Spatial encoding in the visual pathway arises in cortex and depends on active navigation

E. Mika Diamanti[1,2*], Charu Reddy[1], Sylvia Schröder[1], Kenneth D. Harris[3], Aman B. Saleem[4], and Matteo Carandini[1]

[1]UCL Institute of Ophthalmology, University College London, London, EC1V 9EL, UK;
[2]CoMPLEX, Department of Computer Science, University College London, London, WC1E 7JG, UK
[3]UCL Institute of Neurology, University College London, London, WC1N 3BG UK;
[4]UCL Institute of Behavioural Neuroscience, University College London, London, WC1H 0AP, UK
*Correspondence: emdiamanti@princeton.edu

During navigation, responses in primary visual cortex (V1) are modulated by the animal's position. Here, we show that this modulation is present across multiple higher visual cortical areas but largely absent in geniculate inputs to V1. Spatial modulation is stronger during active navigation than during passive viewing. Moreover, navigation activates different neurons than classical drifting gratings, and promotes the reliability of neural responses in parietal and medial cortical areas.

There is increasing evidence that the activity of the primary visual cortex (V1) is influenced by navigational signals[1–3]. During navigation, indeed, the responses of V1 neurons to identical visual stimuli are strongly modulated by the animal's spatial position[4]. It is not known, however, how this modulation varies along the visual pathway. Is it present upstream of V1, as has been suggested[5]? Is it stronger downstream of V1, e.g. in parietal visual areas where selectivity for position is well established[6–12]? And does it require active navigation, a condition that enhances spatial representations in hippocampus[13–15]?

To characterize the influence of spatial position on visual responses, we used a virtual reality (VR) corridor with two visually matching segments[4] (Fig. 1). We used two-photon microscopy to record calcium activity of LGN boutons in layer 4 of area V1 (Fig. 1a, and Suppl. S1a-d), of V1 neurons (Fig. 1b), and of neurons in higher visual areas (Fig. 1c). Mice were head-fixed and ran on a treadmill to traverse a virtual corridor made of two consecutive 40 cm segments that contained the same visual textures[4] (Fig. 1d). A purely visual neuron would respond similarly in these matching segments, while a neuron modulated by spatial position would respond more strongly in one segment.

Spatial position powerfully modulated the visual responses of V1 neurons[4], but barely affected those of LGN afferents (Fig. 1e, f, h, i). LGN boutons gave similar responses in the two segments of the corridor (Fig. 1e), indicating that spatial position had little effect on the LGN responses. By contrast, and as previously reported[4], V1 neurons tended to respond more strongly at a single location in the corridor (Fig. 1f). To quantify the degree of spatial modulation, we defined a spatial modulation index (SMI) as the normalized difference of responses at the two visually matching positions (preferred minus non–preferred, divided by their sum; the preferred position was defined by held-out data). The median SMI across sessions for responsive LGN boutons was barely above zero, at 0.07 ± 0.05 (n = 19 sessions), markedly smaller than the median for responsive V1 neurons, which was 0.41 ± 0.17 (n = 32 sessions; LGN SMI significantly smaller than V1: $p = 10^{-6}$,left-tailed two-sampled *t*-test, Fig. 1h, i).

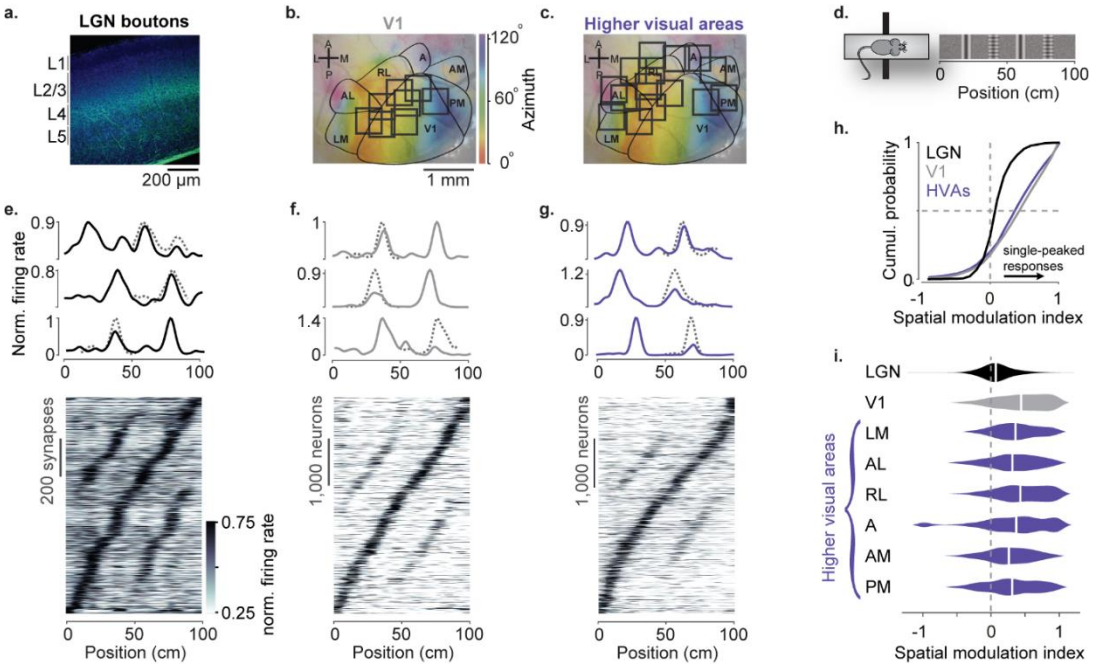


Fig. 1: Spatial position modulates visual activity in visual cortex but not in thalamic LGN.
a. Merged confocal images of somata in V1 (Nissl staining; *blue*) and of LGN axonal projections expressing GCaMP (GFP; *green*). GCaMP expression is densest in layer IV (L4). **b.** Example retinotopic map of the cortical surface with recording sessions targeting (fully or partly) V1 (*black* squares, field of view: 500µm x 500 µm). *Black* contours show borders between areas. **c.** Same retinotopic map as in **b**. with recording sessions targeting (fully or partly) higher visual areas (*black* squares, field of view: 500µm x 500 µm). **d.** Head-restrained mice traversed a virtual corridor by running on a Styrofoam wheel without a specific task. The corridor had two landmarks that repeated after 40 cm, creating visually matching segments. **e.** LGN response profiles as a function of position chosen from the 25th, 50th and 75th percentile of the metric defined in **h** (*top*) and response profiles across the population (*bottom*). Response profiles are obtained from even trials and are normalized and ordered by the position of the maximum response estimated from odd trials. Only response profiles with variance explained ≥ 5% are included (1,140 of 3,182 synapses). Dotted lines are predictions assuming identical responses in the visually matching segments. **f.** Same as in **e.** for response profiles of V1 neurons (4,602 of 16,238 V1 neurons with variance explained ≥ 5%). **g.** Same as in **e.** and **f.** for response profiles of neurons across all 6 higher visual areas (4,381 of 18,142 neurons with variance explained ≥ 5%). **h.** Cumulative distribution of the spatial modulation index across visual areas (SMI) in even trials: SMI = 1 means single-peaked response; SMI = 0 means double-peaked response. *Black:* LGN, *Gray:* V1; *Purple:* mean cumulative distribution across higher visual areas (HVAs). Only neurons with responses within the visually-matching segments are included (LGN: 840/1,140, V1: 2,602/ 4,602, higher visual areas: 2,384/4,381) **i.** Violin plots showing the SMI distribution and median SMI (*white vertical line*) for each visual area (median ± m.a.d. LGN: 0.07 ± 0.11; V1: 0.44 ± 0.31; LM: 0.37 ± 0.25; AL: 0.32 ± 0.28; RL: 0.44 ± 0.31; A: 0.38 ± 0.34; AM: 0.27 ± 0.26; PM: 0.32 ± 0.32). Same color code for LGN, V1 and higher visual areas as in **h.**

Spatial modulation was broadly similar across higher visual areas, and not significantly larger than in V1 (Fig. 1g-i, Suppl. Fig. 1g, h). We measured activity in six visual areas that surround V1 (LM, AL, RL, A, AM, and PM), and found strong modulation by spatial position (Fig. 1g). Pooling across these areas, the median SMI across sessions was 0.40 ± 0.13 ($p = 10^{-10}$, right-tailed Wilcoxon signed rank test, Fig. 1h) and not significantly different from V1 (two-sample *t*-test: $p = 0.65$). Spatial modulation was present in each of the six areas (Suppl. Fig. 1g), and could not be explained by other factors such as running speed, reward events, pupil size and eye position (Suppl. Fig. 2).

