Three types of remapping with linear decoders: a population-geometric perspective

Guillermo Martín-Sánchez, Christian K. Machens*, William F. Podlaski*,†

Champalimaud Centre for the Unknown,
Champalimaud Foundation, Lisbon, Portugal

March 14, 2025

Abstract

Hippocampal remapping, in which place cells form distinct activity maps across different environments, is a robustly-observed phenomenon with many theories and interpretations. Some theories view remapping as the mechanism behind reduced interference between competing spatial memories, whereas others associate it with changes in an underlying latent state representation. However, it remains unclear how these interpretations of remapping relate to one another, and what types of activity changes they are compatible with. To shed some light on these questions, here we propose a neural coding and population geometry perspective to unify and elucidate the mechanisms behind remapping. Assuming that hippocampal population activity can be understood through a linearly-decodable latent space, we show that there are three possible mechanisms to induce activity changes that resemble remapping. Remapping can be due to (i) a true change in the mapping between neural and latent space, (ii) modulation of activity due to non-spatial mixed selectivity of place cells, or (iii) neural variability outside of the latent space that reflects a redundant code. We simulate and visualize examples of these remapping types in a network model, and relate the resultant remapping behavior to various models and experimental findings in the literature. Overall, our work serves as a unifying framework with which to visualize, understand, and compare the wide array of theories and experimental observations about remapping, and may serve as a testbed for understanding neural response variability under various experimental conditions.

Introduction

Place cells in areas CA1 and CA3 of the hippocampus exhibit localized, spatial firing patterns, which are thought to contribute to a cognitive map of space (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978). However, hippocampal activity is known to be modulated by many other variables besides the animal's position (Jeffery, 2007; O'Keefe and Krupic, 2021) — this includes other types of spatial information (e.g., head direction), sensory information, task-related information, internal states, and changes in context. Perhaps the most striking manifestation of this modulation is remapping — the phenomenon in which place cell activity appears to form distinct representations, even in response to relatively minor environmental changes (Colgin et al., 2008; Fenton, 2024). Remapping comes in different flavors — it can be 'complete', in which the activity appears to change randomly and globally across an entire environment (Muller and Kubie, 1987; Muller et al., 1987; Bostock et al., 1991), or it may be 'partial', in which some neurons change their spatial preference, with others only exhibiting minor changes in firing rate (i.e., rate remapping; Quirk et al. (1990); Shapiro et al. (1997); Tanila et al. (1997); Leutgeb et al. (2005)). The implications of remapping on the hippocampal code are still unclear.

Historically, explanations for remapping have focused on the perspectives of spatial navigation and memory (Colgin et al., 2008; Soldatkina et al., 2021). The definitive theory of such a perspective is that of the "multi-chart" attractor model (Samsonovich and McNaughton, 1997; Agmon and Burak, 2023), in which the hippocampal network is hypothesized to store a set of spatial maps, or "charts", as attractors of the network dynamics. Importantly, this theory dictates that each chart should be random and uncorrelated with all others in order to promote high memory capacity and reduced interference between different maps (Treves and Rolls, 1994). The multi-chart model thus predicts random global remapping, consistent with several experimental studies (e.g., Fyhn et al. (2007); Alme et al. (2014)). However it is less clear how this spatial memory perspective can account for partial or rate remapping, mixed-selective and purely non-spatially selective cells (e.g.,

*Co-senior authors
†Correspondence: william.podlaski@research.fchampalimaud.org

Wood et al. (1999); MacDonald et al. (2011); Aronov et al. (2017); Gauthier and Tank (2018); Sun et al. (2020)), as well as place cells with multiple fields (Fenton et al., 2008; Rich et al., 2014).

An alternative perspective generalizes the notion of the cognitive map beyond physical space, and attempts to reconcile the spatial view with other known functions of the hippocampus such as episodic memory (Eichenbaum and Cohen, 2014). In recent years, this has manifested into theories suggesting that the hippocampus builds a latent state space of the environment in order to solve tasks (Whittington et al., 2022). While this latent space view predicts similar spatial tuning profiles as the spatial memory perspective, it has the benefit of providing a more explicit functional role for place cell mixed selectivity, and is also compatible with non-spatially-selective cells and multiple fields. This perspective includes some theories that view place cells as representing a conjunction of space and other (e.g., sensory) variables (Whittington et al., 2020), or others that view space as implicitly constructed through a latent sequence of sensory inputs (Benna and Fusi, 2021; Recanatesi et al., 2021; Raju et al., 2024; Levenstein et al., 2024; Wang et al., 2024). This perspective can also account for different types of remapping (Sanders et al., 2020), and may lead to more structured, non-random remapping effects as compared to the predictions of the spatial memory perspective (Whittington et al., 2020). However, such remapping questions have been relatively less explored from this perspective, and it remains to be seen how specific its predictions are, as well as their compatibility with the spatial memory picture.

While place fields and remapping have traditionally been understood at the single-neuron level, the recent shift in focus towards a population-level view of neural coding (Saxena and Cunningham, 2019; Chung and Abbott, 2021; Langdon et al., 2023) may help to clarify and unify these varying perspectives on hippocampal activity and function (Eichenbaum, 2018). The population view has been applied to hippocampal representations (Meshulam et al., 2017; Gava et al., 2021; Nieh et al., 2021), but its implications for remapping are not yet clear (but see Fenton (2024); see Discussion). In this work, we use a neural coding and population geometry perspective to develop a theoretical framework to understand hippocampal remapping. Under the assumption of a linearly-decodable latent space, we show that there are three fundamental mechanisms that can explain remapping, which we term encoder-decoder, mixed-selective, and nullspace remapping. We then demonstrate each of these types of remapping, and explain how various experimental and theoretical findings partition into them, including the two general perspectives mentioned above. Rather than ruling one mechanism out in favor of another, we suggest that all three remapping mechanisms are likely to be accurate depictions of hippocampal activity changes under different conditions, and propose our framework as a useful perspective with which to understand the variability of hippocampal representations within and across environments.

Results

We start from the standard setup of spatial navigation and remapping, in which an animal navigates through several environments or contexts while hippocampal place cell activity is monitored (e.g., schematized for two 1-d linear track environments A and B in Fig. 1a,b). We consider that these environments are represented by the animal through a common set of environmental variables — this includes (1-d or 2-d) spatial position, $\mathbf{p}$, but also may include one or more additional cognitive variables, $\mathbf{c}$, such as internal (e.g., behavioral) or external (e.g, sensory) states. For simplicity, we assume that these cognitive variables have a deterministic relationship to position, and we thus denote them as $\mathbf{c}(\mathbf{p})$. From a neural coding perspective, our aim will be to model how the environmental variables are *encoded* into hippocampal population activity across environments, and analogously how estimates of these variables may be *decoded* from such activity (Fig. 1a). We will pursue this aim in two steps. First, we will need to specify how firing rates vary as a function of position in each individual environment. To do so, we will define a particular constrained model for how the environmental variables relate to neural activity through an intermediate latent space, resulting in a firing-rate map, $\mathbf{r}(\mathbf{p})$. Second, we will then apply this model to understand remapping by studying how the constraints that we impose in turn control activity changes across environments (i.e., $\mathbf{r}^A(\mathbf{p})$ vs. $\mathbf{r}^B(\mathbf{p})$ in Fig. 1b, right).

Mapping spatial position to neural activity through a latent state space

There are numerous possible encoder and decoder mappings that could describe how environmental variables such as spatial position may relate to neural activity (Fig. 1a; Zhang et al. (1998); Dayan and Abbott (2005); Quiroga and Panzeri (2009); Mathis et al. (2024); see Discussion). The central feature of our model is the introduction of a low-dimensional latent space, $\mathbf{z}(\mathbf{p})$, as an intermediate representation between the environmental variables and neural activity (Fig. 1c). This latent space can be interpreted as an internal representation of the environmental variables in the hippocampus. Building on this latent space representation, we will now specify two biologically-motivated mappings for our model: first, a fixed angular code between the environmental variables and latent space, and second, a flexible linear readout to latent space from neural activity (Fig. 1c). As we shall see, each of these two mappings has unique and important implications for the resultant model.

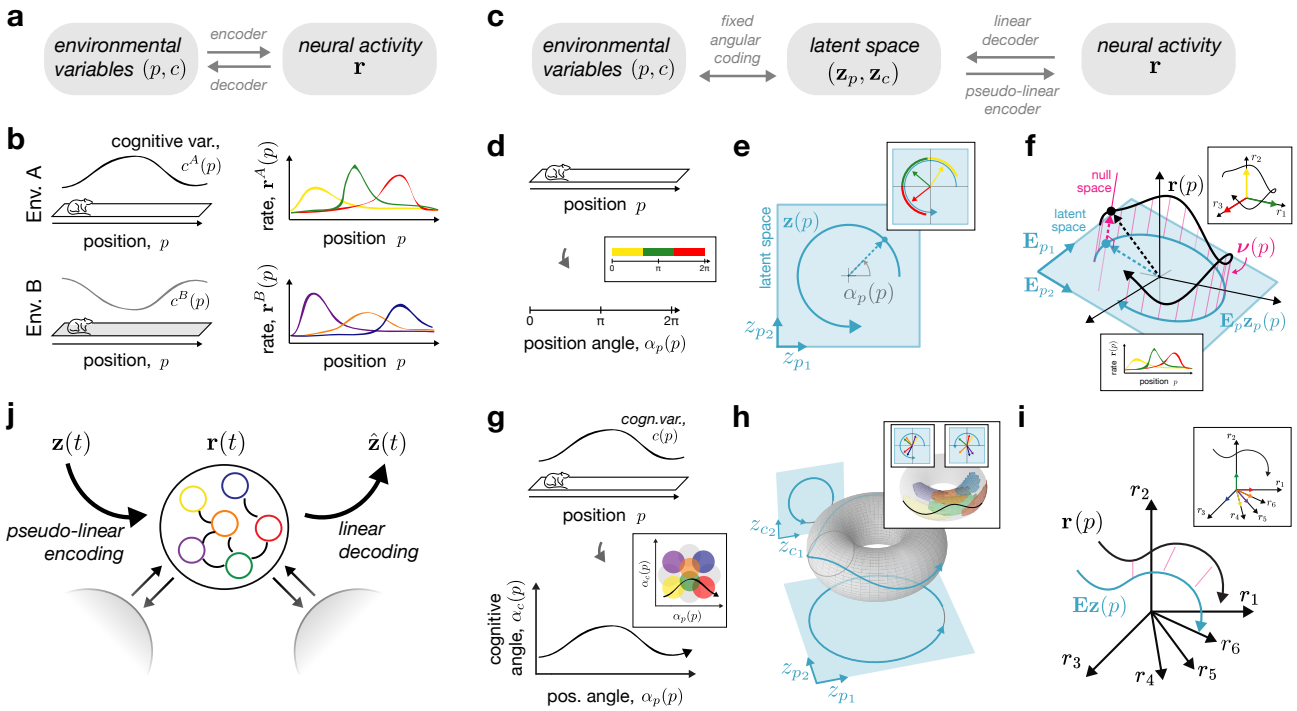

Figure 1: **A neural coding and population-geometric perspective on hippocampal place fields.** **a,b**: General overview of spatial coding and remapping; two linear track environments, A and B, are characterized by a common set of environmental variables (position, p, and cognitive variable(s), c, which may be environment-specific, $c^A(p)$ and $c^B(p)$); spatial and cognitive variables are related to place cell firing rate maps ($\mathbf{r}^A(p)$ and $\mathbf{r}^B(p)$) through encoder and decoder mappings. **c**: The constrained model considered in this work, featuring an intermediate latent space, $\mathbf{z} = (\mathbf{z}_p, \mathbf{z}_c)$, with two important assumptions: (1) fixed angular coding from environmental variables to latent space, and (2) linear decoding from neural activity to latent space. **d**: 1-d position p is encoded as the position angle $\alpha_p(p)$. **e**: This gives rise to circular trajectories in a 2-d latent space $(z_{p_1}(p), z_{p_2}(p))$. **f**: The activity trajectory in neural state space ($\mathbf{r}(p)$, black) can be seen as the combination of a linear encoding of latent space ($\mathbf{E}_p\mathbf{z}_p(p)$, blue), plus a null space component ($\boldsymbol{\nu}(p)$, pink). **d,e,f insets**: Sequential place field activity (**f**, bottom inset) can be seen as the alignment of the trajectory with each neuron's axis (**f**, top inset); the linear decoder results in a rotation of each neural axis in latent space (**e**); each area of the latent trajectory (**e**) or angle space (**e**) is colored by the most active neuron (see Methods). **g,h,i**: Same as panels (**d,e,f**) but for a pair of position and cognitive variables (p, c), leading to the pair of angles $(\alpha_p(p), \alpha_c(p))$, a four-dimensional latent space with trajectories confined to a 4-d torus (shown as a 3-d embedding), and the same neural trajectory in rate space (black) composed of the combination of latent space (blue) and null space (pink). **g,h,i insets**: As above, neural tuning can be visualized in latent and angle space, this time resulting in localized surface patches for each neuron on the torus or in angle space (see Methods). **j**: Place-field representations are modeled in an autoencoder recurrent neural network (RNN) which follows the pseudo-linear encoder and linear decoder constraints of Eqs. 1 and 2 (see Methods).

First, the mapping between environmental variables and latent space accounts for the fact that the internal representation of spatial position and other quantities need not reflect Euclidean coordinates, but can be deformed or nonlinear. In fact, many previous constructive models of place and grid cells consider position as a curved space parameterized by one or more angles (Herz et al. (2017); see Discussion). We adopt this convention here both for spatial position and the other environmental variables (see Methods). Concretely, for a 1-d linear track, spatial position is described by an angle ($\alpha_p(p)$, Fig. 1d), which then maps onto a circular trajectory in two-dimensional latent space ($\mathbf{z}_p(p)$, Fig. 1e). A linear track with an additional cognitive variable $c(p)$ (e.g., odor concentration), can be described in two-dimensional angle space ($(\alpha_p(p), \alpha_c(p))$, Fig. 1g), which then maps onto a four-dimensional latent space $(\mathbf{z}_p(p), \mathbf{z}_c(p))$ with trajectories confined to a torus (shown as a 3-d nonlinear embedding of the 4-d space in Fig. 1h). This picture then generalizes to higher dimensions, with d environmental variables mapping onto hypertoroidal trajectories in $Z = (2d)$-dimensional latent space. While the choice of angular coding is not strictly necessary for our theory (see Discussion), we shall see that it is a simple way of generating localized place-like firing fields. Furthermore, because this mapping specifies the "coordinate system" of the internal representation, we consider it to be fixed across all environments.

