Analysis of response profiles for electrophysiological data

To calculate the response profile of each neuron as a function of position in the virtual corridor we used a local smoothing method[1–3]. First, the spike count was calculated in a 250 ms window. We then discretized the position of the animal in 2 cm bins, yielding 50 bins and we calculated the spike count map and occupancy map for each neuron. Both the spike count and occupancy maps were smoothed by convolving them with a common Gaussian window whose width was optimized to maximize reliability (see below), and the response profile was calculated as the ratio of the smoothed spike count map and the occupancy map[4].

Response profile reliability was calculated as the fraction of variance in firing rate explained by the response profile:

$$\text{Reliability} = 1 - \frac{\sum_t (y(t) - y'(t))^2}{\sum_t (y(t) - \mu)^2}$$

where $y(t)$ is the firing rate of the neuron at time t, $y'(t)$ was the prediction by the place-field for the same time bin and μ is the mean firing rate of the training data[4]. We used five-fold cross-validation to calculate place fields and reliability. Only neurons with a reliability greater than 0.01 were considered for further analysis.

Analysis of response profiles for two-photon data

To obtain response profiles as a function of position along the corridor, we discretized the position of the animal in 1 cm bins, yielding 100 bins and we calculated the fluorescence map and occupancy map for each neuron. Both the spike count and occupancy maps were smoothed by convolving them with a fixed Gaussian window of 5 cm. Only time points with running speeds greater than 1 cm/s were included in further analyses. For consistency with the response profiles obtained from electrophysiological data, we only looked at responses for which the cross-validated reliability was higher than 0.01. These cells were considered to have activity significantly modulated by position in the corridor. To model single-cell activity under the assumption that responses are identical in the two segments of the corridor, we fit (using least squares) a model function to the response profile along the visually-matching segment where the cell peaked. The model function was the sum of two Gaussians that meet at the peak. To obtain a prediction along the whole corridor, we then duplicated the fitted response at ± 40 cm away from the maximum.

The *Spatial modulation ratio* was measured by splitting the dataset into odd and even trials. For each cell, the position of the peak response was measured from the response profile averaged across odd trials. We then computed the ratio between responses at this position and the visually-identical position 40 cm away, using the response profile averaged across even trials. Cells which had a maximal response too close to the start or the end of the corridor (0-15 cm or 85-100 cm) were not considered for analysis of the ratio of responses. Therefore, this excluded cells which responded too close to the start or the end of the corridor, which were outside the visually-matching segments. Two-dimensional response profiles with respect to position and speed (Extended Data Figure 5c) were calculated as previously described[4].

General linear model analyses

To assess the spatial modulation of V1 neurons while jointly accounting for all other visual and behavioural factors, we fitted each cell's calcium fluorescence trace (a time-dependent continuous variable) to three different general linear (or multilinear ridge regression) models of the form: $\hat{y} = X\hat{\beta}$, where X is an T-by-M matrix with T time points and M predictors, $\hat{y}$ is the predicted calcium trace (T-by-1 array). Optimal coefficient estimates $\hat{\beta}$ (M-by-1 array) that minimise the sum squared error were obtained using: $\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$, where λ is the ridge-regression coefficient.

In the full model, the predictor matrix X contains several sets of columns: a set of spatial basis functions $I_i(x_t)$; pupil position ex_t, ey_t; the speed s_t at 5 time lags; pupil diameter p_t again at 5 time lags; and a step function r_t indicating reward with 4 time lags. A model using all these predictors has the form:

$$y_t = \beta_0 + \sum_{i=1}^{12} \beta_i I_i(x_t) + \beta_{ex} ex_t + \beta_{ey} ey_t + \sum_{j=1}^{5} \beta_{s_j} s_{t+\tau_j} + \sum_{j=1}^{5} \beta_{p_j} p_{t+\tau_j} + \sum_{k=1}^{4} \beta_{r_k} r_{t+\tau_k}$$

where β_0 is a constant.