We observed some differences in spatial modulation between cortical areas, but these differences were small and might be explained by biases in retinotopy[16–18] (Fig. 1i, Suppl. Fig. 1). In V1, neurons with central receptive fields showed stronger spatial modulation than neurons with peripheral receptive fields (Suppl. Fig 1, e, f.) A similar effect was seen across higher areas, with lower SMI in areas biased towards the periphery (AM, PM) than in areas biased towards the central visual field (LM, RL) (Fig. 1i, Suppl. Table 1a).

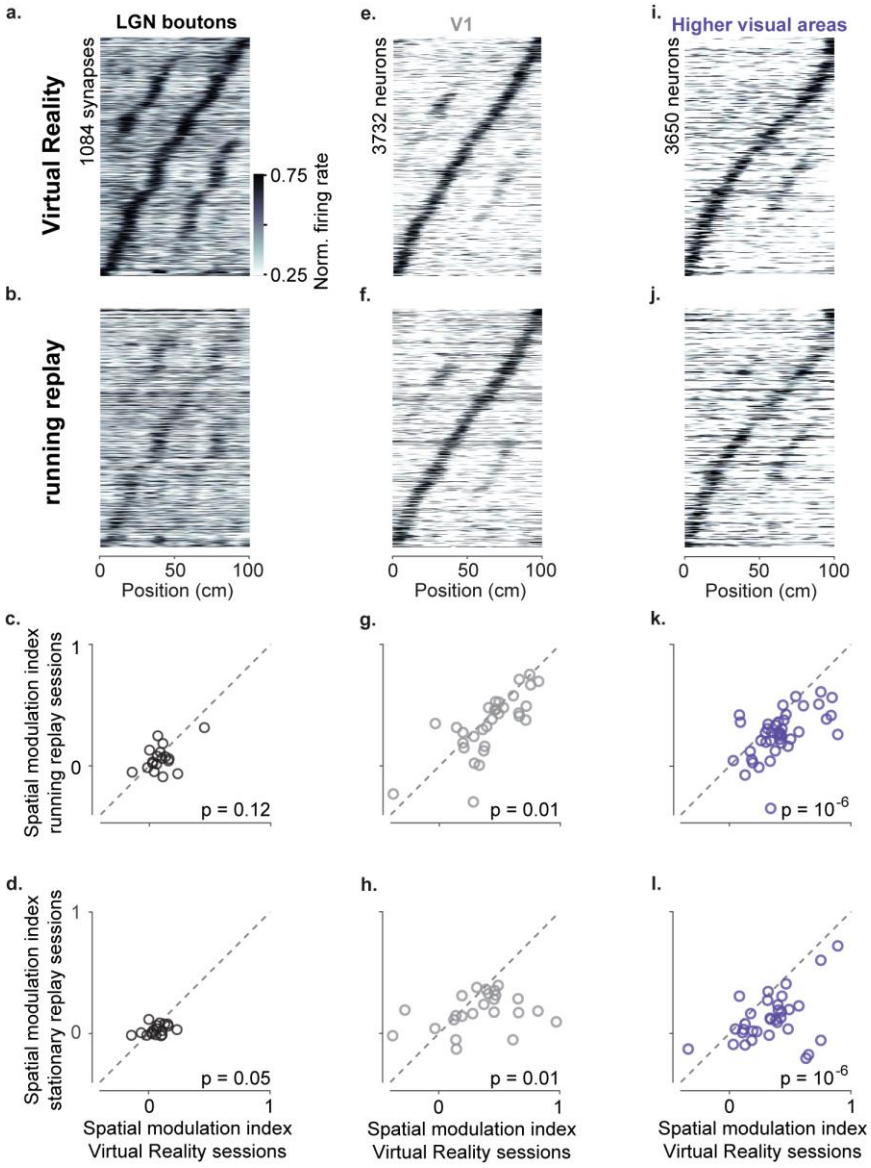


Fig. 2: Active navigation enhances modulation by spatial position in visual areas.
a. Response profiles of LGN boutons in Virtual Reality that also met the conditions for running replay (at least 10 running trials per recording session), estimated as in Figure 1e. **b.** Response profiles of LGN boutons during running replay, ordered by the position of their maximum response estimated from odd trials in Virtual Reality (same order and normalization as in **a**). **c.** Median spatial modulation index per recording session in Virtual Reality versus running replay for LGN (each circle corresponds to a single session; p values from paired-sample right-tailed *t*-test). **d.** Same as **c** for stationary replay. **e-h**. Same as in **a-d** for V1 neurons. **h-k.** Same as in **a-d** for neurons in higher visual areas.

We next asked if spatial modulation requires active navigation. During passive viewing, only a fraction of hippocampal place cells maintain their fields relative to active navigation[13–15]. If the spatial modulation seen in visual cortex mirrors that observed in hippocampus, it should also decrease during passive viewing. To test this prediction, after recordings in the normal virtual environment settings ('VR mode'), we played back the same video regardless of the mouse's movements ('replay mode').

Many cortical neurons, particularly in in higher visual areas, showed weaker modulation by spatial position during replay than in VR (Fig. 2). We focused at first on replay sessions when the mouse was running (at least 10 trials with running speed > 1 cm/s, 'running replay'). In this condition, LGN decreased baseline activity but did not change

spatial modulation: overall median SMI across sessions did not significantly differ from that in VR mode (Fig. 2a-c, n = 18 sessions, p = 0.12, paired-sample right-tailed *t*-test). By contrast, replay was associated with a small reduction of SMI in area V1 (Fig. 2e-g, n = 31 sessions, p = 0.015) and a large reduction in SMI in higher visual areas (Fig. 2i-k, n = 41 sessions, p = 10^{-6}). This decrease in spatial modulation during replay was not associated with firing rate differences, because mean firing-rates remained unchanged (Suppl. Fig. 3, Suppl. Table 1b). Moreover, the cortical decrease in SMI was even more pronounced when the animal was still ('stationary replay', Fig. 2d, Suppl. Fig. 4, V1: n = 24 sessions, p = 0.01; higher visual areas: n = 33 sessions, p = 10^{-6}).