Next, the mapping between the latent space and neural activity specifies how this internal latent representation is embedded into neural state space (Fig. 1f). While many choices exist for this mapping, our key

assumption here is that latent variable estimates, $\hat{\mathbf{z}}$, can be *linearly decoded* from neural activity as

$$\hat{\mathbf{z}}(\mathbf{p}) = \mathbf{D}\mathbf{r}(\mathbf{p}), \tag{1}$$

where $\mathbf{D}$ is a $Z \times N$ linear decoding matrix that maps the N neurons into the Z latent variables. This choice is analogous to a linear population vector code (Dayan and Abbott, 2005), and can be viewed as projecting the full N-dimensional neural trajectory onto a linear subspace. Considering the two-dimensional positional latent space embedded into a three-dimensional neural state space (Fig. 1f), we see that the linear decoder effectively collapses the full neural trajectory (black curve) onto a 2-d plane (blue curve), thereby removing variability in the third, nullspace direction (pink). This intuition also holds for larger numbers of environmental variables and neurons, where the neural, latent, and null spaces may be higher dimensional (Fig. 1i). This linear decoding assumption, though simplistic, is plausible in that its weighted sum operation is similar to what a downstream neuron can read out from the network (Fusi et al. (2016); Keemink and Machens (2019); see Discussion).

There are two additional reasons that make linear decoding a particularly convenient choice. First, it provides a straightforward way of visualizing neural firing fields in the space of environmental variables. To see this, we can return to the picture of two-dimensional latent space in three-dimensional neural state space (Fig. 1d-f), and observe that the sequential place-field activity (Fig. 1f, bottom inset) can be explained by the temporary alignment of the neural trajectory with each neuron's axis (Fig. 1f, top inset, colored arrows). The linear decoder then projects these axes onto latent space (Fig. 1e, inset, colored arrows). Then, we employ a method to color each section of the latent trajectory (Fig. 1e, inset, multi-colored circle) or environmental angle space (Fig. 1d, inset) with the most active neuron, resulting in a map of each neuron's tuning preference overlayed on the environmental and neural trajectories (see Methods). From this view, it becomes clearer why an angular encoding is so suitable for generating place fields, as the curved latent trajectory temporarily aligns with particular neurons' preferred tuning vectors in localized areas. Similar visualizations can be made for higher-dimensional representations (Fig. 1g-i, insets), resulting in a tessellated pattern of tuning preferences on the toroidal manifold in latent space (Fig. 1h, inset, colored circles), and in angle space (Fig. 1g, inset, colored circles). This visualization allows us to directly relate the environmental variables in angle coordinates (Fig. 1g, inset, trajectory and colored circles) to the firing rates in each environment (Fig. 1f, bottom inset), and we will make extensive use of it (see Methods; see Nieh et al. (2021) for a similar visualization).

The second convenience of linear decoding is that, under the assumption of accurate decoding ($\hat{\mathbf{z}}(\mathbf{p}) \approx \mathbf{z}(\mathbf{p})$), it allows us to specify a more explicit *encoding* model by inverting Eq. 1. We can write this nonlinear encoder, which we will refer to as a *pseudo-linear encoder* (see Methods), as

$$\mathbf{r}(\mathbf{p}) = \mathbf{E}\mathbf{z}(\mathbf{p}) + \boldsymbol{\nu}(\mathbf{z}(\mathbf{p})), \tag{2}$$

where $\mathbf{E} = \mathbf{D}^{+}$ is the right pseudo-inverse of the decoding matrix from Eq. 1, and $\boldsymbol{\nu}(\cdot)$ is an arbitrary nonlinear function in the nullspace of $\mathbf{D}$ that depends on the particular mechanistic network model (see below). Given that typically there will be more neurons than latent variables ($N > Z$), the first term on the right-hand-side of Eq. 2 constrains Z dimensions of $\mathbf{r}$ to be linearly related to $\mathbf{z}$, and the second term on the right-hand-side specifies the other $N - Z$ dimensions and accounts for any nonlinearities in the encoding (e.g., non-negativity; Keemink and Machens (2019); Okazawa et al. (2021)). Thus, rather than viewing the neural trajectory in Fig. 1f as being projected into latent space, we can also view it from a constructive, encoding perspective, in which it is composed of the latent trajectory embedded in neural state space (Fig. 1f, blue), plus the addition of an orthogonal, nullspace component (Fig. 1f, pink). As before, similar intuitions apply for higher-dimensional representations, where both the latent and null spaces may be high dimensional (Fig. 1i).

The neural coding perspective leads to a natural and simplified way of modeling the place field representations that result from this latent space formulation, which is to consider an autoencoder network. This network nonlinearly encodes a latent variable input, $\mathbf{z}$, and then enables an estimate $\hat{\mathbf{z}}$ to be linearly decoded from the activity (Fig. 1j). For reasons of biological plausibility, we will make use of an established recurrent neural network (RNN) model (see Methods; Boerlin et al. (2013); Barrett et al. (2013); Calaim et al. (2022)), which comes with additional biologically-plausible characteristics (see Discussion). Importantly, this model satisfies the encoding and decoding constraints above in Eqs. 1 and 2, and in particular, it specifies the nonlinear nullspace term ($\boldsymbol{\nu}$) by constraining the neural firing rates to be non-negative with a ℓ_1-cost on activity (see Methods). We stress that this autoencoder perspective does not require thinking about the latent variables as the true inputs and outputs of the hippocampus, but instead serves as a useful abstract model of hippocampal representations consistent with internally-generated (e.g., attractor) dynamics or other computations (see Discussion).

Three types of remapping

Now that we have established a constrained model for how environmental variables map to neural activity through latent space, we can consider how this activity remaps across environments. But what exactly is remapping? From an experimental point of view, remapping refers to the observed phenomenon that spatial

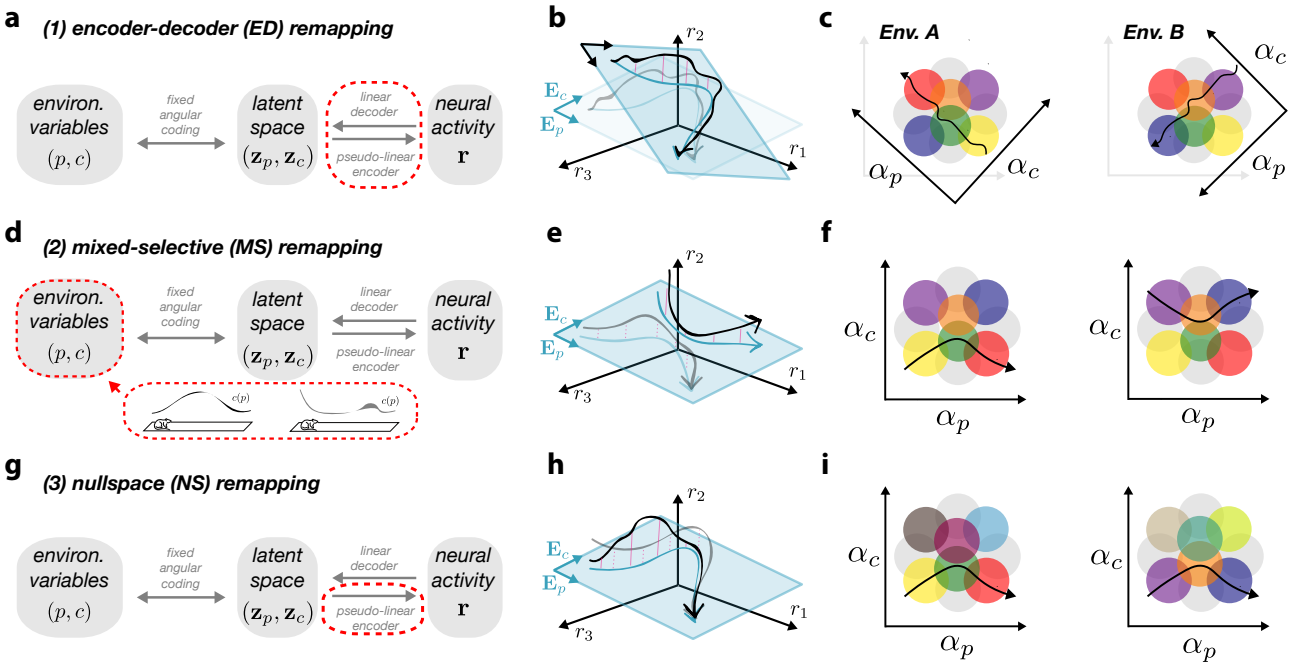

Figure 2: **Three types of remapping with linear decoders. a-c**: encoder-decoder (ED) remapping induces changes in the mapping between latent and neural spaces (**a**, red dashed line), which can be visualized as rotating the latent space inside of a larger embedding space (**b**), or rotating the angle axes (**c**), such that the same latent trajectory passes through a different place cell sequence. **d-f**: mixed-selective (MS) remapping assumes changes to the cognitive environmental variables (**d**) and results in a different latent trajectory with a common positional readout (**e**), which passes through different place fields in angle space (**f**). **g-i**: nullspace (NS) remapping features changes only to the nonlinear part of the pseudo-linear encoder but keeps decoding the same (**g**), resulting in different trajectories with the same underlying latent sequence (**h**), and thus can be seen as changing the place fields that support this trajectory (**i**).

position is coded by a distinct sequence of neural activity in each environment. We will make this more precise by defining remapping across any two environments or contexts A and B as the case in which $\mathbf{r}^A(\mathbf{p}) \neq \mathbf{r}^B(\mathbf{p})$ for one or more spatial positions $\mathbf{p}$. This broad definition not only includes complete and partial remapping, but also any other more subtle variations in neural activity, and will be useful in attempting to capture the most general picture of hippocampal activity variability. However, as our focus here is on comparing firing-rate maps across environments, we will not consider any trial-to-trial or time-dependent variability in the mappings between environmental, latent, and neural spaces (but see Discussion), and instead focus on how environment-specific changes to the deterministic mappings of the model may lead to variability that resembles remapping.

Towards this aim, it is useful to restate the pseudo-linear encoding model from Eq. 2 with an environment-specific index attributed to each variable that could be subject to change across environments. This results in an environment-specific pseudo-linear encoder (see Methods) for a given environment A as

$$\mathbf{r}^A(\mathbf{p}) = \mathbf{E}^A\mathbf{z}^A(\mathbf{p}) + \boldsymbol{\nu}^A(\mathbf{z}^A(\mathbf{p})). \tag{3}$$

Now, with this equation, we can see that there are three possible ways for neural activity to remap as a function of environment: (i) changes to the linear encoder matrix, $\mathbf{E}^A$, (ii) changes to the latent variables themselves, $\mathbf{z}^A(\mathbf{p})$, and (iii) changes to the nonlinear nullspace function $\boldsymbol{\nu}^A(\cdot)$. We will consider each of these in turn.

The first case to consider is an environment-specific setting of the encoder matrix $\mathbf{E}^A$, which will change the mapping from latent space to neural space for each environment (Fig. 2a). We note that this affects both the encoder and decoder, and we thus refer to it as *encoder-decoder remapping*. We can understand this case as causing shifts or rotations of the axes representing the latent variables, which can be visualized in neural state space (Fig. 2b), or can be understood intuitively as a shift in the axes in angle space (Fig. 2c), which will cause a different sequence of place cell activations (cf. Fig. 1b, right). This case is in fact equivalent to the classic view of hippocampal remapping as a set of distinct maps (Samsonovich and McNaughton, 1997), each requiring a unique decoder for spatial position.

The second case fixes the latent axes, and instead considers changes to the latent trajectory through the environmental variables (Fig. 2d), which we restrict to the non-spatial, cognitive variables, $\mathbf{c}^A$, and their corresponding latent representation $\mathbf{z}_c^A$. While changes to spatial position can seemingly lead to remapping if each environment uses a non-overlapping portion of space (e.g., $p \in [0,1]$ for env. A and $p \in [1,2]$ for env. B), this

does not fall under our definition of remapping. Under changes to $\mathbf{z}_c^A$, we can see that the latent trajectory changes along the cognitive axis, despite the positional axis remaining unchanged (Fig. 2e). This can also be visualized in angle space, where we see that the neurons' mixed selectivity dictates that they will only be active when the latent trajectory crosses the right conjunction of spatial and non-spatial information (Fig. 1f; Rigotti et al. (2013); Fusi et al. (2016)). We thus term this *mixed-selective remapping*. This type of remapping relates to more recent latent state space models of the hippocampus (Whittington et al., 2022), and suggests that environmental differences in external (e.g., sensory) and internal (e.g., behavioral state) variables can lead to remapping even with a fixed decoder for spatial position.

Finally, in the third case, we consider the possibility for remapping to be explained solely through changes to the final, non-linear term $\boldsymbol{\nu}^A(\cdot)$, which only affects encoding (Fig. 2g). We refer to this case as *nullspace remapping*, as activity changes will be fully contained within the nullspace of the latent readouts[1] (Fig. 2h). Therefore, latent trajectories in angle space will appear identical across environments, and instead the neural firing fields supporting these trajectories will be modulated, indicating changes in tuning or excitability not captured by the linearly-decodable latent space (see Fig. 2i). While the idea of nullspace activity has been discussed in other contexts (e.g., Churchland and Shenoy (2024)), it has not been, to the best of our knowledge, related to remapping before (see Discussion).

We thus see that our constrained model has enabled us to distinguish three distinct coding mechanisms underlying remapping. Importantly, this result only depends on one of the two model assumptions — linear decoding from neural activity to latent space. In this sense, the three types of activity changes that we describe here will hold for any network with a linearly-decodable latent space, and thus can serve as a useful framework to model neural variability in general (see Discussion). In contrast, the assumption of angular coding is rather a hippocampus-specific assumption which is sufficient to generate localized place-like firing fields, as well as an overall nonlinear relationship between the environmental variables and neural activity, despite linear decoding. Our focus for the remainder of the paper will be to demonstrate examples from each of these three types of remapping, to compare them to common models in the literature, and to discuss how they relate to experiments.

Encoder-decoder remapping

We first considered encoder-decoder (ED) remapping, characterized by environment-dependent changes in the mapping between latent and neural spaces (cf. Fig. 2a-c). We considered two choices for how to model these modulations of the $\mathbf{z} \leftrightarrow \mathbf{r}$ mapping, which we will refer to as the multi-chart model and the grid realignment model, due to their respective correspondence to established models of place field remapping in the literature (Samsonovich and McNaughton, 1997; Fyhn et al., 2007). For simplicity, we temporarily omitted the non-spatial environmental variables and their corresponding latent variables, and thereby aimed to explain remapping solely through changes in how $\mathbf{z}_p$ is encoded into the neural activity $\mathbf{r}$ through the matrix $\mathbf{E}$ (Eq. 3).