Predictors were defined similarly to Ref. [5]. The first spatial basis functions corresponded to regions prior to the visually identical segments:

$$I_1(x) = \begin{cases} 1 \; if \; x \in [0,5] \\ 0 \; otherwise \end{cases}$$

$$I_2(x) = \begin{cases} 1 \; if \; x \in [5,10] \\ 0 \; otherwise \end{cases}$$

The basis functions corresponding to visually matching segments consisted of double step functions with weights a and b (defined below):

$$I_3(x) = \begin{cases} a \; if \; x \in [10,15] \\ b \; if \; x \in [50,55] \\ 0 \; otherwise \end{cases}$$

Basis functions $I_4 \ldots I_{10}$ were defined similarly, to cover the x range 15… 50 and 55…90. The final two functions I_{11} and I_{12} covered the non-repeating x ranges [90,95] and [95,100].

The pupil diameter and eye position predictors were all scaled to lie in a range [-1,1], while speed was scaled to the range [0,1]. The pupil diameter and speed predictors were lagged with 5 time shifts: $\tau_j \in \{-1000\; ms\; , -500\; ms\; , 0\; ms, 500\; ms, 1000\; ms\}$. The reward predictor r_t was defined to be 1 within a 500 ms window of the reward, 0 otherwise, and was lagged by $\tau_k \in \{-1000\; ms\; , -500\; ms\; , 0\; ms, 500\; ms\}$

The three prediction models used different combinations of these predictors. The first model, *the visual model*, relied on just visual predictors I_i,with the constraint that $a = b = \sqrt{2}$ such that the basis functions have unit norm. Therefore, this model resulted in responses that repeated in the visually identical segments perfectly.

The second model, *the non-spatial model*, added the influence of all the behavioural factors we measured: running speed ($s_{t+\tau_k}$), reward events ($r_{t+\tau_k}$), pupil size ($p_{t+\tau_k}$), and the horizontal (ex_t) & vertical (ey_t) pupil position.

Finally, the third model, *spatial model*, allowed for an independent scaling of the two identical sections of the room. This model allowed $a \neq b$, subject to the constraint that the spatial basis functions had unit norm. To achieve this, we used exhaustive search over a parameter $\alpha \in [0,1]$, with $a = \frac{\alpha}{\sqrt{\alpha^2+(1-\alpha)^2}}$ and $b = \frac{1-\alpha}{\sqrt{\alpha^2+(1-\alpha)^2}}$. Note that $\alpha = 0.5$ would correspond to a purely visual representation with spatial modulation ratio close to 1, while $\alpha = 1$ or $\alpha = 0$ would correspond to a response only in the first or second segment, and a spatial modulation ratio close to 0.

To fit the parameters, we used the ridge regression coefficient, λ that maximized the percentage of variance explained using five-fold cross-validation, searching the values λ =0.01, 0.05, 0.1, 0.5 or 1. In the spatial model (where $a \neq b$), we performed multiple ridge regression fits, searching for the optimal value of α using a step size of 0.1, for each λ.

The single cell responses predicted by these models were then processed similarly to the original recorded responses to obtain the response profiles and spatial modulation ratio predicted by the three models. The deviation of the model predictions from the original data were evaluated by fitting an ellipse to the distribution and quantified using the angle of its major axis (the first eigenvector of the covariance matrix).

Decoding population activity and calculating correlations between V1 and CA1

Population activity was decoded using an independent Bayes decoder[6]. For every time bin, we calculated the probability of being at a location x given population response R as:

$$P(x|R) = \frac{1}{Z}\, P(x) \left(\prod_{i=1}^{M} f_i(x)^{r_i} \right) exp\left(-t \sum_{i=1}^{M} f_i(x) \right)$$

where $f_i(x)$ is the response profile and r_i is the spike count of the i^{th} neuron in a time bin, M is the number of neurons and t is the time window. Z is a normalizing constant, which makes the probabilities across all positions sum to one[6,7]. The probability of being in the reward zone was calculated by summing the posterior probabilities in the reward zone, and normalised relative to the value in correct trials (Figure 3c and 3f).