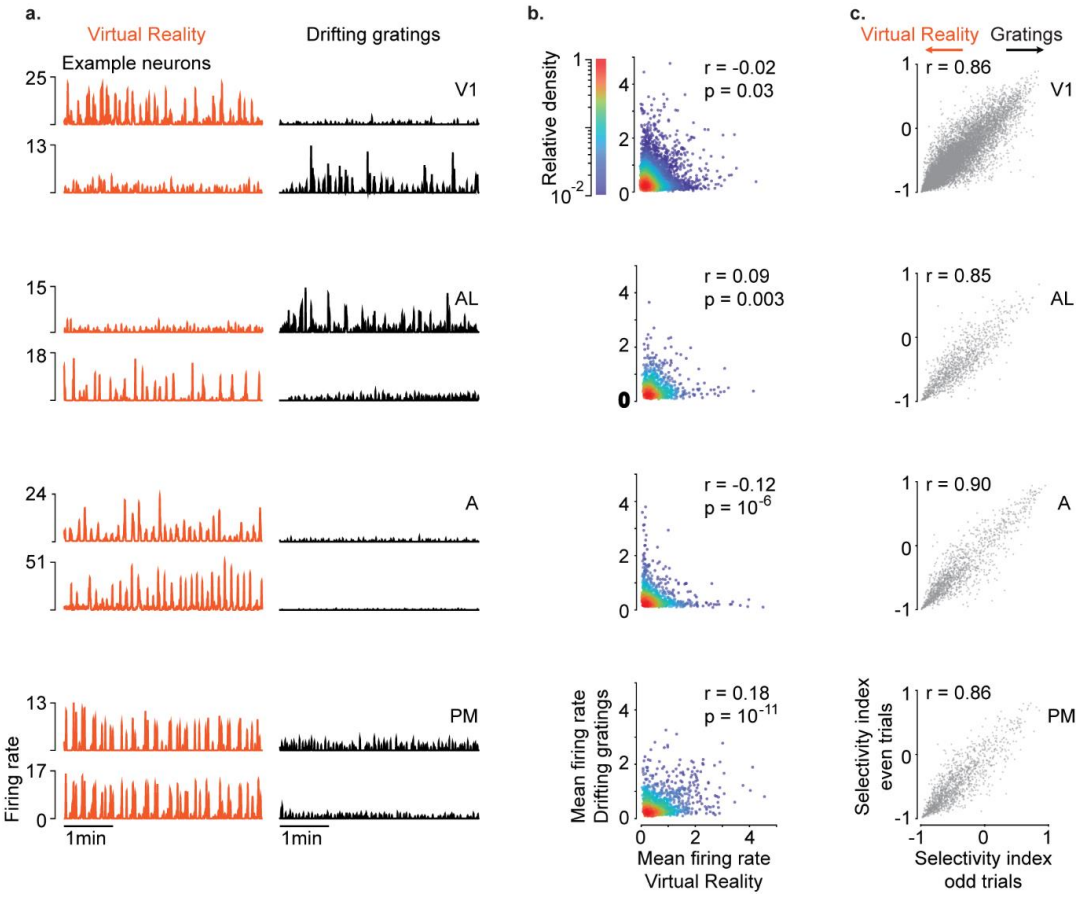


Fig. 3: Navigation and drifting gratings activate different neurons in visual areas
a. Example neuron traces in Virtual Reality and during presentation of vertical drifting gratings for four example areas (V1, AL, A and PM). **b.** Comparison of mean firing rate in Virtual Reality vs. responses to the drifting gratings for the same areas as in **a.** Each dot is a neuron, except in regions of high density, where color indicates density (color bar). (r and p: Pearson correlation coefficient with p-value) **c.** We measured the relative preference for gratings vs. Virtual Reality through a selectivity index. The index obtained from odd trials vs. even trials was highly consistent for the four areas as in **a**, **b.** (r: Pearson correlation coefficient; p<10^{-10} in all areas, Suppl. Table 1f)

Having observed many neurons across visual areas that are modulated by spatial position, we next asked how these neurons would respond to typical laboratory stimuli: drifting gratings (Fig. 3). In a subset of experiments, after recording activity in VR mode and in replay mode, we presented vertical drifting gratings centered on the average receptive field of the imaged neurons. We chose vertical gratings because they move horizontally, like most features that appear on the screens during navigation in VR. We presented the gratings at different spatial and temporal frequencies to accommodate the preferences of different visual areas for these parameters[19–24] (Suppl. Fig. 5).

Surprisingly, many neurons fired either in VR or in response to drifting gratings, but not in both (Fig. 3 a, b and Suppl. Fig. 6 a, b). Most visual areas contained neurons that fired robustly in one condition or the other (Fig. 3a). To compare responses in the two conditions, we first considered a neuron's mean firing rate. Mean firing rates in VR or in responses to gratings were independent in areas V1, RL, AM (Spearman correlation coefficient $\rho \sim 0$ and non-significant $p > 0.01$), and anticorrelated in areas LM and A ($\rho<0$, $p<10^{-5}$; Suppl. Table 1d). Similar results were seen regardless of whether the animal ran during presentation of the gratings (Suppl. Fig. 6a, b). We defined a selectivity index as the peak response to the gratings minus response at the preferred position in VR, divided by their sum. This measure was robust: it was highly correlated in even vs. odd trials (Fig 3c, Suppl. Table 1f).

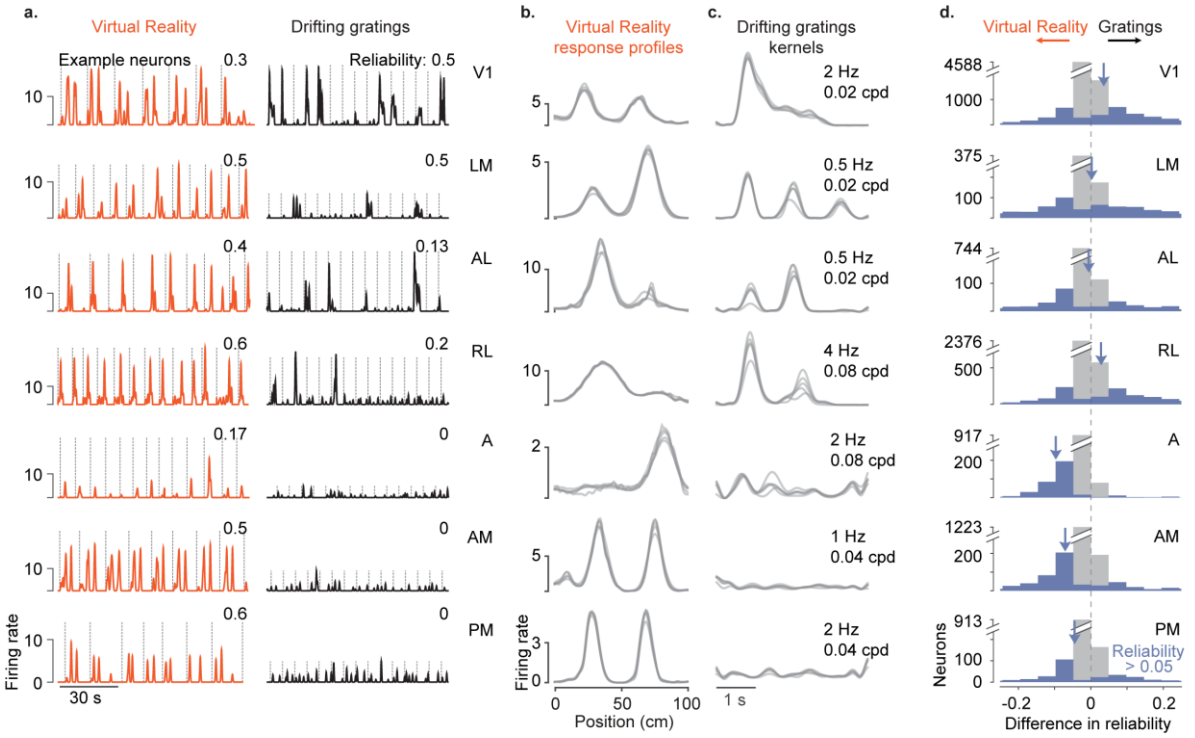


Figure 4: Active navigation enhanced response reliability in parietal and medial areas A, AM, PM.
a. Example neuron traces in Virtual Reality and drifting gratings across a series of trials for each cortical area (*dotted* line: trial onset). **b.** Response profiles in Virtual Reality for the example cells in **a.** Each trace corresponds to a single cross-validation fold (5-folds in total) **c.** Example linear kernels for the cells in **a.** 5-fold cross-validation was used to estimate one kernel per grating. Traces for V1, LM, AL and RL correspond to single-fold kernels for the preferred stimulus. **d.** Distributions of difference in response reliability to drifting gratings or in Virtual Reality estimated for each cortical area. *Gray*: all recorded responses; *blue*: responses with reliability higher than 5%. Arrow: mean difference in reliability for the distribution in *blue*.