In the multi-chart model, we modeled changes in $\mathbf{E}$ as a random rotational mapping of the latent variables in a higher-dimensional subspace that we refer to as the *embedding* space (see Methods). For the case of a linear track with a 2-d latent space, we can visualize this case as a random projection into a 3-d embedding space in which each trajectory takes on a random orientation (Fig. 3a). All trajectories will thus be confined to a sphere, with firing fields determined by the alignment of each neuron's preferred tuning vector with a particular environment's rotation of the latent space (Fig. 3a, bottom). We simulated this toy model in a network of $N = 24$ neurons, and visualized the circular latent input in embedding space (Fig. 3b). We then visualized firing fields (Fig. 3c; cf. Fig. 1g; see Methods) and plotted firing rates as a function of position (Fig. 3d), noting two observations. One, place cells had a neighborhood structure in embedding space, meaning that cells that were active together in one environment were likely appear again in another environment (Fig. 3d). Two, despite this structure, this model allows for arbitrary affine transformations of spatial coordinates — i.e., place fields could remap to arbitrary positions across environments, due to the fact that two latent trajectories could pass through the same point on the sphere, but at different corresponding positions (Fig. 3c,d).

To quantify these effects, we scaled up the multi-chart model to the case of a network of $N = 2048$ neurons coding for a 2-d square environment (4-d latent space embedded into a 128-d embedding space), and characterized the place fields using standard statistical measures (rate maps for two environments shown in Fig. 3e; see Methods). We found that the overlap in population firing-rates between environments was significantly larger than a shuffle control (Fig. 3f, top), confirming the neighborhood structure of the low-dimensional embedding space. In contrast, we then found that the average spatial correlation of each neuron's rate map across environments was no different from the shuffle control (Fig. 3f, bottom), again confirming the fact that position can map arbitrarily in each environment. Lastly, we suspected that the overlap could be made fully random (as in the original model, Samsonovich and McNaughton (1997)) by increasing the embedding dimensionality, which we verified in a simulation of a "full-dimensional" embedding space (Fig. S2).

[1] We note that the previous two types of remapping will, in general, also be accompanied by changes in $\boldsymbol{\nu}^A$, due to constraints on $\mathbf{r}^A(p)$ such as non-negativity and ℓ_1 cost (see Fig. S1). The difference here is that changes are restricted only to this term.

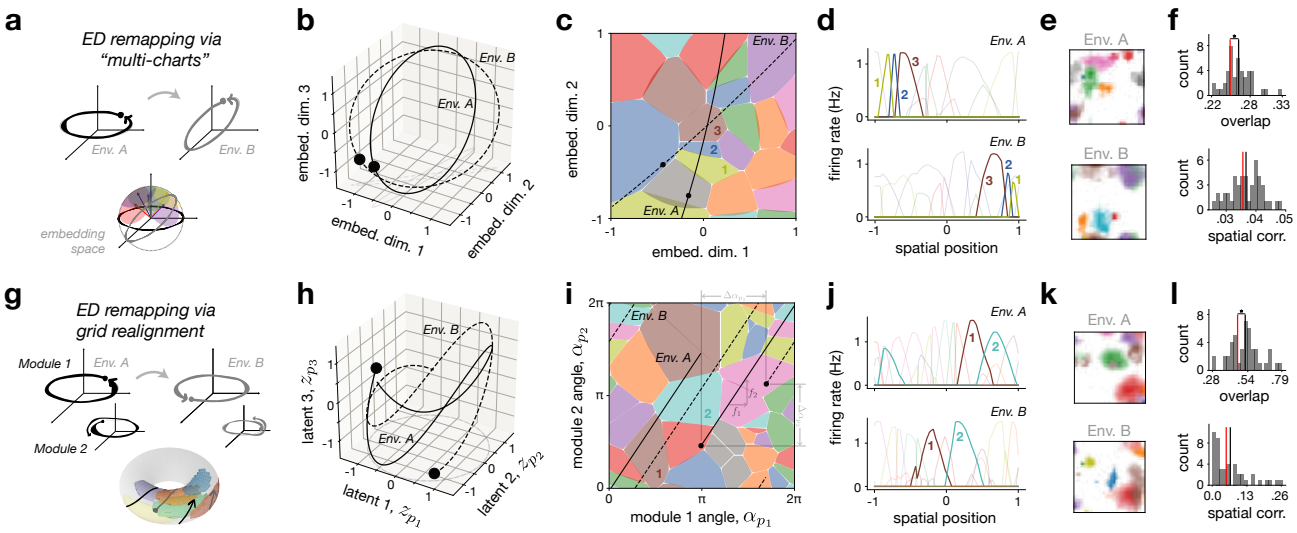

Figure 3: **Encoder-decoder (ED) remapping.** **a**: ED remapping via the classic "multi-chart" model. Considering a linear track with 2-d latent space, the latent space can be seen as being randomly rotated into a larger 3-d embedding space, which forms a sphere. Place cell remapping reflects the different alignment of neural tuning vectors with each environment's slice through the sphere. **b,c,d**: A network of $N = 24$ neurons is simulated with the setup from panel (**a**) for two environments A and B; latent inputs follow rotated circular trajectories (**b**), and place fields are visualized in a 2-d projection of the 3-d embedding space (**c**) or as a function of position (**d**); three place cells are highlighted ($1, 2, 3$ in (**c**, **d**)). **e,f**: Scaled up multi-chart model with 2-d spatial position (4-d latent space) in a network of $N = 2048$ neurons and a 128-dimensional embedding space, with place field rate maps from two example environments (**e**) and overlap and spatial correlation distributions (mean in black) compared with a shuffle control (red) (**f**); star indicates statistical significance (see Methods). **g**: ED remapping via grid realignment. Considering a linear track with two grid modules (4-d latent space), the latent space can be seen as being realigned or rotated within itself in each environment, resulting in shifted trajectories on the very same toroidal latent manifold. **h,i,j**: A network of $N = 24$ neurons is simulated with the setup from panel (**g**) and analogous to panels (**b,c,d**); latent inputs are shown in 3 of the 4 latent dimensions (**h**), and place fields are visualized in angle space (**i**) and as a function of position (**j**) with two place cells highlighted (1 and 2); realignment is due to phase changes in the starting position of each angle, and the common slope is due to the ratio of frequencies, f_2/f_1. **k,l**: Scaled up grid realignment model with 2-d spatial position in a network of $N = 96$ neurons and 3 grid modules (12-d latent space), with place field rate maps from two example environments (**k**) and overlap and spatial corr. distributions (mean in black) compared with shuffle (red, **l**).

At face value, an effective multi-chart model requires substantial environment-dependent changes in input or connectivity, with questionable plausibility (see Discussion). We thus considered an alternative approach based on grid module realignment, a well-characterized phenomenon in which different phase relationships between grid codes in the entorhinal cortex may lead to different place cell activation patterns (Solstad et al., 2006; Fyhn et al., 2007; Monaco and Abbott, 2011). We can envision this as a four-dimensional latent space for a 1-d linear track position, consisting of two grid cell module inputs that form a torus (Fig. 3g). For each individual module, rather than rotating the latent space in a higher-dimensional space, the circular trajectory is simply rotated on its own axis — i.e. the phase is realigned — resulting in different trajectories and place field activations on the toroidal latent manifold (Fig. 3g, see Methods).

We simulated a toy version of this case (following Fig. 3g) in a network of $N = 24$ neurons and 2 modules, and visualized the trajectories in latent space (Fig. 3h), and in angle space with the place fields visualized (Fig. 3i,j). When looking at the trajectories in three of the four latent dimensions (Fig. 3h; linear embedding of the 4-d torus; see Methods), we observed that the circular positional encoding of the start of the linear track (black dots) was shifted both in the first module (given by the first two dimensions) and in the second module (visualized in the difference of height in the third dimension), resulting in distinct trajectories. In angle space, we observed that grid realignment results in random starting locations of the trajectories (horizontally by the phase shift of the first module and vertically by the second), despite a fixed slope that reflects the fixed ratio of periods of the two modules (Fig. 3i). In this toy model, as well as in a scaled up case (3 modules in a 12-d latent space; Fig. 3k,l), we found qualitatively similar results to the multi-chart model — this included a neighborhood structure marked by a population firing-rate overlap significantly greater than shuffle, but arbitrary positional remapping marked by a random spatial correlation (Fig. 3l). Similarly to increasing the embedding dimensionality of the multi-chart model, this could be made more random with the addition of

more grid modules. Furthermore, this grid-like input also allows for place cells with multiple fields (Fig. 3j, 2-cyan). In sum, the multi-chart and grid realignment models can thus be seen as two qualitatively similar implementations from the same class of remapping, whose properties crucially depend on the dimensionality of the neural coding space (see Discussion).

Mixed-selective remapping

We then moved on to mixed-selective (MS) remapping, which uses a common latent space across all environments, but also includes cognitive (i.e., non-spatial) environmental variables that can change within and across environments (cf. Fig. 2d-f). Given the generality of this model, we focused on two particular cases of interest, which we will call space-feature coding and implicit-space coding, respectively, to help highlight the flexibility and constraints of this class of remapping and its relation to recent models from the literature.

The first case that we considered was space-feature coding, which was characterized by a fixed latent trajectory for spatial position across all environments, $\mathbf{z}_p(p)$, in addition to environment- and position- dependent cognitive variables $\mathbf{z}_c^A(p)$ (Fig. 4a). Neurons were tuned to be conjunctive to particular combinations of spatial position and cognitive variables, similar to other models in the field (see Methods; Whittington et al. (2022)). We first modeled a toy version of this setup, which featured a four-dimensional latent space composed of a pair of latents coding for linear track position, along with another pair of latents coding for a single cognitive variable (Fig. 4a), and we visualized trajectories and place fields in latent space and angle space (Fig. 4b-d). Across the three environments shown, trajectories evolved through the same positional trajectory (circular projection onto horizontal plane, Fig. 4b), with variability due to the additional cognitive information (vertical axis).

To understand the effects of variability in the cognitive variables, we then considered two trajectories in the same neighborhood of latent space (Fig. 4b,c; Envs. A and B, solid and dashed), as well as an additional distinct trajectory (Fig. 4b,c, Env. C, dotted). We found that the more similar environments (Envs. A and B) exhibited partial remapping, with coherent rate remapping in some positions (neurons 1, 2, and 3), and more drastic changes in other positions (4 and 5; Fig. 4c,d). Comparing more distinct environments (Env. C) led to global remapping with non-overlapping groups of place cells. We scaled up this model by considering 2-d spatial trajectories along with random 60-dimensional cognitive trajectories generated by a Gaussian Process (see Methods; Fig. 4e,f). As expected, overlap was more random than ED remapping, since the appearance of two neurons did not necessarily predict the appearance of a third, due to the flexibility of the contextual variables (Fig. 4f, top). However, in contrast to ED remapping, we found that spatial correlation was significantly different from shuffle, a consequence of neurons being conjunctive for a particular fixed place (Fig. 4f, bottom).

Given the natural importance of the spatial positional variable in determining the extent of remapping for space-feature coding, we next wondered what would happen if the non-spatial cognitive variables were made relatively more important as compared to position. Inspired by some related models in the literature (Benna and Fusi, 2021; Raju et al., 2024), we then considered an extreme case of this that we call the implicit-space coding model, which did not contain any explicit spatial position variables at all (Fig. 4g). Instead, in this case latent trajectories were composed solely of cognitive environmental variables and their corresponding latent variables, and so each position p could be associated with arbitrary locations in latent space (Fig. 4h). Similar to space-feature coding, however, this model again depended on the richness of the cognitive variable trajectories, which we explored in a toy model with two cognitive variables coded by a four-dimensional latent space.

As expected, we found that when the latent trajectories were distinct, they could result in complete random remapping (Fig. 4h-j, Envs. A and C). However, if two trajectories were too similar, e.g., if they shared the same cognitive state in particular areas, they could display similar place field activities (Fig. 4h-j, Envs. A and B). We then scaled up this implicit-space coding model and again used random Gaussian Process trajectories for each environment (Fig. 4k,l) — importantly, the parameter controlling variability in these trajectories determines how structured or random the remapping will be. For sufficient variability (see Methods), we found that this setup resulted in complete random remapping with respect to both overlap and spatial correlation (Fig. 4l). In summary, we see that space-feature and implicit-space coding represent two types of mixed-selective remapping, which differ in the relative importance of an explicit code for spatial position relative to other non-spatial variables. Furthermore, the randomness of remapping in both models crucially relies upon the richness and variability of the non-spatial information encoded into place cell activity, acting as a strong assumption for this class of remapping to explain experimental activity changes.

Nullspace remapping

Lastly, we considered nullspace remapping, in which all changes to neural activity occur in the nullspace of the latent space (cf. Fig. 2g-i). We can visualize this case in two ways — nullspace changes do not actually change the trajectories in latent space, and thus can be seen as modulating the neural firing fields that are contained within this space (cf. Fig. 2i); equivalently, we can also identify particular nullspace directions outside of the latent space, and view trajectories as changing in this space (Fig. 5a). In fact, the other two types of

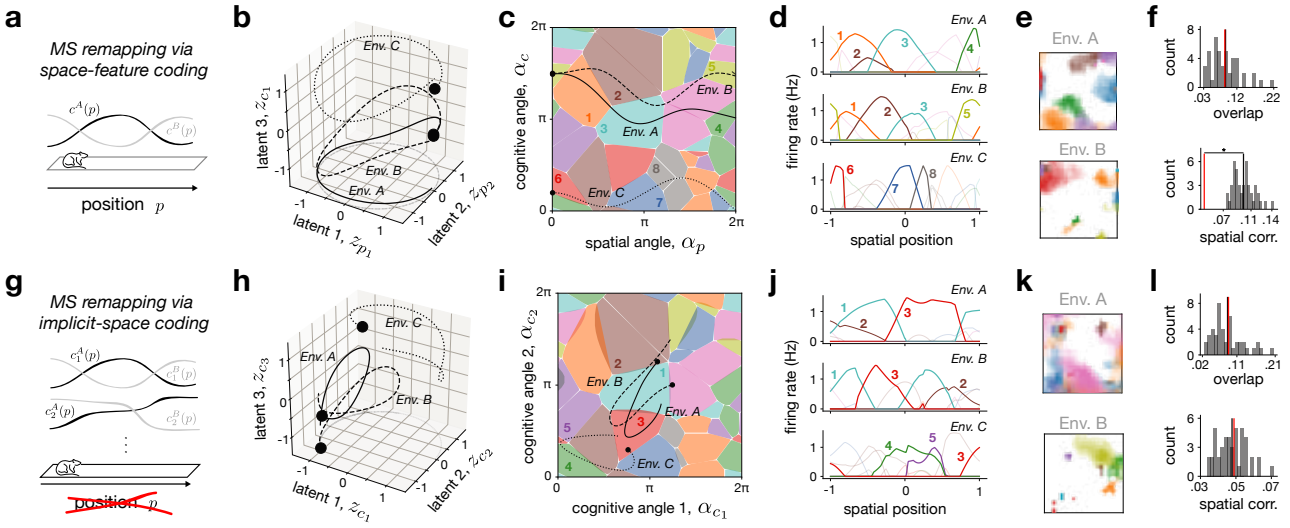

Figure 4: **Mixed-selective (MS) remapping.** **a**: MS remapping via space-feature coding; each environment features a shared position variable p, plus an environment-specific and position-dependent cognitive variable ($c^A(p)$ and $c^B(p)$). **b,c,d**: A network of N = 32 neurons is simulated using the space-feature coding setup from panel (**a**) for three environments; latent input trajectories visualized in 3 of the 4 latent dimensions shows circular positional trajectories (**b**) with variability in the third, cognitive direction ((z_1, z_2) plane); trajectories and place fields are visualized in angle space (**c**) and as a function of position (**d**), with several place fields highlighted (1-8). **e,f**: Scaled up example of space-feature coding model with N = 1024 neurons and 32 environmental variables (including 2-d position), with example place field maps from two environments (**e**) and overlap and spatial correlation distributions (mean in black) compared with a shuffle control (red) (**f**). Star indicates statistical significance (see Methods). **g**: MS remapping via implicit-space coding; similar to the position-dependent model, but with the spatial variables omitted, so that hippocampal activity merely reflects the conjunction of a set of cognitive variables (two shown for each environment, $c_1^A(p)$ and $c_2^A(p)$ for env. A). **h,i,j**: A network of N = 32 neurons is simulated using the implicit-space coding setup from panel (**g**) for three environments; latent input trajectories visualized in 3 of the 4 cognitive latent dimensions shows unconstrained trajectories (**b**); trajectories and place fields are visualized in angle space (**i**) and as a function of position (**j**), with several place fields highlighted (1-5). **k,l**: Scaled up example of implicit-space coding with N = 1024 neurons and 32 environmental variables (*excluding* 2-d position), with example place field maps from two environments (**k**) and overlap and spatial corr. distributions (mean in black) compared with shuffle (red) (**l**).

remapping implicitly already featured nullspace remapping, in that changes to the linear mapping in the first term of Eq. 3 will often necessarily be accompanied by changes to the nullspace term $\boldsymbol{\nu}$, in order to retain non-negativity in firing rates (Fig. S1). However, although the redundancy of the code dictates that there should be different combinations of activity patterns that represent the same latent space, finding a mechanism that can constrain changes solely to the nullspace is not a trivial task (see Discussion).

