment by vertices whose locations are determined by the basis vectors and their translations. The lattice parameters were determined by finding values that minimize a cost function that quantifies the fit between the template and the data. The cost function is given by the sum of the squared distances from every data peak to the nearest vertex in the template lattice, weighted by the autocorrelation amplitude at that data peak:

$$C = \sum_{i=1}^n w_i \|p_i - v_i\|_2$$

where p_i is the (x, y) location of the data peak, v_i is the vector for the point in the lattice nearest to the i th data peak, w_i is the correlation coefficient at p_i , and n is the total number of peaks in the autocorrelogram.

The central peak of a cross-correlogram is typically not at the origin but shifted by some displacement vector \vec{d} from the origin. Thus, two additional parameters $\vec{d} = (d_x, d_y)$ were estimated simultaneously with (λ_1, λ_2) and (ψ_1, ψ_2) by minimizing the same cost function to find the best-fit template lattice of the cross-correlogram.

Relative phase, relative phase difference and phase magnitude. Let $\vec{\phi}^\alpha$ represent the phase of cell α , where the component ϕ_i^α is the phase along the i th lattice basis vector. The relative phase between cells α and β is then

$$\delta^{\alpha\beta} \equiv (\vec{\phi}^\alpha - \vec{\phi}^\beta) \bmod 1$$

where 'mod 1' is understood to apply to each component. The relative phase between a pair of cells is closely related to the shift \vec{d} in the peak of their spatial cross-correlogram. If the two cells have precisely the same lattice parameters,

then \vec{d} will equal the relative shift of the two cells' grid patterns. The oblique projection of \vec{d} onto the two primary lattice vectors \vec{e}_1 and \vec{e}_2 produces the components (d_1^{proj}, d_2^{proj}) . When the components are normalized by λ_1 and λ_2 , respectively, and considered modulo 1, we get the relative phase:

$$\begin{aligned} (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta}) &= (\phi_1^\alpha - \phi_1^\beta, \phi_2^\alpha - \phi_2^\beta) \bmod 1 \\ &= (d_1^{proj}/\lambda_1, d_2^{proj}/\lambda_2) \bmod 1 \end{aligned}$$

If the relative phase between the same two cells is measured in two different conditions (such as in distinct trials), we can define the relative phase difference as the difference in $\vec{\delta}^{\alpha\beta}$ between the two conditions. We will denote this as $\Delta_C(\vec{\delta}^{\alpha\beta})$, where C labels the condition. Relative phase differences are equivalent modulo 1 and representing them on the unit lattice cell with components in the interval [0,1) has the consequence that values close to zero will appear in the four corners of the unit cell (**Supplementary Fig. 6a,b**). To avoid this and map values close to zero together, we remapped the unit lattice cell to the equivalent unit cell with components in the interval [-0.5,0.5). Thus, the relative phase difference at distinct times t_1, t_2 is given by:

$$\Delta_t(\vec{\delta}^{\alpha\beta}) = f[(\vec{\delta}_{t_1}^{\alpha\beta} - \vec{\delta}_{t_2}^{\alpha\beta}) \bmod 1]$$

$$f(\vec{x}) = \begin{cases} (x_1 - 1, x_2 - 1) & \text{if } x_1 \geq 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1 - 1, x_2) & \text{if } x_1 \geq 0.5 \text{ and } x_2 < 0.5 \\ (x_1, x_2 - 1) & \text{if } x_1 < 0.5 \text{ and } x_2 \geq 0.5 \\ (x_1, x_2) & \text{if } x_1 < 0.5 \text{ and } x_2 < 0.5 \end{cases}$$

where $\vec{\delta}_{t_1}^{\alpha\beta} = (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta})_{t_1}$ and $\vec{\delta}_{t_2}^{\alpha\beta} = (\delta_1^{\alpha\beta}, \delta_2^{\alpha\beta})_{t_2}$ are relative phases between cells α, β at time t_1 and t_2 . The function $f(\vec{x})$ maps (x_1, x_2) within the interval [0,1] into [-0.5,0.5) by subtracting 0.5 from each component, if that component is greater than or equal to 0.5 (**Supplementary Fig. 6c**).