In parietal and medial visual areas A, AM, and PM, responses during active navigation were more reliable than responses to gratings (Fig. 4 and Suppl. 6c, d). We calculated the reliability of neuronal responses based on variance explained (Fig 4, Suppl. Fig. 6e, f). For VR, we estimated variance explained from cross-validated predictions of responses based on spatial position (Fig 4b). For gratings, we estimated variance explained from cross-validated predictions of responses based on the visual stimuli presented (allowing one response per stimulus, Fig 4c). Parietal areas A, AM and medial area PM responded more reliably during active navigation than when viewing gratings (Fig 4d), regardless of running (Suppl. 6c, d; note the smaller sample for PM during running). Moreover, the Spearman correlation between reliability in VR and in

responses to gratings was less than 0.1 in these areas, and non-significant for areas A and PM (Suppl. table 1e). Areas A, AM and PM are therefore more strongly engaged by active navigation, consistent with the view that they sit higher along the visual processing hierarchy[25,26].

Together, these results show that the spatial modulation observed[2-4] in V1 is not inherited from LGN inputs, but rather arises in cortex and persists across higher visual areas. We found that LGN input to V1 is minimally influenced by navigational signals, and the little modulation that we did see might perhaps arise within cortex, from presynaptic modulation[28]. Instead, all cortical visual areas contained neurons that were influenced by navigational signals and that gave stronger responses during active behavior than during passive viewing. Notably, many neurons with reliable firing rates in VR failed to respond to drifting gratings, and vice versa. This result extends the finding that cortical neurons can respond very differently to different stimuli[29]. Most neurons in parietal areas A and AM responded reliably only during active behavior, consistent with a central role of parietal cortex during navigation[6,7,30,8–12]. Medial area PM also preferred active navigation, in agreement with anatomical[25] and functional[24,31] evidence suggesting that it may play a role in navigation. Our results illustrate how active navigation can shape responses along the visual pathway: it enhances the reliability of responses and the modulation by navigational signals in cortex, while exerting negligible effect on visual thalamus.

Acknowledgements

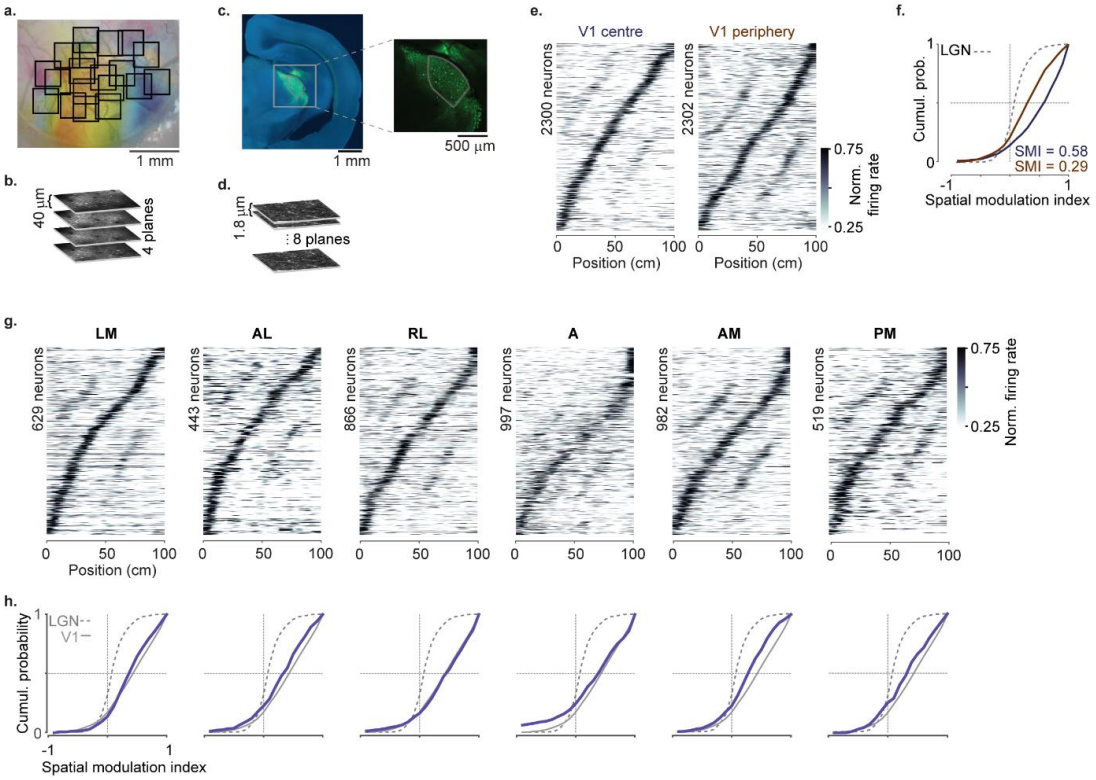
We thank Julien Fournier for helpful discussions, Michael Krumin for assistance with two-photon imaging, and Karolina Socha for advice on imaging LGN boutons. This work was supported by EPSRC (PhD award F500351/1351 to E.M.D.), by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (200501 to A.B.S.), and by the Wellcome Trust (grant 205093 to M.C. and K.D.H.). M.C. holds the GlaxoSmithKline / Fight for Sight Chair in Visual Neuroscience.

References

1. Ji, D. & Wilson, M. A. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* **10,** 100–107 (2007).
2. Haggerty, D. C. & Ji, D. Activities of visual cortical and hippocampal neurons co-fluctuate in freely moving rats during spatial behavior. *Elife* **4,** (2015).
3. Fiser, A. *et al.* Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* **19,** 1658–1664 (2016).
4. Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D. & Carandini, M. Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature* **562,** 124–127 (2018).
5. Hok, V., Jacob, P.-Y., Bordiga, P., Poucet, B. & Save, E. A spatial code in the dorsal lateral geniculate nucleus. doi:10.1101/473520
6. McNaughton, B. L. *et al.* Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4,** 27–39 (1994).
7. Nitz, D. A. Tracking route progression in the posterior parietal cortex. *Neuron* **49,** 747–756 (2006).
8. Save, E. & Poucet, B. Role of the parietal cortex in long-term representation of spatial information in the rat. *Neurobiol. Learn. Mem.* **91,** 172–178 (2009).
9. Whitlock, J. R., Pfuhl, G., Dagslott, N., Moser, M.-B. & Moser, E. I. Functional split between parietal and entorhinal cortices in the rat. *Neuron* **73,** 789–802 (2012).
10. Wilber, A. A., Clark, B. J., Forster, T. C., Tatsuno, M. & McNaughton, B. L. Interaction of Egocentric and World-Centered Reference Frames in the Rat Posterior Parietal Cortex. *J. Neurosci.* **34,** 5431–5446 (2014).
11. Krumin, M., Lee, J. J., Harris, K. D. & Carandini, M. Decision and navigation in mouse parietal cortex. *Elife* **7,** (2018).
12. Minderer, M., Brown, K. D. & Harvey, C. D. The Spatial Structure of Neural Encoding in Mouse Posterior Cortex during Navigation. *Neuron* **102,** 232–248.e11 (2019).
13. Chen, G., King, J. A., Burgess, N. & O'Keefe, J. How vision and movement combine in the hippocampal