Though it is not often framed as such, one prominent source of nullspace remapping can actually be found in a competitive code, such as those induced through sparse coding and lateral inhibition, in which neurons compete to become active (Rozell et al., 2008). To see this, let us consider two place cells that share the same spatial localization (and thus latent tuning preference), and which are connected through lateral inhibition (Fig. 5b,c). With our linear decoder assumption, this means that we can visualize the latent readout in two-dimensional neural state space on the diagonal between the two axes (Fig. 5b, blue), along with an orthogonal nullspace axis (Fig. 5b, pink). Then, assuming inhibitory competition between the two neurons, we can see that there are three qualitatively different activity regimes which each result in the same latent readout — either one of the two neurons outcompetes the other and fires alone, or the two jointly fire together (Fig. 5c). In fact, the recurrent network model that we have employed here has the exact lateral-inhibitory structure needed for this effect (Boerlin et al. (2013); Denève and Machens (2016); Podlaski and Machens (2024a); see Discussion), and redefines trial-to-trial variability as a minimal type of nullspace "remapping".

Interestingly, it is also known that such nullspace changes can be induced through the inhibition of particular neurons — in our example (Fig. 5b,c), this would correspond to one neuron being inhibited, and thus the other one compensating for this change (Barrett et al., 2016; Calaim et al., 2022). Therefore, to study a more interesting and dramatic example, we will consider a recent experimental result in which a default place cell code was tagged and silenced in an awake behaving animal (Trouche et al., 2016). Interestingly, when this occurred, a new, alternative place code emerged in the network, which the authors attributed to inhibitory

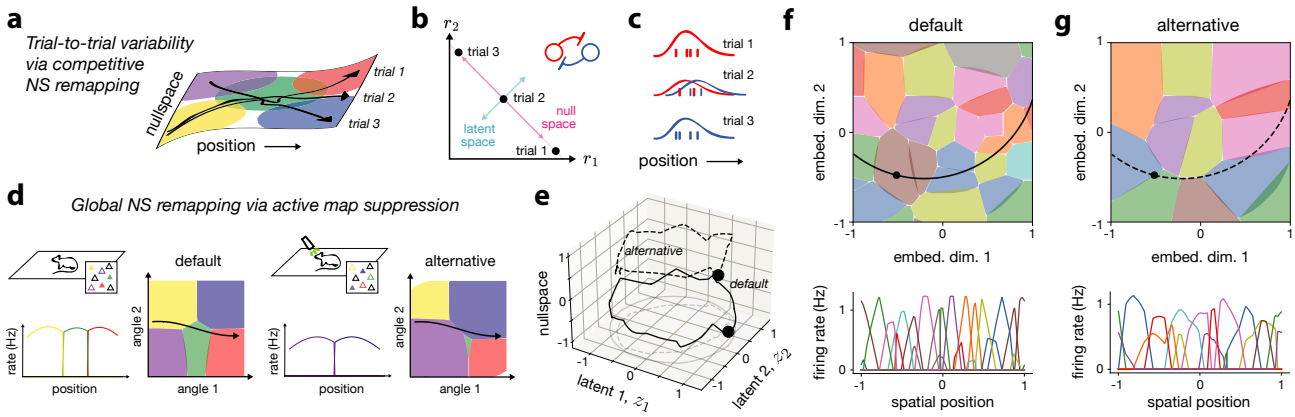

Figure 5: **Nullspace (NS) remapping.** **a**: Schematic of how variability in a nullspace direction may cause place fields to be modulated, resulting in trial-to-trial variability (cf. Fig. 2i). **b,c**: In two-dimensional (r_1, r_2) neural space, a single latent variable is coded (blue), with an orthogonal nullspace direction (pink); a competitive architecture with lateral inhibition results in trials with distinct activity for neuron 1 (red) and neuron 2 (blue). **d**: Illustration of experimental setup from Trouche et al. (2016) and underlying competitive mechanism, in which a default stable place map was inhibited, resulting in an alternative map becoming active; illustration of competition in angle space, in which each field competes for coding space, and inhibition of the active neurons results in the blue and purple neurons taking over. **e,f,g**: Simulated example of the setup from panel (**d**), using a single environment of the multi-chart ED model from Fig. 3b-d (N = 48 neurons). Circular positional trajectories are plotted with the maximum-variance nullspace direction (**e**; see Methods); the default and alternative place field maps are shown in (**f**) and (**g**) in angle space (top) and as a function of position (bottom). Note that there are no shared place neurons between the two environments due to the complete inhibition of the default map.

competition but without a clear mechanism (Fig. 5d). Here, we can propose a more explicit mechanistic explanation based on nullspace remapping, which is that an alternative set of neurons can support the very same positional code. We can visualize the effects of neural inhibition in the firing field maps in angle space as the competition between neighboring maps (Fig. 5d). Active neurons in an environment, given by where the latent trajectory crosses (Fig. 5d, left), shrink and disappear due to the experimental inhibition, leading the neighboring fields to grow and fill in the space (Fig. 5d, right). Intriguingly, this mechanism comes with an additional experimental prediction, which is that partial inhibition of an active map should also cause the non-inhibited neurons to increase their place field sizes to account for "gaps" in the representation.

To confirm these effects in simulations, we utilized the multi-chart example of ED remapping from Fig. 3a-d. Following a simulation of one environment of the network, we then tagged the active neurons, and ran another simulation with these neurons silenced (Fig. 5e-g). We found that a qualitatively similar circular positional trajectory was found in both maps, with a shift being primarily confined to the nullspace (Fig. 5e). Furthermore, as expected, the competitive nature of the recurrent network resulted in a deletion of the active cells firing field and an expansion of non-active ones, which translated into a new sequence of place field activations (Fig. 5f,g). Interestingly, this type of nullspace remapping supports a view of the hippocampus as having competitive dynamics (Wills et al., 2005) — however, instead of highlighting competition between distinct maps, this view suggests that individual neurons also compete to play a role in the same active map (see Discussion).

Case study: reward-modulated place fields

We have now presented examples of three types of remapping in networks that represent a linearly-decodable latent space. To conclude, we considered an additional example that highlights the benefits of visualizing remapping as latent trajectories crossing neurons' firing fields. Specifically, we considered a task in which an animal navigates through a linear track environment with a goal or reward that can potentially change location (Fig. 6a). Such a setup of goal-directed firing in the hippocampus has been the focus of several studies (Dupret et al., 2010; Gauthier and Tank, 2018), but some questions arise as to the nature of the representations that it yields (O'Keefe and Krupic, 2021). To probe these representations, we can compare the case in which neurons are conjunctive to specific combinations of position and cognitive information, to the case of random tuning, resulting in some neurons with selectivity only to one of the two variable types (i.e., pure selectivity; Fig. 6b). e To model this, we considered the mixed-selective remapping setup with space-feature coding (cf. Fig. 4a-e), with a pair of angles coding for linear track position and reward level, respectively. We first considered conjunctive neurons, i.e., with equal tuning to a particular position and reward value (Fig. 6b, left; see Methods). We then simulated two distinct reward locations and observed the resultant place field rate maps that resulted from this setup, which yielded several interesting results (Fig. 6c,d). One, we observed that place fields exhibited partial

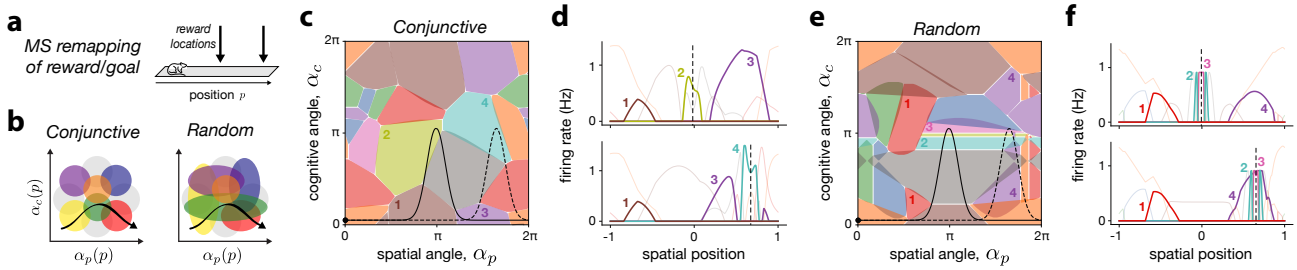

Figure 6: **Two types of reward coding with mixed-selective remapping.** **a**: Experimental setup, in which an animal navigates through a linear track environment, with two distinct reward locations. **b**: Two distributions of neural tuning, in which neurons are tuned with equal weighting towards position and reward intensity (conjunctive, left), or in which neurons are tuned randomly, on a spectrum from purely selective to mixed-selective (random, right) (see Methods). **c,d**: MS remapping with conjunctive selectivity in a network of N = 16 neurons and 4-d latent space; two latent trajectories follow the reward location, overlapping with different firing fields in angle space (**c**) and resulting in a different sequence of place cell activity (**d**, four neurons highlighted 1-4). **e,f**: Same as panels (**c,d**) but with random tuning, resulting in some neurons with pure selectivity for either position or reward, and others conjunctive.

remapping — this included a more complete change in the neurons coding for the reward location (Fig. 6c,d, 2-yellow, 4-cyan), with fields farther away from the reward remaining unchanged (1-brown) and other neurons in between showing partial modulation (3-purple). Two, we additionally observed an increase in the number of neurons coding for the goal location (as seen experimentally; Hollup et al. (2001)), which we could attribute to the longer trajectory in angle (or latent) space around the reward locations.

Given the evidence for purely reward-selective neurons in the literature (Gauthier and Tank, 2018), we then considered the case of random tuning (Fig. 6b, right) in which some neurons appear to be more pure selective (yellow - pure selective to position, green - pure selective to reward). We then simulated the same setup as before, and noted some interesting differences in activity profiles (Fig. 6e,f). First we noticed that, due to the presence of some non-pure neurons, we could capture the same qualitative effects of the previous model. However, due to the presence of pure selective cells, we also found some reward cells that shifted their responses between the two locations (Fig. 6e, 2-cyan, 3-pink), and some spatial cells that fired independently of the reward (Fig. 6e, 4-purple). While other neurons could appear to be pure-selective cells to, e.g., position, our experimental setup could not be sufficiently rich to distinguish it. We thus conclude that the flexibility of this model is able to capture a host of experimental observations, and provides mechanistic insights that may explain when they occur and when they do not occur (see Discussion).

Discussion

In this work, we have presented a unifying view of hippocampal place field remapping from the perspectives of neural coding and population geometry. Our work is one among several recent studies that employ a population-level perspective on spatial representations (Meshulam et al., 2017; Low et al., 2018; Stefanini et al., 2020; Nieh et al., 2021; Gardner et al., 2022; Levy et al., 2023), though our focus on remapping is relatively unique (but see Low et al. (2023); Fenton (2024)). By assuming a linearly-decodable latent state space, we demonstrated three mechanisms — encoder-decoder (ED), mixed-selective (MS), and nullspace (NS) remapping — that each may underlie neural activity changes, including complete and partial, as well as random and structured remapping. We observed that rate overlap and spatial correlation can often serve as useful measures to distinguish ED and MS remapping (Figs. 3,4), though in extreme cases they may become undistinguishable through these measures alone. However, we stress that our aim was not to determine which of our three remapping mechanisms was "correct" — rather, we suspect that all three may be accurate depictions of activity changes in the hippocampus under different experimental conditions. Our modeling perspective can thus serve as a testbed to understand how various experimental manipulations may affect the hippocampal population code.

Our framework allowed us to re-contextualize various remapping findings and theories. Our two proposed implementations of ED remapping enabled us to link multi-chart (Samsonovich and McNaughton, 1997) and grid realignment (Fyhn et al., 2007; Monaco and Abbott, 2011) models in analogous low-dimensional population spaces. Though a multi-chart model based on rotation of the latent space may have some plausibility (e.g., Naumann et al. (2022)), it can also be seen as a useful abstraction of grid realignment that captures a grid code's ability to efficiently tile a high-dimensional space (Sharma et al., 2022; Chandra et al., 2025). A similar intuition was found for MS remapping, in which the richness of sensory and cognitive variables may play a similar role, and future work should more explicitly study how various models fit into this perspective (Whittington

et al., 2020; Benna and Fusi, 2021; Raju et al., 2024). Additionally, as explored for the case of reward coding (cf. Fig. 6), the interaction of a latent trajectory with the nature of the place cells' mixed-selectivity (Rigotti et al., 2013) can explain various experimental findings, including splitter cells (Frank et al., 2000; Wood et al., 2000), contextual modulation (Chiossi et al., 2024), and behavioral modulation (Liberti III et al., 2022).