For relative phase, we took the relative 'phase magnitude' to be of the form $\|\vec{\delta}^{\alpha\beta}\| = (\delta_1^{\alpha\beta})^2 + (\delta_2^{\alpha\beta})^2 + \delta_1^{\alpha\beta} \delta_2^{\alpha\beta}$, and similarly for relative phase difference magnitudes.

Error analysis. The error bars (ϵ) in **Figures 3c,d** and **4d,e** for the ratio of two uncertain quantities λ_1/λ_2 were given by the standard method of error propagation given a covariance matrix Σ for the uncertainties in λ_1 and λ_2 . Σ was estimated via bootstrap resampling: given an original spike (discharge) map of M total spike locations, we created 100 new spike maps of M total spikes each, by picking spike locations from the original map one at a time, at random, and with replacement. Next, we used these spike maps to generate rate maps using the same procedure as for the original, with the original trajectory data (that is, with a normalization given by the same visitation frequency as the original spike map) and estimate grid parameters from the spatial autocorrelogram with the template matching algorithm. This procedure generated 100 samples of the grid parameters, from which we computed the covariance matrix as an estimate for Σ . The error for the ratio λ_1/λ_2 is given by:

$$\frac{\lambda_1}{\lambda_2} \pm \epsilon = \frac{\lambda_1}{\lambda_2} \pm \frac{\lambda_1}{\lambda_2} \sqrt{\frac{\sum_{11}}{\lambda_1^2} + \frac{\sum_{22}}{\lambda_2^2} - 2 \frac{\sum_{12}}{\lambda_1 \lambda_2}}$$

The radius (r) of the red circles in **Figures 2a,b, 3f** and **5**, signifies the measurement error of the differences across trials, in phase per cell (**Fig. 2b**) or in relative phase per cell pair (**Figs. 2a, 3f** and **5**). This measurement error is estimated by applying the same bootstrapping technique to every cell (pair), computing now the (relative) phase difference as defined above for each bootstrap sample, and then subtracting the mean of sampled (relative) phase differences for each cell (pair). The error in (relative) phase difference across (pairs of) cells is given by collapsing all the bootstrap samples of zero-mean (relative) phase differences and fitting a 2D Gaussian with a multiple of the identity matrix as a covariance to those samples (by expectation-maximization algorithm). The radius in the figures is the square root of the estimated covariance.

Analysis of velocity-driven perturbation. We defined cardinal direction labels as north (45°–135°), west (135°–225°), south (225°–315°) or east (315°–45°). Each time point t in the rat's trajectory was labeled by the rat's velocity at that

time (given by the vector difference quotient between the position at $t + \Delta t$ and t). Spikes that occurred at time t were labeled by the trajectory direction label at that time. This produced four sets of trajectories and for each cell, four corresponding sets of spike maps, labeled by north, south, east and west. For each direction, we generated rate maps and relative phases (as we did earlier for the full-trajectory and full-spike maps). For each cell pair, we thus obtained four 'direction-labeled relative phases' given by $\vec{\delta}_{dir}^{\alpha\beta} = (\vec{\phi}_{dir}^\alpha - \vec{\phi}_{dir}^\beta)$, where $dir \in \{\text{north, south, east, west}\}$ and α, β refer to the cells. We denoted the relative phase for the full spike maps, obtained earlier, as $(\vec{\delta}_{full}^{\alpha\beta})$. The 'direction-labeled relative phase differences' for each cell pair were defined as the differences between the direction-labeled relative phase and the full relative phase (that is, $\Delta_{dir}(\vec{\delta}^{\alpha\beta}) = \vec{\delta}_{dir}^{\alpha\beta} - \vec{\delta}_{full}^{\alpha\beta}$). The mean value of the direction-labeled relative phase difference, with the average taken over all simultaneously recorded cell pairs, is shown in **Figure 7c**. It is written as $\langle \Delta_{dir}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$.