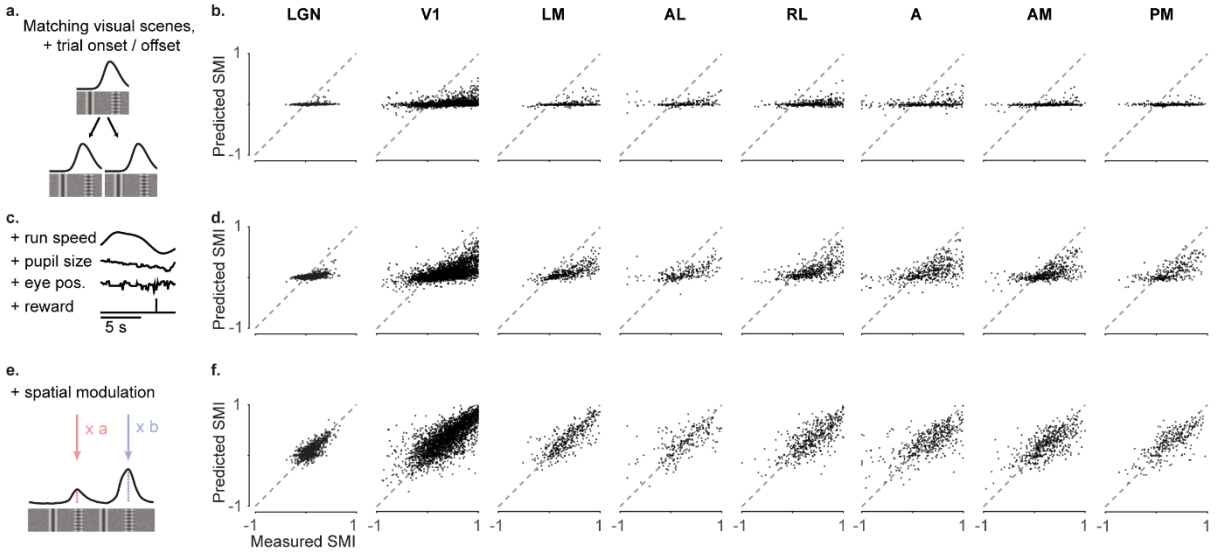
place code. *Proc Natl Acad Sci U S A* **110,** 378–383 (2013).

14. Song, E. Y., Kim, Y. B., Kim, Y. H. & Jung, M. W. Role of active movement in place-specific firing of hippocampal neurons. *Hippocampus* **15,** 8–17 (2005).
15. Terrazas, A. *et al.* Self-Motion and the Hippocampal Spatial Metric. *J. Neurosci.* **25,** 8085–8096 (2005).
16. Zhuang, J. *et al.* An extended retinotopic map of mouse cortex. *Elife* **6,** 1–29 (2017).
17. Garrett, M. E., Nauhaus, I., Marshel, J. H. & Callaway, E. M. Topography and Areal Organization of Mouse Visual Cortex. *J. Neurosci.* **34,** 12587–12600 (2014).
18. Wang, Q. & Burkhalter, A. Area map of mouse visual cortex. *J. Comp. Neurol.* **502,** 339–357 (2007).
19. 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).
20. Marshel, J. H., Garrett, M. E., Nauhaus, I. & Callaway, E. M. Functional specialization of seven mouse visual cortical areas. *Neuron* **72,** 1040–1054 (2011).
21. Murakami, T., Matsui, T. & Ohki, K. Functional Segregation and Development of Mouse Higher Visual Areas. *J. Neurosci.* **37,** 731–17 (2017).
22. Smith, I. T., Townsend, L. B., Huh, R., Zhu, H. & Smith, S. L. Stream-dependent development of higher visual cortical areas. *Nat. Neurosci.* **20,** 200–208 (2017).
23. Tohmi, M., Meguro, R., Tsukano, H., Hishida, R. & Shibuki, K. The extrageniculate visual pathway generates distinct response properties in the higher visual areas of mice. *Curr. Biol.* **24,** 587–597 (2014).
24. Roth, M. M., Helmchen, F. & Kampa, B. M. Distinct Functional Properties of Primary and Posteromedial Visual Area of Mouse Neocortex. *J. Neurosci.* **32,** 9716–9726 (2012).
25. Wang, Q., Sporns, O. & Burkhalter, A. Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J. Neurosci.* **32,** 4386–4399 (2012).
26. Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* (2019). doi:10.1038/s41586-019-1716-z
27. Haggerty, D. C. & Ji, D. Activities of visual cortical and hippocampal neurons co-fluctuate in freely moving rats during spatial behavior. *Elife* **4,** (2015).
28. Gil, Z., Connors, B. W. & Amitai, Y. Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* **19,** 679–686 (1997).
29. de Vries, S. E. J. *et al.* A large-scale, standardized physiological survey reveals higher order coding throughout the mouse visual cortex. *bioRxiv* 359513 (2018). doi:10.1101/359513
30. Nitz, D. A. Spaces within spaces: Rat parietal cortex neurons register position across three reference frames. *Nat. Neurosci.* **15,** 1365–1367 (2012).
31. Funamizu, A., Kuhn, B. & Doya, K. Neural substrate of dynamic Bayesian inference in the cerebral cortex. *Nat. Neurosci.* **19,** 1682–1689 (2016).
32. Shamash, P., Carandini, M., Harris, K. & Steinmetz, N. A tool for analyzing electrode tracks from slice histology. *bioRxiv* 447995 (2018). doi:10.1101/447995
33. Schröder, S. *et al.* Retinal outputs depend on behavioural state. *bioRxiv* 638049 (2019). doi:10.1101/638049
34. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* **65,** 472–479 (2010).
35. 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).
36. Madisen, L. *et al.* Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron* **85,** 942–958 (2015).
37. Wekselblatt, J. B., Flister, E. D., Piscopo, D. M. & Niell, C. M. Large-scale imaging of cortical dynamics during sensory perception and behavior. *J. Neurophysiol.* **115,** 2852–2866 (2016).
38. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: Flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).
39. Ratzlaff, E. H. & Grinvald, A. A tandem-lens epifluorescence macroscope: Hundred-fold brightness advantage for wide-field imaging. *J. Neurosci. Methods* **36,** 127–137 (1991).
40. Carandini, M. *et al.* Imaging the Awake Visual Cortex with a Genetically Encoded Voltage Indicator. *J. Neurosci.* **35,** 53–63 (2015).
41. Kalatsky, V. A. & Stryker, M. P. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. *Neuron* **38,** 529–45 (2003).
42. Yang, Z., Heeger, D. J. & Seidemann, E. Rapid and Precise Retinotopic Mapping of the Visual Cortex Obtained by Voltage-Sensitive Dye Imaging in the Behaving Monkey. *J. Neurophysiol.* **98,** 1002–1014 (2007).
43. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv* 61507 (2016). doi:10.1101/061507