Nullspace remapping enabled us to not only provide an alternative explanation and unique predictions for suppression-induced remapping (Trouche et al., 2016), but also presents intriguing connections to models of variable activity in working memory (Druckmann and Chklovskii, 2012) and motor control (Churchland and Shenoy, 2024). In a sense, MS and NS remapping are not so different, in that changes occur in the nullspace of spatial position, with the difference being that in MS remapping these changes occur in an experimentally controllable or identifiable direction, and in NS remapping they do not (cf. Fig. 4b & Fig. 5e). Thus, ED remapping is the only case that can truly be considered a change in the cognitive map for space. Given the fixed positional readout for MS and NS remapping, these cases may also help to explain experiments in which remapping occurs, but behavioral performance is unchanged (Jeffery et al., 2003).

Our framework featured two important assumptions. Our assumption of angular coding was meant to serve as a convenient example mapping between the latent space and the spatial and cognitive variables, and can be seen as an explicit modeling choice that induces constant population activity (as observed in the hippocampus; Buzsáki et al. (2002)), localized firing fields (Herz et al., 2017), and the possibility for each cell to have multiple (but not necessarily periodic) fields (Rich et al., 2014). Our choice of a linear decoder from neural to latent space was the only necessary assumption to derive the three remapping mechanisms, and is backed up by substantial motivation (Rigotti et al., 2013; Fusi et al., 2016; Bernardi et al., 2020). Furthermore, we stress that this did not mean that space could be linearly decoded from neural activity (due to the nonlinear angle coding), and also that it did not imply that neural activity linearly encodes the latent variables (as demonstrated in Fig. S1). We summarize these two assumptions in the following way: angular coding was a sufficient but not strictly necessary choice for inducing localized place-fields; linear decoding was a necessary choice for deriving the three types of remapping, and does not depend on the angular coding or place-like firing.

An important limitation of our work was the choice to model the hippocampus as an autoencoder network, without modeling internal dynamics, coupling with a grid cell network, nor any unsupervised or task-based learning (Rennó-Costa and Tort, 2017; Weber and Sprekeler, 2018; Whittington et al., 2020; Agmon and Burak, 2020; Tessereau et al., 2021; Benna and Fusi, 2021; Morris and Derdikman, 2023; Raju et al., 2024; Pettersen et al., 2024; Wang et al., 2024). Our choice to do so was motivated by the desire to understand hippocampal representations in the most general sense (and without any learning or plasticity), as modeling a particular learning process or task would have narrowed the generality of the results. That said, our results can neither speak to the utility of particular representations, nor their plausibility, and it will be interesting to consider dynamics, both as a way of forming internally-generated trajectories in latent space (Whittington et al., 2022), as well as a potential additional mechanism for remapping in the case of toggling or switching maps (Gothard et al., 1996; Sheintuch et al., 2020). Interestingly, the architecture we employed here has already been applied to model path integration (Boerlin et al., 2013), attractor memories (Podlaski and Machens, 2024b), and other dynamics (Alemi et al., 2018; Nardin et al., 2021), which could be explored in concert with hippocampal representations in the future. Alternatively, our model can also be used as a framework with which to understand and interpret the representations produced by these other trained hippocampal models.

We chose to limit our study to deterministic mappings from position to activity, but variability and stochasticity have been of interest in the field (Fenton and Muller, 1998; Olypher et al., 2002; Low et al., 2018; Nieh et al., 2021; Fenton, 2024). Our model can be extended to account for this, either through stochasticity of the environment variables (Renart and Machens, 2014; Nieh et al., 2021), or internally-generated trial-to-trial variability such as through a spiking implementation of the model (Podlaski and Machens (2024a); cf. Fig. 5a-c). We note that such spiking variability would also open up our framework to the analysis of shorter timescale co-firing patterns which are known to be relevant for hippocampal coding (El-Gaby et al., 2021; Levy et al., 2023; Nardin et al., 2023; Fenton, 2024). One additional type of variability is the drift or instability of place cell activity over time (Ziv et al., 2013), which has been linked to remapping elsewhere (Whittington et al., 2022). From the perspective of our theoretical framework, this drift can be attributed to each of the three remapping types — learning or plasticity-induced modifications may change the latent mapping (ED remapping) (Sanders et al., 2020; Sun et al., 2025), changes in behavioral or cognitive variables may seemingly result in drift (MS remapping) (Sadeh and Clopath, 2022), or non-coding excitability changes, neurogenesis, or cell death may cause nullspace (NS) remapping (Delamare et al., 2024).

Lastly, in line with the previous point, this work is more general than remapping, and describes how mixed selectivity and linear decoding lead to an interpretable understanding of neural representations and how they vary. Many of the theoretical results were obtained without any explicit assumptions of place or spatial tuning (see Methods). Therefore, this framework could be useful in characterizing geometry and variability in other areas (e.g., Genkin et al. (2023); El-Gaby et al. (2024)). Thus, our work fits in with the recent trend in the literature of neural manifolds and population geometry (Gallego et al., 2017; Chung and Abbott, 2021; Jazayeri and Ostojic, 2021; Langdon et al., 2023), and may be a useful perspective on neural variability in general.

References

Agmon, H. and Burak, Y. (2020). A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *Elife*, 9:e56894.

Agmon, H. and Burak, Y. (2023). Simultaneous embedding of multiple attractor manifolds in a recurrent neural network using constrained gradient optimization. *arXiv preprint arXiv:2310.18708*.

Alemi, A., Machens, C., Deneve, S., and Slotine, J.-J. (2018). Learning nonlinear dynamics in efficient, balanced spiking networks using local plasticity rules. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Alme, C. B., Miao, C., Jezek, K., Treves, A., Moser, E. I., and Moser, M.-B. (2014). Place cells in the hippocampus: eleven maps for eleven rooms. *Proceedings of the National Academy of Sciences*, 111(52):18428–18435.

Aronov, D., Nevers, R., and Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722.

Barrett, D. G., Denève, S., and Machens, C. K. (2013). Firing rate predictions in optimal balanced networks. *Advances in Neural Information Processing Systems*, 26.

Barrett, D. G., Deneve, S., and Machens, C. K. (2016). Optimal compensation for neuron loss. *Elife*, 5:e12454.

Benna, M. K. and Fusi, S. (2021). Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence. *Proceedings of the National Academy of Sciences*, 118(51):e2018422118.

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C. D. (2020). The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell*, 183(4):954–967.

Boerlin, M., Machens, C. K., and Denève, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology*, 9(11):e1003258.

Bostock, E., Muller, R. U., and Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. *Hippocampus*, 1(2):193–205.

Buzsáki, G., Csicsvari, J., Dragoi, G., Harris, K., Henze, D., and Hirase, H. (2002). Homeostatic maintenance of neuronal excitability by burst discharges in vivo. *Cerebral cortex*, 12(9):893–899.

Calaim, N., Dehmelt, F. A., Gonçalves, P. J., and Machens, C. K. (2022). The geometry of robustness in spiking neural networks. *Elife*, 11:e73276.

Chandra, S., Sharma, S., Chaudhuri, R., and Fiete, I. (2025). Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, pages 1–13.

Chiossi, H. S., Nardin, M., Tkacik, G., and Csicsvari, J. L. (2024). Learning reshapes the hippocampal representation hierarchy. *bioRxiv*, pages 2024–08.

Chung, S. and Abbott, L. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. *Current opinion in neurobiology*, 70:137–144.

Churchland, M. M. and Shenoy, K. V. (2024). Preparatory activity and the expansive null-space. *Nature Reviews Neuroscience*, pages 1–24.

Colgin, L. L., Moser, E. I., and Moser, M.-B. (2008). Understanding memory through hippocampal remapping. *Trends in neurosciences*, 31(9):469–477.

Dayan, P. and Abbott, L. F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press.

Delamare, G., Zaki, Y., Cai, D. J., and Clopath, C. (2024). Drift of neural ensembles driven by slow fluctuations of intrinsic excitability. *Elife*, 12:RP88053.

Denève, S. and Machens, C. K. (2016). Efficient codes and balanced networks. *Nature neuroscience*, 19(3):375–382.

Diamond, S. and Boyd, S. (2016). Cvxpy: A python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5.

Druckmann, S. and Chklovskii, D. B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, 22(22):2095–2103.

Dupret, D., O'neill, J., Pleydell-Bouverie, B., and Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8):995–1002.

Eichenbaum, H. (2018). Barlow versus hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? *Neuroscience letters*, 680:88–93.

Eichenbaum, H. and Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83(4):764–770.

El-Gaby, M., Harris, A. L., Whittington, J. C., Dorrell, W., Bhomick, A., Walton, M. E., Akam, T., and Behrens, T. E. (2024). A cellular basis for mapping behavioural structure. *Nature*, pages 1–10.

El-Gaby, M., Reeve, H. M., Lopes-dos Santos, V., Campo-Urriza, N., Perestenko, P. V., Morley, A., Strickland, L. A., Lukács, I. P., Paulsen, O., and Dupret, D. (2021). An emergent neural coactivity code for dynamic memory. *Nature neuroscience*, 24(5):694–704.

Fenton, A. A. (2024). Remapping revisited: how the hippocampus represents different spaces. *Nature Reviews Neuroscience*, pages 1–21.

Fenton, A. A., Kao, H.-Y., Neymotin, S. A., Olypher, A., Vayntrub, Y., Lytton, W. W., and Ludvig, N. (2008). Unmasking the ca1 ensemble place code by exposures to small and large environments: more place cells and multiple, irregularly arranged, and expanded place fields in the larger space. *Journal of Neuroscience*, 28(44):11250–11262.

Fenton, A. A. and Muller, R. U. (1998). Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proceedings of the National Academy of Sciences*, 95(6):3182–3187.

Frank, L. M., Brown, E. N., and Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27(1):169–178.

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current opinion in neurobiology*, 37:66–74.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., and Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194.

Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, 94(5):978–984.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., and Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128.

Gauthier, J. L. and Tank, D. W. (2018). A dedicated population for reward coding in the hippocampus. *Neuron*, 99(1):179–193.

Gava, G. P., McHugh, S. B., Lefèvre, L., Lopes-dos Santos, V., Trouche, S., El-Gaby, M., Schultz, S. R., and Dupret, D. (2021). Integrating new memories into the hippocampal network activity space. *Nature neuroscience*, 24(3):326–330.

Genkin, M., Shenoy, K. V., Chandrasekaran, C., and Engel, T. A. (2023). The dynamics and geometry of choice in premotor cortex. *BioRxiv*.

Gothard, K. M., Skaggs, W. E., Moore, K. M., and McNaughton, B. L. (1996). Binding of hippocampal ca1 neural activity to multiple reference frames in a landmark-based navigation task. *The Journal of neuroscience*, 16(2):823.

Herz, A. V., Mathis, A., and Stemmler, M. (2017). Periodic population codes: from a single circular variable to higher dimensions, multiple nested scales, and conceptual spaces. *Current Opinion in Neurobiology*, 46:99–108.

Hollup, S. A., Molden, S., Donnett, J. G., Moser, M.-B., and Moser, E. I. (2001). Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5):1635–1644.

Jazayeri, M. and Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current opinion in neurobiology*, 70:113–120.

Jeffery, K. J. (2007). Integration of the sensory inputs to place cells: what, where, why, and how? *Hippocampus*, 17(9):775–785.

Jeffery, K. J., Gilbert, A., Burton, S., and Strudwick, A. (2003). Preserved performance in a hippocampal-dependent spatial task despite complete place cell remapping. *Hippocampus*, 13(2):175–189.

Keemink, S. W. and Machens, C. K. (2019). Decoding and encoding (de) mixed population responses. *Current Opinion in Neurobiology*, 58:112–121.

Langdon, C., Genkin, M., and Engel, T. A. (2023). A unifying perspective on neural manifolds and circuits for cognition. *Nature Reviews Neuroscience*, 24(6):363–377.

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., and Moser, M.-B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734):619–623.

Levenstein, D., Efremov, A., Eyono, R. H., Peyrache, A., and Richards, B. (2024). Sequential predictive learning is a unifying theory for hippocampal representation and replay. *bioRxiv*, pages 2024–04.

Levy, E. R. J., Carrillo-Segura, S., Park, E. H., Redman, W. T., Hurtado, J. R., Chung, S., and Fenton, A. A. (2023). A manifold neural population code for space in hippocampal coactivity dynamics independent of place fields. *Cell reports*, 42(10).

Liberti III, W. A., Schmid, T. A., Forli, A., Snyder, M., and Yartsev, M. M. (2022). A stable hippocampal code in freely flying bats. *Nature*, 604(7904):98–103.

Low, I. I., Giocomo, L. M., and Williams, A. H. (2023). Remapping in a recurrent neural network model of navigation and context inference. *Elife*, 11:e86943.2.

Low, R. J., Lewallen, S., Aronov, D., Nevers, R., and Tank, D. W. (2018). Probing variability in a cognitive map using manifold inference from neural dynamics. *BioRxiv*, page 418939.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749.

Mancoo, A., Keemink, S., and Machens, C. K. (2020). Understanding spiking networks through convex optimization. *Advances in Neural Information Processing Systems*, 33:8824–8835.

Mastrogiuseppe, F. and Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623.

Mathis, M. W., Rotondo, A. P., Chang, E. F., Tolias, A. S., and Mathis, A. (2024). Decoding the brain: From neural representations to mechanistic models. *Cell*, 187(21):5814–5832.

Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W., and Bialek, W. (2017). Collective behavior of place and non-place neurons in the hippocampal network. *Neuron*, 96(5):1178–1191.

Monaco, J. D. and Abbott, L. F. (2011). Modular realignment of entorhinal grid cell activity as a basis for hippocampal remapping. *Journal of Neuroscience*, 31(25):9414–9425.

Morris, G. and Derdikman, D. (2023). The chicken and egg problem of grid cells and place cells. *Trends in Cognitive Sciences*, 27(2):125–138.

Muller, R. U. and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–1968.

Muller, R. U., Kubie, J. L., and Ranck, J. B. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *Journal of Neuroscience*, 7(7):1935–1950.

Nardin, M., Csicsvari, J., Tkačik, G., and Savin, C. (2023). The structure of hippocampal ca1 interactions optimizes spatial coding across experience. *Journal of Neuroscience*, 43(48):8140–8156.

Nardin, M., Phillips, J. W., Podlaski, W. F., and Keemink, S. W. (2021). Nonlinear computations in spiking neural networks through multiplicative synapses. *Peer Community Journal*, 1.

Naumann, L. B., Keijser, J., and Sprekeler, H. (2022). Invariant neural subspaces maintained by feedback modulation. *Elife*, 11:e76096.

Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., Pinto, L., Gauthier, J. L., Brody, C. D., and Tank, D. W. (2021). Geometry of abstract learned knowledge in the hippocampus. *Nature*, 595(7865):80–84.

Okazawa, G., Hatch, C. E., Mancoo, A., Machens, C. K., and Kiani, R. (2021). Representational geometry of perceptual decisions in the monkey parietal cortex. *Cell*, 184(14):3748–3761.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'Keefe, J. and Nadel, L. (1978). The hippocampus as a cognitive map.

Olypher, A., Lánskỳ, P., and Fenton, A. (2002). Properties of the extra-positional signal in hippocampal place cell discharge derived from the overdispersion in location-specific firing. *Neuroscience*, 111(3):553–566.