To assess whether the shifts in relative phase as a function of motion direction were meaningful, we created a null hypothesis distribution by segmenting the full trajectory into continuous pieces of a length consistent with the lengths of the continuous pieces generated in the direction-labeled trajectory segmentation described above (the fragment length was set equal to the correlation time of the rat's heading direction; a representative value of the heading direction correlation time across experiments is approximately 0.65 s, and we chose 1.6 s to provide a window of at least two time constants). However, the segmentation did not correlate with movement along a specific direction. The segments resulting from this process were divided, randomly, into four sets of equal size. We considered one such set of directionally mixed or random segments, and labeled it ' R ', in contrast to the 'north, south, east and west' labels of the directional trajectory segments. The relative phase difference for this one set of directionally mixed trajectory segments was denoted $\Delta_R(\vec{\delta}^{\alpha\beta}) = \vec{\delta}_R^{\alpha\beta} - \vec{\delta}_{full}^{\alpha\beta}$. Averaging this relative phase difference for one set of trajectory segments, across all simultaneously recorded cell pairs, gives one sample of $\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$, which can be seen as one gray vector in **Figure 7c**. Repeating this procedure 400 times produced the 400 different gray vectors in **Figure 7c**. The lengths of these vectors represent the expected magnitude of deviation from the full phase simply due to subsampling errors from subdividing the full trajectory into four sets, independent of directional effects. Thus, these vectors provide the null hypothesis for no directional motion effects on relative phase. We approximated the sampled distribution of vectors $\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$ by a symmetric 2D Gaussian with s.d. σ^2 . Using this distribution, the P value of the mean direction-labeled relative phase differences, $l = \langle \Delta_{dir}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$, is given by:

$$P = \int_2^\infty \exp(-\frac{r^2}{2\sigma^2}) dr$$

In **Figure 7d**, we tested the opponency of shifts in relative phase, for opposite movement directions, by computing the magnitude of the difference between opposing mean labeled relative phase differences, $\|\langle \Delta_{North}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} - \langle \Delta_{South}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} \|$ and $\|\langle \Delta_{West}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} - \langle \Delta_{East}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} \|$. The magnitude of opponency expected under the null hypothesis is given by the expected magnitude of the differences between the gray bars in **Figure 7c**, that is, by averaging $\|\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}^i - \langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}^j \|$, where i and j index values of $\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}$ from the 400 samples shown in **Figure 7c**. This gives the height of the "random" bar in **Figure 7d**. The fraction of samples with $\|\langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}^i - \langle \Delta_R(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta}^j \|$

that is larger in size than $\|\langle \Delta_{North}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} - \langle \Delta_{South}(\vec{\delta}^{\alpha\beta}) \rangle_{\alpha\beta} \|$ gives the P value that the direction-labeled relative phase differences can be accounted for by the null hypothesis.

Analysis of relaxation from the perturbation. We tracked the mean shifts in relative phase for each cardinal direction as a function of time to examine the dynamics of perturbation off the attractor. First, we used the northward direction-labeled trajectory fragments to define a corresponding set of time windows to select spikes for analyzing relative phase along that movement direction. Next, we slid the same set of time windows forward, so that instead of being centered in time at the northward fragment, they were centered 1 s after the center (in time) of each northward fragment, and so on, in steps of 1 s. We did the same in the opposite direction, sliding the windows back. For each position of the windows, we computed the relative phase for spikes obtained from those windows

and subtracted from it relative phase obtained from the full trajectory. For each time shift, we averaged the result across all 223 cell pairs in our data set.

Analysis of stochastic dynamics. To investigate the variance of spiking in grid cells, we computed the distribution of the interspike interval using firing times $\{\tau_i\}$ that were rescaled from the original firing times $\{t_i\}$ via $\tau_i = \int_0^{t_i} \lambda(t)dt$, where $\lambda(t)$ is the time varying firing rate of the cell. $\lambda(t)$ was approximated by sliding a rectangular window function along the spike train. The coefficient of variation in

the left part of **Supplementary Figure 3** was derived from the rescaled interspike interval with $\Delta t = 0.5$ s. The coefficient of variation as a function of window size for four representative cells is shown on the right in **Supplementary Figure 3**.

51. Langston, R.F. *et al.* Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
52. Brandon, M.P. *et al.* Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* **332**, 595–599 (2011).