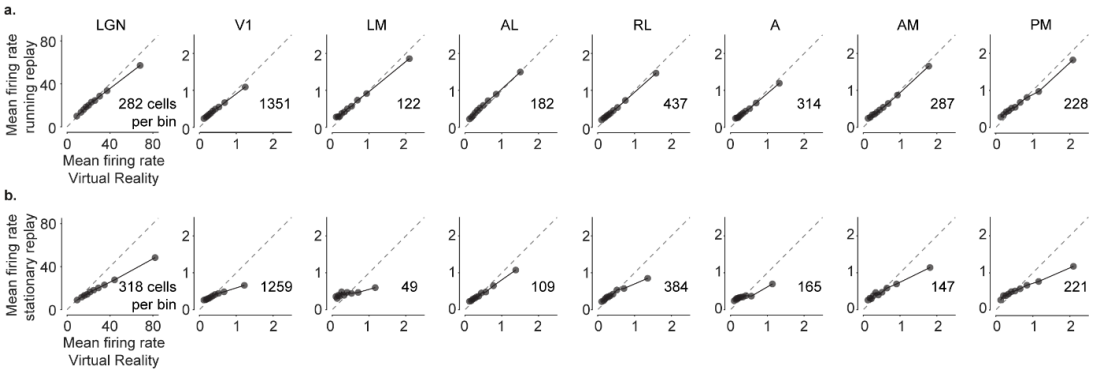


Suppl. Fig. 1: Imaging responses across the visual pathway during navigation

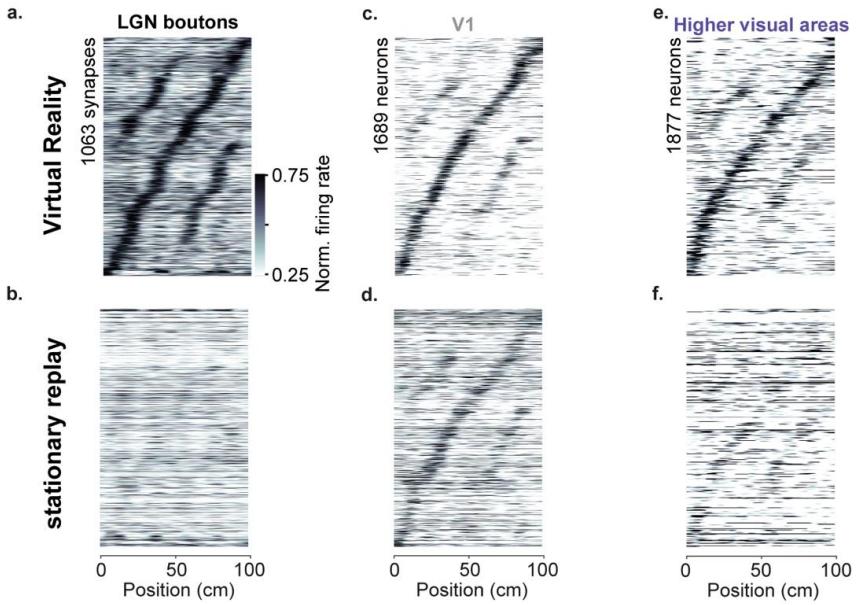
a. Two-photon imaging of cell bodies in visual cortex across multiple days: example retinotopic map with single-session fields of view (*black squares*). Combination of Fig 1b and c. **b.** Schematics of the volumetric imaging performed in visual cortex (total of 4 planes set 40 µm apart). **c.** Example image showing viral expression of GCaMP (*green*) in the visual thalamus of the right hemisphere (stained with Nissl in *blue*). *Inset:* zoomed-in image showing GCaMP expression in LGN and the surrounding nuclei. The LGN border (*gray contour*) was determined using SHARP-Track[32]. **d.** Schematics of the volumetric imaging of LGN boutons performed for motion-correction in the vertical dimension[33] (total of 8 planes set 1.8 µm apart). **e.** Response profile patterns obtained from even trials (ordered and normalized based on odd trials) for portions of V1 with average receptive fields in the center ('V1 center'; <40 deg azimuth angle; *left*) or in the periphery ('V1 periphery'; *right*). **f.** Cumulative distribution of the spatial modulation index in even trials for 'V1 center' (*purple*) or 'V1 periphery' (*orange*); *dotted line:* Distribution of LGN boutons (same as in **1h**). **g.** Response profile patterns obtained from even trials (ordered and normalized based on odd trials) for all probed higher visual areas. Only response profiles with variance explained ≥ 5% are included (LM: 629/1503 AL: 443/1774 RL: 866/5192 A: 997/4126 AM: 982/3278 PM: 519/2509) **h.** Cumulative distribution of the spatial modulation index in even trials for each higher visual area (*purple*). *Dotted line:* LGN, *Gray:* V1 (same as in **1h**).



Suppl. Fig. 2: Spatial modulation is not explained by other behavioral and visual factors.
We constructed three models to predict the activity of individual neurons from successively larger sets of predictor variables[4] **a, b.** Measured spatial modulation index for each visual area versus predictions of the simplest model ('purely visual model'). The 'purely visual model' considers only the repetitions of the visual scenes, trial onset and offset, and as expected, fails to predict the SMIs estimated from the data. Each point represents a neuron. **c, d.** Measured spatial modulation index for each cortical visual area versus predictions of the 'non-spatial' model. The 'non-spatial' model also includes the contribution by behavioral factors that can differ within and across trials: speed, reward times, pupil size, and eye position. Adding these factors improves predictions compared to the 'purely visual' model but fails to match the measured SMIs. Therefore, the joint contribution of all task-related and visual factors is not sufficient to explain the observed spatial modulation **e, f.** Measured spatial modulation index for each visual area versus predictions of the 'spatial' model. The spatial model allows the peaks in the visually matching segments to vary independently. It provides a much better match to the data.

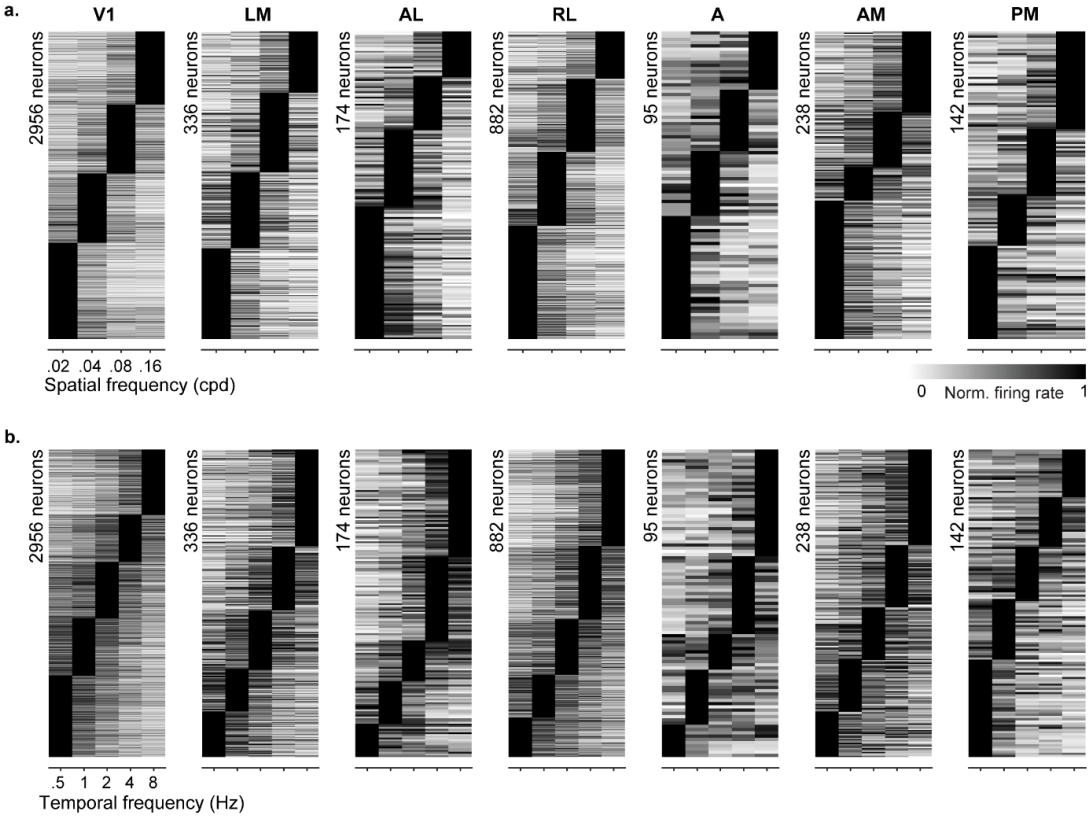


Suppl. Fig. 3: Comparison of firing rates in Virtual Reality and replay conditions
a. Comparison of the mean firing rate between Virtual Reality and running replay for all probed visual areas, split into 10 percentile bins. **b.** Same as in **a.** for stationary replay. Mean firing rates were similar between Virtual Reality and running replay, but decreased significantly during stationary replay in many areas, consistent with the well-established influence of running speed on visual responses[34,35]. (paired-sample right-tailed *t*-test: running replay: LGN: p = 0.003; p > 0.05 in all cortical areas; stationary replay: LGN: 10^{-65}, V1: 10^{-24}, LM: 0.93, AL: 0.15, RL: 0.06, A: 0.56, AM: 0.03, PM: 10^{-09}).