O'Keefe, J. and Krupic, J. (2021). Do hippocampal pyramidal cells respond to nonspatial stimuli? *Physiological Reviews*, 101(3):1427–1456.

Pettersen, M., Rogge, F., and Lepperød, M. E. (2024). Learning place cell representations and context-dependent remapping. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Podlaski, W. F. and Machens, C. K. (2024a). Approximating nonlinear functions with latent boundaries in low-rank excitatory-inhibitory spiking networks. *Neural Computation*, 36(5):803–857.

Podlaski, W. F. and Machens, C. K. (2024b). Storing overlapping associative memories on latent manifolds in low-rank spiking networks. *arXiv preprint arXiv:2411.17485*.

Quirk, G. J., Muller, R. U., and Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, 10(6):2008–2017.

Quiroga, R. Q. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173–185.

Raju, R. V., Guntupalli, J. S., Zhou, G., Wendelken, C., Lázaro-Gredilla, M., and George, D. (2024). Space is a latent sequence: A theory of the hippocampus. *Science Advances*, 10(31):eadm8470.

Recanatesi, S., Farrell, M., Lajoie, G., Deneve, S., Rigotti, M., and Shea-Brown, E. (2021). Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature communications*, 12(1):1417.

Renart, A. and Machens, C. K. (2014). Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–220.

Rennó-Costa, C. and Tort, A. B. (2017). Place and grid cells in a loop: implications for memory function and spatial coding. *Journal of Neuroscience*, 37(34):8062–8076.

Rich, P. D., Liaw, H.-P., and Lee, A. K. (2014). Large environments reveal the statistical structure governing hippocampal representations. *Science*, 345(6198):814–817.

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590.

Rozell, C. J., Johnson, D. H., Baraniuk, R. G., and Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, 20(10):2526–2563.

Sadeh, S. and Clopath, C. (2022). Contribution of behavioural variability to representational drift. *Elife*, 11:e77907.

Samsonovich, A. and McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15):5900–5920.

Sanders, H., Wilson, M. A., and Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *Elife*, 9:e51140.

Saxena, S. and Cunningham, J. P. (2019). Towards the neural population doctrine. *Current opinion in neurobiology*, 55:103–111.

Shapiro, M. L., Tanila, H., and Eichenbaum, H. (1997). Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7(6):624–642.

Sharma, S., Chandra, S., and Fiete, I. (2022). Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. In *International Conference on Machine Learning*, pages 19658–19682. PMLR.

Sheintuch, L., Geva, N., Baumer, H., Rechavi, Y., Rubin, A., and Ziv, Y. (2020). Multiple maps of the same spatial context can stably coexist in the mouse hippocampus. *Current biology*, 30(8):1467–1476.

Soldatkina, O., Schönsberg, F., and Treves, A. (2021). Challenges for place and grid cell models. *Computational Modelling of the Brain: Modelling Approaches to Cells, Circuits and Networks*, pages 285–312.

Solstad, T., Moser, E. I., and Einevoll, G. T. (2006). From grid cells to place cells: a mathematical model. *Hippocampus*, 16(12):1026–1031.

Stefanini, F., Kushnir, L., Jimenez, J. C., Jennings, J. H., Woods, N. I., Stuber, G. D., Kheirbek, M. A., Hen, R., and Fusi, S. (2020). A distributed neural code in the dentate gyrus and in ca1. *Neuron*, 107(4):703–716.

Sun, C., Yang, W., Martin, J., and Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature neuroscience*, 23(5):651–663.

Sun, W., Winnubst, J., Natrajan, M., Lai, C., Kajikawa, K., Michaelos, M., Gattoni, R., Stringer, C., Flickinger, D., Fitzgerald, J. E., et al. (2025). Learning produces a hippocampal cognitive map in the form of an orthogonalized state machine. *Nature*, pages 1–11.

Tanila, H., Shapiro, M. L., and Eichenbaum, H. (1997). Discordance of spatial representation in ensembles of hippocampal place cells. *Hippocampus*, 7(6):613–623.

Tessereau, C., O'Dea, R., Coombes, S., and Bast, T. (2021). Reinforcement learning approaches to hippocampus-dependent flexible spatial navigation. *Brain and Neuroscience Advances*, 5:2398212820975634.

Treves, A. and Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4(3):374–391.

Trouche, S., Perestenko, P. V., van de Ven, G. M., Bratley, C. T., McNamara, C. G., Campo-Urriza, N., Black, S. L., Reijmers, L. G., and Dupret, D. (2016). Recoding a cocaine-place memory engram to a neutral engram in the hippocampus. *Nature neuroscience*, 19(4):564–567.

Wang, Z., Di Tullio, R. W., Rooke, S., and Balasubramanian, V. (2024). Time makes space: Emergence of place fields in networks encoding temporally continuous sensory experiences. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Weber, S. N. and Sprekeler, H. (2018). Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *Elife*, 7:e34560.

Whittington, J. C., McCaffary, D., Bakermans, J. J., and Behrens, T. E. (2022). How to build a cognitive map. *Nature neuroscience*, 25(10):1257–1272.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. (2020). The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., and O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876.

Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613–616.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., and Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3):623–633.

Zhang, K., Ginzburg, I., McNaughton, B. L., and Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *Journal of neurophysiology*, 79(2):1017–1044.

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., Gamal, A. E., and Schnitzer, M. J. (2013). Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3):264–266.

Materials and methods

The Methods is divided into four sections. First, in the General formulation (Section 1), we describe the constrained model that maps environmental variables to neural activity through a latent space (Fig. 1), followed by how this leads to three distinct types of remapping (Fig. 2). In Examples of remapping (Section 2), we describe how this model is adapted to each of the examples described in Figs. 3-6. Then, in Simulations (Section 3), we describe the specifics of the RNN model that we used, and how it is adapted to produce the simulation results in Figs. 3-6. Lastly, in Data analysis (Section 4), we describe the methods used to compute statistical measures of overlap and spatial correlation on the simulated data.

1 General formulation

1.1 Mapping spatial position to neural activity through a latent state space

To constrain the types of remapping and allow for visualizations, we introduce the latent variables $\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c})$. Then, we consider the transformation from $\mathbf{p}, \mathbf{c}$ to the latent variables $\mathbf{z_p}, \mathbf{z_c}$ (nonlinear, fixed across environments) and from the latent variables $\mathbf{z_p}, \mathbf{z_c}$ to the rates $\mathbf{r}$ (linear decoding, variable across environments).

1.1.1 Nonlinear (angular) encoding

For the fixed non-linear mapping between the environmental variables $\mathbf{p}, \mathbf{c}$ and the latent variables $\mathbf{z_p}, \mathbf{z_c}$, we choose an angular encoding. We consider the environmental variables to be normalized as $\forall i,\ p_i, c_i \in [-1, 1]$. Then we have an encoding and decoding for each component:

Angular encoding:

$$\begin{aligned} \mathbf{z}_{p_{2i,2i+1}}(p_i) : [-1,1] &\to [0, 2\pi] \to S_1 \\ p_i \mapsto \alpha_i = \pi(p_i + 1) &\mapsto \mathbf{z}_{p_{2i,2i+1}}(p_i) = (\cos(\alpha_i), \sin(\alpha_i)) \end{aligned} \tag{4}$$

Angular decoding:

$$\begin{aligned} p_i(\mathbf{z}_{p_{2i,2i+1}}) : S_1 &\to [0, 2\pi] \to [-1, 1] \\ \mathbf{z}_{\mathbf{p}_{2i,2i+1}} \mapsto \alpha_i = \operatorname{atan2}(\mathbf{z}_{\mathbf{p}_{2i+1}}, \mathbf{z}_{p_{2i}}) + \pi &\mapsto p_i(\mathbf{z}_{p_{2i,2i+1}}) = 2\frac{\alpha_i}{2\pi} - 1 \end{aligned} \tag{5}$$

where $S_1 \subset \mathbb{R}^2$ is the unit circle in the plane and $\operatorname{atan2}(y, x)$ returns the angle of the vector (x, y) in the range $[-\pi, \pi]$. This is done equivalently for $\mathbf{z_c}$ from and to $\mathbf{c}$.

Generally, we can consider P positional variables and C cognitive variables. For position, we have $\mathbf{p} \in [-1, 1]^P$, leading to $\mathbf{z_p} \in S_1 \times S_1 \times ... \times S_1 \subset \mathbb{R}^{2P}$. For cognitive variables, we have $\mathbf{c} \in [-1, 1]^C$, leading to $\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c}) \in S_1 \times S_1 \times ... \times S_1 \subset \mathbb{R}^{2P+2C}$. For most of the visualizations we consider a one-dimensional position and contextual variables $p, c \in [-1, 1]$ and therefore $\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c}) \in S_1 \times S_1 \subset \mathbb{R}^4$, a torus.

1.1.2 Linear decoding and pseudo-linear encoding

We now consider the variable linear mapping between the latent variables $\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c})$ and rates $\mathbf{r}$.

Linear decoding:

$$\mathbf{z}(\mathbf{r}) : \mathbb{R}^N \to \mathbb{R}^Z \tag{6}$$

$$\mathbf{r} \mapsto \mathbf{z} = \mathbf{D}\mathbf{r} \tag{7}$$

Pseudo-linear encoding:

$$\mathbf{r}(\mathbf{z}) : \mathbb{R}^Z \to \mathbb{R}^N \tag{8}$$

$$\mathbf{z} \mapsto \mathbf{r} = \mathbf{D}^R\mathbf{z} + \boldsymbol{\nu}(\mathbf{z}) \tag{9}$$

where $\mathbf{D} \in \mathbb{R}^{Z \times N}$ is the decoding matrix, $\mathbf{E} = \mathbf{D}^R \in \mathbb{R}^{N \times Z}$ is its right inverse (also referred to as $\mathbf{D}^+$ in the main text), and $\boldsymbol{\nu}(\mathbf{z}) \in \mathbb{R}^N$ is the non-linear component in the nullspace of $\mathbf{D}$. To see that the pseudo-linear encoder is equivalent to the linear decoder, we can multiply Eq. 9 by $\mathbf{D}$, to recover Eq. 7, noting that $\mathbf{D}\mathbf{D}^R = \mathbf{I}$ and $\mathbf{D}\boldsymbol{\nu}(\mathbf{z}) = \mathbf{0}$.

1.2 Three types of remapping

The angular coding is assumed to be fixed, as it defines the coordinates of the internal representation of the environmental variables. Thus we can consider the transformation from $\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c})$ to rates $\mathbf{r}$ to be variable across environments. In doing so, we can use an environment-specific pseudo-linear encoder equation to get:

$$\mathbf{r}^A(\mathbf{p}) = \mathbf{E}^A \mathbf{z}^A(\mathbf{p}) + \boldsymbol{\nu}^A(\mathbf{z}^A(\mathbf{p})) \tag{10}$$

which allows us to distinguish the different kinds of remapping.

Note that the non-linear part of the encoding $\boldsymbol{\nu}^A$ will also change with $\mathbf{E}^A$, as changing the linear encoding will also affect the nullspace. Therefore, this type differentiates itself from nullspace remapping in that it considers changes both to $\mathbf{E}^A$ and $\boldsymbol{\nu}^A$, while nullspace remapping considers changes exclusively to $\boldsymbol{\nu}^A$.

2 Examples of remapping

Here we will use the general formulation defined in Section 1 to study the three different types of remapping described in the main text.

2.1 Encoder-decoder (ED) remapping

We first consider encoder-decoder (ED) remapping, i.e. changes in the environment identity, leading to changes in $\mathbf{E}^A$, and thus changes in $\mathbf{r}^A$. In these cases, for simplicity, we assume there are no cognitive variables being encoded, i.e. $C = 0$ and therefore $Z = P$ and $\mathbf{z} = \mathbf{z_p}$. This results in the stable map:

$$\mathbf{r}^A(\mathbf{p}) = \mathbf{E}^A \mathbf{z_p}(\mathbf{p}) + \boldsymbol{\nu}^A(\mathbf{z_p}(\mathbf{p})) \tag{11}$$

and equivalently for $\mathbf{r}^B(p)$.

In practice, to implement this without changing the connectivity of the network, we do not directly change the matrix $\mathbf{D}$ (and $\mathbf{E}$) of the RNN model. Instead we define an *embedding space*, $\mathbf{y}$, that contains all subspaces $\mathbf{E}$ for all environments. To do this, we multiply the latent variable trajectory by a transformation matrix to obtain the trajectory in embedding space as

$$\mathbf{y}^A(\mathbf{p}) = \mathbf{R}^A \mathbf{z_p}(\mathbf{p}), \tag{12}$$

which is then given as input to the model. We can thus adapt the pseudo-linear encoder as

$$\mathbf{r}^A(\mathbf{p}) = \mathbf{E_y} \mathbf{y}^A(\mathbf{p}) + \boldsymbol{\nu}^A_{\mathbf{y}}(\mathbf{y}^A(\mathbf{p})) = \mathbf{E_y} \mathbf{R}^A \mathbf{z_p}(\mathbf{p}) + \boldsymbol{\nu}^A(\mathbf{z_p}(\mathbf{p})) \tag{13}$$

We will consider two cases, one with random encoder/decoder matrices (akin to classic models or remapping, e.g., multi-chart continuous attractor neural networks (CANNs)) and another with more than one module in the angular encoding and encoder/decoder matrices restricted per module (akin to grid cell realignment models of remapping).

2.1.1 Classical models of remapping: random R

We want to model the classical way of thinking about remapping as a consequence of having a network that tracks position ($\mathbf{z_p}$ in this case) and whose representation changes **randomly** across environments. Importantly, the identity of the environment is not usually explicitly coded in the network but fed by an exterior component. Therefore, to implement this we sample randomly $\mathbf{R} \sim U_{ortho}(Y \times P)$ where Y is a hyperparameter that sets the *dimensionality* of the embedding space and $U_{ortho}(Y \times P)$ is the uniform distribution of matrices $Y \times P$ orthonormal in the columns. We consider two cases for Y:

High-dimensional case, $Y = N$ (Fig. S2): Like in multi-chart CANNs, we can consider a random association between each neuron and its preferred position per environment. The random preferred position of each neuron in a particular environment can be encoded in the encoding matrix $\mathbf{E}^A = \mathbf{E_y} \mathbf{R}^A$ such that:

$$\mathbf{R} \sim U_{ortho}(N \times P) \tag{14}$$

where $U_{ortho}(N \times P)$ is the uniform distribution of matrices $N \times P$ orthonormal in the columns. Two additional technical comments about this model and the relationship with multi-chart CANNs are:

- First, if we only set $Y = N$ i.e. if the rows of $\mathbf{E_y}$ are not also forced to be linearly independent, we would be inducing relationships between neurons such that the remapping would not be fully random. Therefore $\mathbf{E_y}$ has to be fully orthonormal, and as such, for simplicity, we choose $\mathbf{E_y} = Id_{N \times N}$ and even though we do not change the parameters of the network, we get functionally $\mathbf{E}^A = \mathbf{R}^A$.