When calculating joint distributions (Figures 2 and 3), we smoothed the distribution by a Gaussian window with a width of 4 spatial bins. To account for the effects of position and speed on calculating the correlations between V1 and CA1 decoding errors, we shuffled the data within the time points when the animal was at the same position (within 2cm) and ran in a specific speed range (5 cm/s bins: 5-10 cm/s to 30-35 cm/s).

To determine whether correlations between the V1-encoded and CA1-encoded positions could arise from variables such as speed and reward, we asked whether the prediction of position encoded in one brain region was improved by the position encoded in the other region, even after accounting for all other visual and non-visual variables (Extended Data Figure 10). Specifically, we used a non-linear decoder[8,9] (Tree Bagger regression implementation of random forests from the statistics toolbox of Matlab), to evaluate how well we could predict the position decoded from V1, based on: (i) the actual position of the animal, (ii) running speed, (iii) animal licks, (iv)

rewards, and (v) position estimated from CA1 neurons. The lick and reward events were smoothed by a 50ms Gaussian window before using them as inputs to the decoder. The maximum number of trees used was 50, as we found that performance saturated by that point. To test if CA1 contributed to the prediction of the V1 estimate, we then predicted the V1 decoded positions without the CA1 estimate as an input to the decoder (Extended Data Figure 10).

Simulation of V1 complex cells

Response profiles expected from purely visual neurons were obtained from simulations of a population of complex receptive fields. Complex receptive fields were modelled as two Gabor filters in spatial quadrature (i.e. shifted in spatial phase by 90°) having the same orientation and spatial frequency. Responses were simulated by convolving the VR images at successive positions along the corridor with the pair of Gabor filters and taking the sum of their squared outputs (energy model[10,11]). The receptive fields were designed so to simulate different orientation selectivity (from 0°, 15°, …, 165°; we overrepresented the cardinal orientations) spatial frequency selectivity (0.04, 0.05, 0.06 and 0.07 cycles/°), which are the ranges typically observed in the mouse visual system[12–14]. The receptive fields were simulated to cover azimuths from 40° to 80°, matching the receptive field position of the cells we focused on in our recordings. In addition, the responses generated by the complex cell model had profiles that were similar to the (purely visual) responses observed in the recorded data (Extended Data Figure 3). We did not observe phase specific responses in our recordings (i.e. multiple sharp peaks within each landmark), which would be expected for simple cells.

Supplementary Methods References

1. Loader, C. R. Local likelihood density estimation. *Ann. Stat.* **24,** 1602–1618 (1996).
2. Harris, K., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424,** 552–556 (2003).
3. Harris, K. D. *et al.* Spike train dynamics predicts theta- related phase precession in hippocampal pyramidal cells. *Nature* **417,** 2116–2118 (2002).
4. Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D. & Carandini, M. Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* **16,** 1864–1869 (2013).
5. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. *Cell* **170,** 986–999.e16 (2017).
6. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79,** 1017–44 (1998).
7. Bendor, D. & Wilson, M. A. Biasing the content of hippocampal replay during sleep. *Nat. Neurosci.* **15,** 1439–1444 (2012).
8. Huber, D. *et al.* Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484,** 473–478 (2012).
9. Breiman, L. Random forests. *Mach. Learn.* **45,** 5–32 (2001).
10. Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.* **283,** 53–77 (1978).
11. Carandini, M. What simple and complex cells compute. *J. Physiol.* **577,** 463–466 (2006).

12. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* **65,** 472–479 (2010).

13. Andermann, M. L., Kerlin, A. M., Roumis, D. K., Glickfeld, L. L. & Reid, R. C. Functional specialization of mouse higher visual cortical areas. *Neuron* **72,** 1025–1039 (2011).

14. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72,** 1040–1054 (2011).