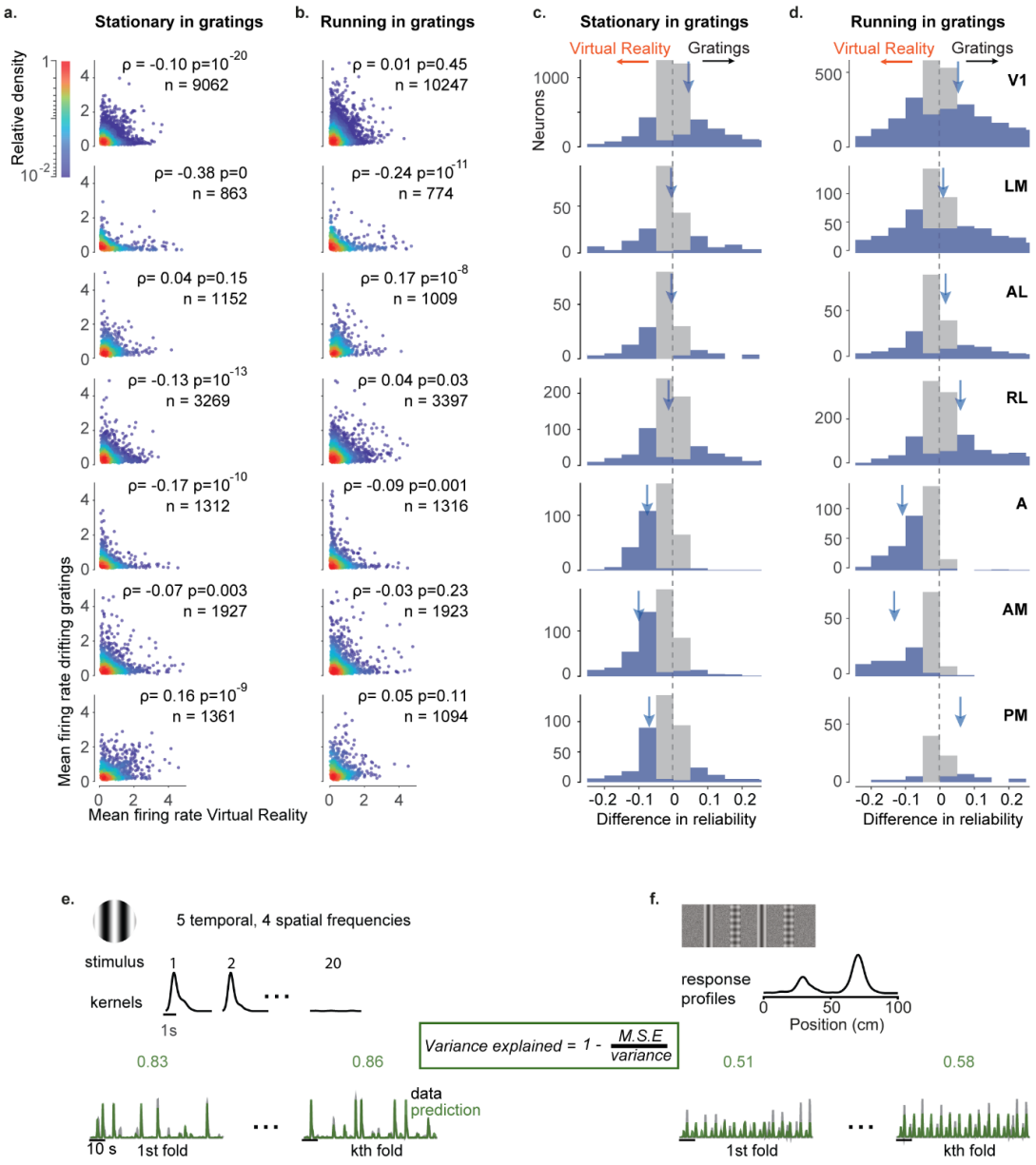
Suppl. Fig. 4: Spatial modulation decreases when animals are still during replay

a. Response profiles of LGN boutons in Virtual Reality that also met the conditions for stationary replay (at least 10 running trials per recording session), estimated as in Figure 1e. **b.** Response profiles of LGN boutons during stationary replay, ordered by the position of their maximum response estimated from odd trials in Virtual Reality (same order and normalization as in **a**). **c,d.** Same as in **a,b** for V1 neurons. **e,f.** Same as in **a,b** for neurons in higher visual areas.



Suppl. Fig. 5: Preferred spatial or temporal frequencies across visual areas.

a. Preferred spatial frequency across the population of each visual cortical area. Each row corresponds to a neuron's average response (2s time window after stimulus onset) to each spatial frequency at it's preferred temporal frequency. Responses are sorted based on each neuron's preferred spatial frequency. **b.** As in **a.** but for temporal frequencies at each neuron's preferred spatial frequency.



Suppl. Fig. 5: Navigation activates different neurons than gratings regardless of running

a. Density scatter plots of mean firing rate in Virtual Reality vs presentation of drifting gratings during periods when animals are stationary. Each panel corresponds to a cortical visual area (ρ, p: Spearman correlation coefficient with p value; n: total number of neurons) **b.** Same as in **a.** for periods during drifting gratings when animals are running. **c.** Histograms of difference in response reliability in Virtual Reality to grating-stationary . Each panel corresponds to a single cortical visual area. *Gray*: all recorded responses; *blue*: responses with reliability higher than 5%. Arrow: mean difference in reliability for the distribution in purple. **d.** Same as in **c.** for grating-running. **e, f.** To measure response reliability we used a measure of variance explained (*inset*). For drifting gratings we used ridge regression to obtain linear filters for each stimulus. For Virtual Reality, we used a local smoothing method. Predictions were five-fold cross-validated and response reliability was defined as the mean variance explained across folds.

Suppl. Table 1: Comparison of different measures across visual areas.

	LGN	V1	LM	AL	RL	A	AM	PM
a. Spatial modulation index (SMI)	0.07	0.44	0.37	0.32	0.44	0.38	0.27	0.32
b. Running replay: Decrease in firing rate, p-value	0.003	1	0.98	1	1	0.99	0.90	0.84
c. stationary replay: Decrease in firing rate, p-value	10^{-65}	10^{-24}	0.93	0.15	0.06	0.56	0.03	10^{-9}
d. Firing rate in Virtual Reality vs gratings: Spearman correlation coefficient ρ and p-value	-	-0.02 0.03	-0.31 10^{-20}	0.09 0.003	-0.02 0.15	-0.12 10^{-6}	-0.05 0.03	0.18 10^{-11}
e. Firing rate in Virtual Reality vs gratings: Spearman correlation coefficient ρ and p-value	-	0.19 10^{-88}	0.21 10^{-13}	0.10 10^{-4}	0.12 10^{-15}	0.04 0.16	0.09 10^{-5}	0.04 0.15
f. Selectivity index in odd vs. even trials: Spearman correlation coefficient ρ and p-value	-	0.86 $<10^{-10}$	0.91 $<10^{-10}$	0.85 $<10^{-10}$	0.85 $<10^{-10}$	0.90 $<10^{-10}$	0.87 $<10^{-10}$	0.86 $<10^{-10}$

Full Methods

All experimental procedures were conducted under personal and project licenses issued by the Home Office, in accordance with the UK Animals (Scientific Procedures) Act 1986.