- Second, in CANNs, neurons appear always equally distributed and with the same place field size (since the way to model it is usually with a grid of space). For this the matrix $\mathbf{R}^A$ sampled needs to have the property that the sampled rows (understood as N P-dimensional vectors) are equally spaced in the latent space $\mathbf{z_p}$. In this work, we will not impose this and be more general.

Low-dimensional case, $Y < N$ (Fig. 3a-f): We then propose a simpler case where the neurons are correlated. The need for $\mathbf{E}^A$ and $\mathbf{E}^B$ to be uncorrelated arises in CANNs from the need to have a stable bump of activity, and to have true random global remapping. So we considered what happens if the different encoding matrices are correlated. This happens if the dimensionality of the embedding space Y is smaller than the number of neurons N, since this enforces that neurons are not orthogonal to each other in this space, such that:

$$\mathbf{R} \sim U_{ortho}(Y \times P) \tag{15}$$

where $U_{ortho}(Y \times P)$ is the uniform distribution of matrices $Y \times P$ orthonormal in the columns. In this case we can visualize the embedding space $\mathbf{y}$ directly (e.g. the sphere in Fig. 3a,b).

2.1.2 Grid cell realignment: more than one module + module phase shift R

We then would like to model remapping as the changes of grid cell input to/conjunctive code with place cells. To simulate grid cell input or participation in the representation, when doing the angular encoding (see Methods Section 1.1.1) instead of using a single angular module to encode $\mathbf{p} \in \mathbb{R}^P$ in $\mathbf{z_p} \in \mathbb{R}^{2P}$ we use $m > 1$ modules, such that $\mathbf{z_p} \in \mathbb{R}^{2mP}$. This gives us a more complex generalization of the angular encoding before (except in this example this reduces to the same with $m = 1$ modules).

Angular encoding with m modules:

$$\begin{aligned}\mathbf{z}_{\mathbf{p}_{2mi,\dots,2m(i+1)-1}}(\mathbf{p}_i) :&[-1,1] \to [0,f_0] \times [0,f_1] \times \dots \times [0,f_{m-1}] \to S_1 \times S_1 \times \dots \times S_1 \\ &\mathbf{p}_i \mapsto (\alpha_{i,j})_{j=0}^{m-1}, \alpha_{i,j} = \left(\frac{3}{2}\right)^{f_j} \pi(\mathbf{p}+1) \mapsto (\cos(\alpha_{i,0}), \sin(\alpha_{i,0}), \dots, \cos(\alpha_{i,m-1}), \sin(\alpha_{i,m-1}))\end{aligned} \tag{16}$$

where f_j is some parameter of the module index, and enforcing that always $f_0 = 0$.

For decoding, while there are complex methods to decode from several angular modules encoding a linear signal, here we decided to use a simplification. We added the restriction that $f_0 = 0$ always, which means the first module will be restricted to $\alpha_0 \in [0, 2\pi]$ and form a bijection $\mathbf{p} \in [-1,1] \longleftrightarrow \mathbf{z} \in S_1$. However, the other modules can have $\alpha_j \in [0, a]$ with $a \neq 2\pi$ and therefore constitute either an arc of S_1 or many revolutions in S_1.

Angular decoding with m modules:

$$\begin{aligned}\mathbf{p}_i(\mathbf{z}_{\mathbf{p}_{2mi,\dots,2m(i+1)-1}}) :&S_1 \times S_1 \times \dots \times S_1 \to [0,2\pi] \to [-1,1] \\ &\mathbf{z}_{\mathbf{p}_{2mi,\dots,2m(i+1)-1}} \mapsto \alpha_i = \text{atan2}(\mathbf{z}_{\mathbf{p}_{2mi}}, \mathbf{z}_{\mathbf{p}_{2mi+1}}) + \pi \mapsto 2\frac{\alpha_i}{2\pi} - 1\end{aligned} \tag{17}$$

Once we have our m grid cell modules, we want to simulate phase shifts in the grid modules relating to remapping, for that we can sample:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{0}_{2\times 2} & \cdots & \mathbf{0} \\ \mathbf{0}_{2\times 2} & \mathbf{R}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_m \end{pmatrix} \quad \text{with} \quad \mathbf{R}_j \sim Rot(2 \times 2) \tag{18}$$

where $Rot(2 \times 2)$ is the uniform distribution over rotational matrices 2×2.

2.2 Mixed-selective remapping

Second, we consider cases that fall into the category of mixed-selective remapping, i.e. changes in the cognitive variables in each environment, lead to changes in $\mathbf{z}_\mathbf{c}^A(\mathbf{p})$, that lead to changes in $\mathbf{r}^A$. This gives us the stable map:

$$\mathbf{r}^A = \mathbf{E}\mathbf{z}^A + \boldsymbol{\nu}(\mathbf{z}^A) = \mathbf{E}\mathbf{z_p} + \mathbf{E}\mathbf{z}_\mathbf{c}^A + \boldsymbol{\nu}(\mathbf{z_p}, \mathbf{z}_\mathbf{c}^A) \tag{19}$$

We will consider two cases: (i) space-feature coding, with both positional and environmental variables $P \neq 0 \wedge C \neq 0$, and (ii) implicit-space coding, with only environmental variables $P = 0 \wedge C \neq 0$ (akin to purely sensory/memory models of remapping).

2.2.1 Space-feature coding: $P \neq 0$

In the first place we state there is an explicitly (albeit complex and non-linear) representation of space, i.e. $P \neq 0$. Then we have to define how we generate the functions $\mathbf{z}_{\mathbf{c}}^{A}(\mathbf{p})$. We define C cognitive variables as:

$$\mathbf{c}(\mathbf{p}) = \mathbf{k} + g(\mathbf{p}) \quad \text{with} \quad \mathbf{k} \sim U([-1,1]^{C}) \quad \text{and} \quad g(\mathbf{p}) \sim GP(\mathbf{0}, K) \tag{20}$$

with GP being a Gaussian process with kernel K.

2.2.2 Implicit-space coding: $P = 0$

In the second we state there is no explicit representation of space, i.e. $P = 0$, and therefore place cells only arise from the high correlation between space and the cognitive variables. This means:

$$\mathbf{r}^{A} = \mathbf{E}\mathbf{z}_{\mathbf{c}}^{A} + \boldsymbol{\nu}(\mathbf{z}_{\mathbf{c}}^{A}) \tag{21}$$

We generate the cognitive variables in the same way as in the True Mixed-selective case.

2.2.3 Space-feature reward coding: $P \neq 0$

For the last example of the paper (Fig. 6), we use space-feature coding to simulate a location reward task. We define a single cognitive variable representing reward as:

$$\mathbf{c}(\mathbf{p}) = \text{pdf}_{\mathcal{N}(\mu_r, \sigma \cdot Id_P)}(\mathbf{p}) / \max_{pdf} - 1 \tag{22}$$

where σ is the spread of the reward and $\mu_r \sim U([-1,1]^{P})$ is the location of the reward in this particular environment. By dividing by $\max_{pdf}$, the maximum value of the pdf for a $\sigma \cdot Id$ covariance, and subtracting 1 we ensure that $\mathbf{c}(\mathbf{p}) \in [-1, 0]$. This is an arbitrary decision for the visualization where -1 represents no reward and 0 represents full reward. These convention could of course be modified and e.g. different reward magnitudes or spreads studied. For the visualization to be clearer we lift it slightly with $\mathbf{c}(\mathbf{p}) \leftarrow \mathbf{c}(\mathbf{p}) + 0.05$.

2.3 Nullspace remapping

Third, we consider cases that fall into the category of null-space remapping, i.e. changes in the network, lead to changes in $\boldsymbol{\nu}^{A}$, that lead to changes in $\mathbf{r}^{A}$. Then the stable maps are:

$$\mathbf{r}^{A} = \mathbf{E}\mathbf{z} + \boldsymbol{\nu}^{A}(\mathbf{z}) \tag{23}$$

We consider a case where we modify the non-linear component $\boldsymbol{\nu}^{A}(\mathbf{z})$ by increasing the thresholds of all neurons active in that environment (akin to experimental inhibition-driven recruitment). See the following section for more details.

3 Simulations

To simulate the cases described above we use a recurrent neural network (RNN) model based on the spike-coding network (SCN), an efficient spiking autoencoder network with competitive interactions between neurons (Boerlin et al., 2013; Calaim et al., 2022). Rather than simulate the spiking version of the network, we opt to simulate a firing-rate version, whose steady-state firing rates for a given input can be found by solving a quadratic program (Barrett et al. (2013, 2016); Mancoo et al. (2020); see below).

We model the network as encoding a latent input signal $\mathbf{y}$, such that the linear read-out $\hat{\mathbf{y}} = \mathbf{D}_{\mathbf{y}}\mathbf{r}$ is closely tracking this input, i.e. $\hat{\mathbf{y}} \approx \mathbf{y}$. The decoder matrix and network connectivity, determined by the matrix $\mathbf{D}$, determines the embedding of the input signal $\mathbf{y}$ into the network. Thus, simulation of the network entails constructing a desired input $\mathbf{y}^{A}$ for each environment A. In this section we discuss how we construct $\mathbf{y}^{A}$, how we set-up the network, and how we do the decoding.

3.1 Construction of the input y

3.1.1 Spatial position, p

We grid the space by considering a number of positions $npts$ that splits the space equally. If $p \in [-1, 1]$ then $j \in [0, ..., npts - 1]$ $p(j) = j(2/(npts - 1)) - 1$. If $\mathbf{p} \in [-1, 1]^{2}$ then $i, j \in [0, ..., \sqrt{npts} - 1]$ $\mathbf{p}(i, j) = (2/(npts - 1)) - 1 \cdot (i, j)$.

3.1.2 Cognitive variables, c

In the cases where $C \neq 0$, i.e. mixed-selective remapping (Section 2.2) we have to construct the environmental functions $\mathbf{c}^A(\mathbf{p})$ for $i = [0, ..., C-1]$.

For mixed-selective remapping, we use the spatial grid computed previously (3.1.1) to define the functions $\mathbf{c}^A(\mathbf{p}_{grid}) = k + g(\mathbf{p}_{grid})$ with $k \sim U([-1,1])$ and $g(\mathbf{p}_{grid}) \sim GP(\mathbf{0}, K_{SE}(\cdot,\cdot;\sigma,l)$ with $K_{SE}(\mathbf{x},\mathbf{x}') = \sigma^2 e^{-\frac{d(\mathbf{x},\mathbf{x}')^2}{2l^2}}$ and $d(\mathbf{x},\mathbf{x}') = \|\mathbf{x}-\mathbf{x}'\|^2$. To make sure these functions are bounded $\mathbf{c}_i \in [-1,1]$ we constrain them consistently with a S_1 latent space, i.e.: $\mathbf{c} \leftarrow (\mathbf{c}+1) \mod 2-1$.

For reward coding we use the grid of the space computed previously to define the function of reward in all the environment $\mathbf{c}^A(\mathbf{p}_{grid}) = \text{pdf}_{\mathcal{N}(\mu_r^A, \sigma \cdot Id_P)}(\mathbf{p}_{grid}) / \max_{pdf} - 1$ with $\mu_r^A \sim U([-1,1]^P)$, the location of the reward in this particular environment.

3.1.3 Latent variables, z

We use the angle encoding (see Section 1.1.1) of the animals path $\mathbf{p}$ and the environmental variables $\mathbf{c}^A$ to construct $\mathbf{z_p}$ and $\mathbf{z_c}^A$ respectively, and therefore $\mathbf{z}^A = (\mathbf{z_p}, \mathbf{z_c}^A)$ (for the cases with $P = 0$ and $C = 0$ we construct only $\mathbf{z_c}^A$ and $\mathbf{z_p}$ respectively).

The only parameter in this step here is the number m of modules used that is always $m = 1$ except for the grid realignment case (Section 2.1.2) and the frequency parameters of the modules $f_1, ..., f_{m-1}$ (reminder that $f_0 = 0$ always).

3.1.4 Input to the network, y

In sum, the input to the network is constructed in the following way for each case:

- Encoding-decoding remapping: $\mathbf{y}^A = \mathbf{R}^A \mathbf{z_p}$ with
  - High dimension: $\mathbf{R} \sim U_{ortho}(N \times N)$
  - Low dimension: $\mathbf{R} \sim U_{ortho}(Y \times N)$, $Y < N$ a hyperparameter.
  - Grid cell phase shift: $\mathbf{R} = diag(\mathbf{R}_j)$ with $\mathbf{R_j} \sim Rot(2 \times 2)$
- Mixed-selective remapping: $\mathbf{y}^A = \mathbf{z}^A$
  - True mixed-selective (and reward coding): $\mathbf{y}^A = \mathbf{z}^A = (\mathbf{z_p}, \mathbf{z_c}^A)$
  - Sensory/memory: $\mathbf{y}^A = \mathbf{z}^A = \mathbf{z_c}^A$
- Nullspace remapping: $\mathbf{y}^A = \mathbf{R}\mathbf{z}^A = \mathbf{R}\mathbf{z_p}^A$ with a fixed random $\mathbf{R} \in \mathbb{R}^{Y \times N}$ with $Y < N$ (as in encoding-decoding low dim. remapping).

3.1.5 Normalization of y

The way that $\mathbf{y}$ is constructed from $\mathbf{z} \in S_1 \times S_1 \times .. \times S_1$, we get that $\forall t \quad \|\mathbf{y}(t)\| = 1$. However, if the dimensionality of $\mathbf{y} \in \mathbb{R}^Y$ is large, the network will not track the signal with high fidelity (Calaim et al., 2022). Because of this we use the scaling

$$\mathbf{y} \leftarrow \sqrt{Y}\mathbf{y} \tag{24}$$

Finally, in cases of MS remapping with space-feature coding (Sections 2.2.1 & 2.2.3) where $\mathbf{y} = \mathbf{z} = (\mathbf{z_p}, \mathbf{z_c})$, we normalize each part separately such that $\|\mathbf{y}_{j \in P}\| = \|\mathbf{y}_{j \in C}\| = 1/\sqrt{2}$.

3.2 Recurrent neural network (RNN) model

3.2.1 Simulation

We simulate a firing-rate version of the spike-coding network (SCN) model (Boerlin et al., 2013), which constrains dynamics to a low-dimensional latent space due to its low-rank connectivity (Mastrogiuseppe and Ostojic, 2018). Interestingly, the steady-state firing rates of the SCN model can be rephrased as a convex optimization problem (Barrett et al., 2013, 2016; Mancoo et al., 2020). Specifically for each position $\mathbf{p}$ we use CVXPY (Diamond and Boyd, 2016) to solve the constrained optimization problem:

$$\begin{aligned} \mathbf{r} = \min \|\mathbf{y} - \mathbf{D}\mathbf{r}\|^2 + 2\mathbf{T}^\top \mathbf{r} \\ \text{subj. to} \quad \mathbf{r} \geq \mathbf{0} \end{aligned} \tag{25}$$

using $\mathbf{y}(\mathbf{p})$ as a parameter and obtaining the feasible optimal solution $\mathbf{r}(\mathbf{p})$.

3.2.2 Network Parameters

The primary network parameters are the decoding weights $\mathbf{D}$. We consider two cases:

- $\mathbf{D} = \mathbf{Id}_{N\times N}$ for the case of high dimensional, classical models, encoding-decoding remapping (see Section 2.1.1)
- $\mathbf{D} \sim U_{norm}(Y \times N)$ for all other cases. Here U_{norm} represents the uniform distribution over the set of matrices that have normalized columns, i.e. $\forall j \in N \quad \|\mathbf{D}_j\| = 1$.

Following this, we apply a normalization to the decoding weights depending on whether or not we are simulating a mixed code (M) or conjunctive code (C) as follows:

- Mixed code (M): no additional normalization
- Conjunctive code (C): we normalise each pair of dimensions, i.e. $\forall j \in [0, ..., Y/2-1] \quad \|\mathbf{D}_{2j,2j+1}\| = 1/\sqrt{Y/2}$
- Conjunctive-conjunctive code (CC): we normalise each pair of dimensions, i.e. $\forall j \in [0, ..., Y/2-1] \quad \|\mathbf{D}_{2j,2j+1}\| = 1/\sqrt{Y/2}$ and then normalise $\|\mathbf{D}_{j\in P}\| = \|\mathbf{D}_{j\in C}\| = 1/\sqrt{2}$
- Conjunctive-mixed code (CM): we normalise each pair of dimensions only of $\mathbf{z_p}$ related dimensions, i.e. $\forall j \in [0, ..., 2P/2-1] \quad \|\mathbf{D}_{2j,2j+1}\| = 1/\sqrt{2P/2}$ and then normalise $\|\mathbf{D}_{j\in P}\| = \|\mathbf{D}_{j\in C}\| = 1/\sqrt{2}$
- Pure and Mixed code (pM): we choose the first $pure_P$ neurons to be pure positional, i.e. $\mathbf{D}_{\{0,...,pure_P-1\},j\in C} = \mathbf{0}$, the next $pure_C$ neurons to be pure cognitive, i.e. $\mathbf{D}_{\{pure_P,...,pure_C-1\},j\in P} = \mathbf{0}$ and the rest we do not modify. We make sure by renormalising that $\forall j$, $\|\mathbf{D}_j\| = 1$.

In the case of null space remapping, we do a single simulation, record the active neurons in that space *active* and then set all the thresholds $\mathbf{T}_{active} = -current$ and repeat the simulation.

3.3 Decoding

Once we obtain a firing-rate trajectory $\mathbf{r}^A$, we decode an estimate $\hat{\mathbf{y}}^A = \mathbf{D}\mathbf{r}^A$. Then from $\hat{\mathbf{y}}^A$ we obtain $\hat{\mathbf{z_p}}^A$ by reversing the construction:

- Encoder-decoder remapping: $\hat{\mathbf{z_p}}^A = \mathbf{R}^T\hat{\mathbf{y}}^A$
- Mixed-selective remapping:
  - Space-feature coding (& reward coding): $\hat{\mathbf{z_p}}^A = \hat{\mathbf{y}}^A_{0,...,2P}$
  - Implicit-space coding: we cannot obtain $\hat{\mathbf{z_p}}^A$ out of it in our case (see Discussion)
- Nullspace remapping: $\hat{\mathbf{z_p}}^A = \mathbf{R}^T\hat{\mathbf{y}}^A$

Finally, by definition of the latent space (see Section 1.1) we can decode the animals position $\hat{p}_0 = atan2(\hat{\mathbf{z_p}}_{0,1})/\pi - 1$ and $\hat{p}_1 = atan2(\hat{\mathbf{z_p}}_{2,3})/\pi - 1$.

3.4 Figures and simulations parameters

	High dim.	Low dim.	Low dim vis.	Grid cells	Grid cells vis.
P	2		1	2	1
Y	128	128	3	12	4
$red = \frac{N}{Y}$	1	16	8	8	8
N	128	2048	24	96	24
$\mathbf{D}$	$\mathbf{Id}_{N\times N}$	$U_{norm}(Y \times N)$		$U_{norm}(Y \times N)$	
C or M	M	M	M	C	C
$\mathbf{z_p}$		$m = 1$		$m = 3, f_j \in [0, 1, -1]$	$m = 2, f_j \in [0, 1]$
$\mathbf{z}$		$\mathbf{z} = \mathbf{z_p}$		$\mathbf{z} = \mathbf{z_p}$	
$\mathbf{y} = \mathbf{R}\mathbf{z}$	$U_{ortho}(N \times N)$	$U_{ortho}(Y \times N)$		$diag(Rot(2 \times 2))$	
envs	10		2	10	2

Table 1: Encoder-decoder (ED) remapping (Fig. 3)

	True mixed	True mixed vis.	Reward vis.	Sensory	Sensory vis.
P	2	1	1	2	1
Y	64	4	4	64	4
$red = \frac{N}{Y}$	16	8	4	16	8
N	1024	32	16	1024	32
$\mathbf{D}$	$U_{norm}(Y \times N)$			$U_{norm}(Y \times N)$	
C or M	CM		CM and pM	M	
$\mathbf{c}$	$\mathbf{k} + GP(\mathbf{0}, K)$, $\sigma = 0.25$, $l = 0.5$		$\text{pdf}_{\mathcal{N}(\mu_r, \sigma Id)}$, $\sigma = 0.01$	$\mathbf{k} + GP(\mathbf{0}, K)$,$\sigma = 0.25$, $l = 0.5$	
$\mathbf{z}$	$\mathbf{z} = (\mathbf{z_p}, \mathbf{z_c})$			$\mathbf{z} = \mathbf{z_c}$	
$\mathbf{y}$	$\mathbf{y} = \mathbf{z}$, $\|\mathbf{y}_{j \in P}\| = \|\mathbf{y}_{j \in C}\| = 1/\sqrt{2}$			$\mathbf{y} = \mathbf{z}$	
envs	10	3	2	10	3

Table 2: Mixed-selective (MS) remapping (Figs. 4 & 6)

	Nullspace	Nullspace vis.
P	2	1
Y	128	3
$red = \frac{N}{Y}$	32	16
N	4096	48
$\mathbf{D}$	$U_{norm}(Y \times N)$	
C or M	M	
$\mathbf{T}$	$\mathbf{T}_{active} = 10.5$	
$\mathbf{z}$	$\mathbf{z} = \mathbf{z_p}$	
$\mathbf{y} = \mathbf{R}\mathbf{z}$	$U_{ortho}(Y \times N)$	
envs	1	1

Table 3: Nullspace (NS) remapping (Fig. 5)

3.5 Code availability

The code for all simulations and analysis presented here can be found in the following GitHub repository: `https://github.com/guillemarsan/RemappingGeometry`

4 Data analysis

For all analysis of place fields from firing rates, we set: $place_field[b] = ratemap[b]$ if $ratemap[b] \geq 10^{-3}$ else 0. This is necessary since the optimization problem can find solutions with arbitrarily small rates. Then, to define the remapping measures we define the following useful functions. First, the *alignment* between two vectors is the cosine angle

$$Alig(x, y) = \frac{x^\top y}{\|x\|\|y\|} = \cos(\alpha), \tag{26}$$

where α is the angle formed by the vectors x and y. Next, given the mean firing rates vector per neuron and per environment $mfr(n, e)$ we define the *overlap* of neuron n_1 in environment e_1 with neuron n_2 in environment e_2 as:

$$Overlap(n_1, e_1, n_2, e_2) = Alig(mfr(n_1, e_1), mfr(n_2, e_2)). \tag{27}$$

Finally, given the place fields per neuron and per environment $place_field(n, e)$ we define the *spatial correlation* of neuron n_1 in environment e_1 with neuron n_2 in environment e_2 as:

$$SpatialCorr(n_1, e_1, n_2, e_2) = Alig(place_field(n_1, e_1), place_field(n_2, e_2)) \tag{28}$$

where α is the angle formed by the vectors $place_field(n_1, e_1)$ and $place_field(n_2, e_2)$.

4.1 Total Overlap

Given two environments e_1, e_2 we consider neurons $j \in \{1, ..., k\} = J$ that are active in **at least one** of the two environments, i.e. $\forall j \in \ \exists i \in \{1, 2\} \ \exists b s.t. place_field(j, e_i, b)! = 0$.

Then we define for that **pair of environments**:

$$TotalOverlap(e_1, e_2) = \sum_{j \in J} Overlap(j, e_1, j, e_2) = \cos(\alpha) \tag{29}$$

Notice that this is equivalent to considering the vectors $mfr(J, e_i) = [mfr(j, e_i) \text{for} j \in J]$ and defining:

$$TotalOverlap(e_1, e_2) = Alig(mfr(J, e_1), mfr(J, e_2)) \tag{30}$$

where α is the angle formed by the k-dimensional vectors $mfr(J, e_1)$ and $mfr(J, e_2)$.

For the shuffling condition we consider 20 random shufflings of the neurons $i \in [0, ..., 19] \quad \sigma_i(J) = \sigma_i(\{1, ..., k\})$ and compute

$$TotalOverlap^{(s)}(e_1, e_2) = \langle \sum_{j \in J} Overlap(j, e_1, \sigma_i(j), e_2) \rangle_i \tag{31}$$

Or equivalently use the vectors of shuffled neurons $mfr^{(s)}(\sigma_i(J), e_2) = [mean_fr(j, e_2) \text{for} j \in \sigma_i(\{1, ..., J\})] \in \mathbb{R}^J$ to compute $TotalOverlap^{(s)}(e_1, e_2) = \langle Alig(mfr(J, e_1), mfr(\sigma(J), e_2)) \rangle_i$.

Finally we compute the mean across pairs of environments $i, j \in 1, ..., envs$ and the standard error of the mean:

$$\begin{aligned} MeanTotalOverlap &= \langle TotalOverlap(e_i, e_j) \rangle_{i,j} \\ MeanTotalOverlap^{(s)} &= \langle TotalOverlap^{(s)}(e_1, e_2) \rangle_{i,j} \\ SEMTotalOverlap &= std_{i,j}(TotalOverlap(e_i, e_j))/envs \\ SEMTotalOverlap^{(s)} &= std_{i,j}(TotalOverlap^{(s)}(e_i, e_j))/envs \end{aligned} \tag{32}$$

4.2 Mean Spatial correlation

In this case, we only consider pairs of environments that have at least one place cell active in common.

Given two environments e_1, e_2 as such, we consider neurons $j \in \{1, ..., k\} = J$ that are active in **both** of the two environments, i.e. $\forall j \ \forall i \in \{1, 2\} \ \exists bs.t.place_field(j, e_i, b)! = 0$.

Then we define for that **pair of environments**:

$$MeanSpaitalCorr(e_1, e_2) = \frac{1}{k} \sum_{j \in J} SpatialCorr(j, e_1, j, e_2) \tag{33}$$

For the shuffling condition we consider 20 random shufflings of **all** the neurons indices $\sigma_i(\{1, ..., N\})$ in the second environment e_2 and then, for each shuffle, consider neurons that are active in **both** environments, i.e. $\exists b_1 s.t.place_field(b_1, j, e_1)! = 0 \wedge \exists b_2 s.t.place_field(b_2, \sigma_i(j), e_2)! = 0$.

$$MeanSpaitalCorr^{(s)}(e_1, e_2) = \frac{1}{k} \sum_{j \in J_i} SpatialCorr(j, e_1, \sigma_i(j), e_2) \tag{34}$$

Finally we compute the mean across pairs of environments $i, j \in 1, ..., envs$ and the standard error of the mean:

$$\begin{aligned} MeanMeanSpatialCorr &= \langle MeanSpatialCorr(e_i, e_j) \rangle_{i \neq j} \\ MeanMeanSpatialCorr^{(s)} &= \langle MeanSpatialCorr^{(s)}(e_i, e_j) \rangle_{i \neq j} \\ SEMMeanSpatialCorr &= std_{i \neq j}(MeanSpatialCorr(e_i, e_j))/envs \\ SEMMeanSpatialCorr^{(s)} &= std_{i \neq j}(MeanSpatialCorr^{(s)}(e_i, e_j))/envs \end{aligned} \tag{35}$$

4.3 Statistical test

For checking the significance of difference to shuffle condition, we use a one-sample t-test, using as data the values of the overlap or mean spatial correlation for each pair of environments (usually 10 environments and therefore ${}^{10}C_2 = 45$ distinct pairs), and as null hypothesis the mean mean spatial correlation from the shuffled data, i.e.:

$$\text{p-value} = \text{T-test}((MeanSpatialCorr(e_i, e_j))_{i \neq j}, \mu_0 = MeanMeanSpatialCorr^{(s)}) \tag{36}$$

Then we conclude it is significant (marked with '*') if the p-value < 0.05.

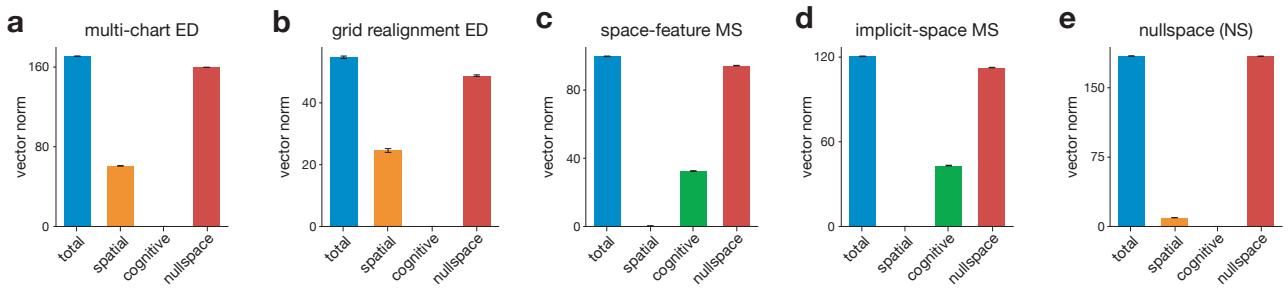

Figure S1: **Comparison of remapping changes in each component of pseudo-linear encoder.** Explanation of how this works, and then shown for multi-chart ED remapping (**a**), grid realignment ED remapping (**b**), space-feature MS remapping (**c**), implicit-space MS remapping (**d**), and nullspace remapping (**e**).

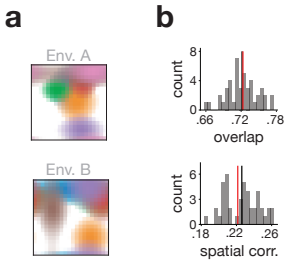

Figure S2: **Multi-chart encoder-decoder (ED) remapping with full-dimensional embedding space.** **a**: Example place field rate maps for two environments. **b**: Overlap and spatial correlation distributions, with mean (black) and comparison with a shuffle control (red).