For calcium imaging experiments in visual cortex, we used double or triple transgenic mice expressing GCaMP6 in excitatory neurons (3 females, 1 male, implanted at 4-8 weeks). The triple transgenics expressed GCaMP6 fast[36] (Emx1- Cre;Camk2a-tTA;Ai93, 3 mice). The double transgenic expressed GCaMP6 slow[37] (Camk2a-tTA;tetO-G6s, 1 mouse). For calcium imaging experiments of LGN boutons, we used 3 C57BL/6 mice (3 females, implanted at 6-9 weeks).

Surgical procedures

For two-photon calcium imaging of activity in visual cortex, 4–10–week–old mice were implanted with an 8 mm circular chamber and a 4 mm craniotomy was performed over the left or right visual cortex as previously described[4]. The craniotomy was performed by repeatedly rotating a biopsy punch and it was shielded with a double coverslip (4 mm inner diameter; 5 mm outer diameter).
For two-photon calcium imaging of activity of LGN boutons, 4–10–week–old mice were first implanted with an 8 mm head plate and a 4mm craniotomy was performed over the right hemisphere, as described above. We next injected 253 nL (2.3 nL pulses with an inter-pulse interval of 5 s, a total of 110 pulses) of the virus AAV9.CamkII.GCamp6f.WPRE.SV40 ($5.0x10^{12}$ GC/ml) into the visual thalamus. To target LGN in the right hemisphere, the pipette was directed at 2.6 mm below the brain surface, 2.3 mm posterior and 2.25mm lateral from bregma. To prevent backflow, the pipette was kept in place for 5 min after the end of the injection.

Perfusion and Histology

Mice were perfused with 4% PFA and the brain was extracted and fixed for 24 hours at 4^{o} C in 4% PFA, then transferred to 30% sucrose in PBS at 4^{o} C. The brain was mounted on a benchtop microtome and sectioned at 60 µm slice thickness. Free-floating sections were washed in PBS, mounted on glass adhesion slides, stained with DAPI (Vector Laboratories, H-1500) and cover with a glass-slip before being imaged.

Imaging of brain slices

Brain sections were initially imaged on a Zeiss AxioScan with a 4x/0.2-NA objective (Nikon CFI Plan Apochromat Lambda) by stitching 3x3 fields of view (total image size: 6mm x 6 mm). Images were obtained in two colors: green for GCaMP and blue for DAPI.

To obtain higher magnification images of GCaMP expression in LGN or of LGN boutons in V1 we performed confocal microscopy on a Zeiss LSM 880 with Airyscan with a 10x/0.3-NA objective (Zeiss EC Plan-Neofluar). We imaged GCaMP using an excitation wavelength of 488nm and DAPI using 405nm, and averaged across 16 frames. We imaged LGN using tiling with 15% overlap.

Two-photon imaging

Two-photon imaging was performed with a standard multiphoton imaging system (Bergamo II; Thorlabs Inc.) controlled by ScanImage4[38]. A 970 nm or 920 nm laser beam, emitted by a Ti:Sapphire Laser (Chameleon Vision, Coherent), was targeted onto L2/3 neurons or L4 LGN boutons through a 16x water-immersion objective (0.8 NA, Nikon). Fluorescence signal was transmitted by a dichroic beam splitter and amplified by photomultiplier tubes (GaAsP, Hamamatsu). The emission light path between the focal plane and the objective was shielded with a custom-made plastic cone, to prevent contamination from the monitors' light. Multiple-plane imaging was enabled by a piezo focusing device (P-725.4CA PIFOC, Physik Instrumente), and an electro-optical modulator (M350-80LA, Conoptics Inc.) which allowed adjustment of the laser power with depth.

For experiments monitoring activity in visual cortex, we imaged 4 planes set apart by 40 µm. Images of 512x512 pixels, corresponding to a field of view of 500x500 µm, were acquired at a frame rate of 30 Hz (7.5 Hz per plane). For experiments monitoring activity of LGN boutons, we imaged 7-10 planes set apart by 1.8 µm. Images of 256x256 pixels, corresponding to a field of view of 100x100 µm, were acquired at a frame rate of 58.8 Hz.

Widefield calcium imaging

To obtain retinotopic maps we used wide-field calcium imaging, as previously described[4]. Briefly, we used a standard epi-illumination imaging system[39,40] together with an SCMOS camera (pco.edge, PCO AG). A 14º-wide vertical window containing a vertical grating (spatial frequency 0.15 cycles/º), swept[41,42] the horizontal position of the window over 135º of azimuth angle, at a frequency of 2 Hz. To obtain maps for preferred azimuth we combined responses to the 2 stimuli moving in opposite direction[41].

Two-dimensional visual stimuli

We presented 20 vertical drifting sinusoidal gratings moving in the nasal to temporal direction at 4 spatial frequencies (0.02, 0.04, 0.08 and 0.16 cycles/deg) and 5 temporal frequencies (0.5, 1, 2, 4 and 8 Hz). Stimulus presentations lasted 4 s, were followed by a 3 s blank period, and were repeated 10 times. On each repeat, the 20 stimuli and a blank were presented in pseudorandom order. Stimulus contrast was 100% and stimulus diameter was 40 deg. Within each session, stimulus position was adjusted to match the center of the receptive fields obtained from the neuropil.

To obtain neuropil receptive fields, on each recording session we presented sparse uncorrelated noise for 5 min. The screen was divided into a grid of squares of 4 x 4 degrees size. Each square was turned on and off randomly at a 10 Hz rate. At each moment in time, 2% of the squares were on. To compute the neuropil receptive fields, the field of view was segmented into 5x5 patches (100 µm x 100 µm surface per patch). For each patch, we first averaged the raw fluorescence across the patch's pixels. We then computed the stimulus-triggered average of the averaged raw fluorescence trace. The response was further smoothed in space and its peak was defined as the patch's receptive field center.

Virtual Reality set-up

Animals were head-restrained in the center of three LCD monitors (IIyama ProLite E1980SD 19'') or three 10-inch LCD screens (LP097QX1-SPAV 9.7'', XiongYi Technology Co., Ltd.) placed

at 90deg angle to each other. The distance from each screen was 19cm for the LCD monitors, or 11cm for the LCD screen, so that visual scenes covered the visual field by 135 deg in azimuth and 42 deg in elevation.

The Virtual Reality environment was a corridor with two visually matching segments (Saleem 2018). Briefly, the corridor was 8cm wide and 100 cm long. A vertical grating or a plaid, 8 cm wide each, alternated in the sequence grating-plaid-grating-plaid at 20, 40, 60 and 80 cm from the start of the corridor.

In Virtual Reality mode, animals traversed the virtual environment by walking on a polystyrene wheel (15 cm wide, 18 cm diameter) which allowed movement along a single dimension (forwards or backwards). Running speed was measured online with a rotary encoder (2400 pulses/rotation, Kübler, Germany) and was used to control the update of visual scenes. Upon reaching the 100th cm of the corridor, animals were presented with a gray screen for an inter-trial period of 3 to 5 s, after which they were teleported back to the beginning of the corridor for the next trial. The duration of each trial depended on how long it took the animal to reach the end of the corridor, typically less than 8 s. Trials in which animals did not reach the end of the corridor within 30 s were timed-out and excluded from further analysis. A typical session consisted of more than 50 trials.

In the passive viewing mode, mice were presented with a previous closed-loop session, while still free to run on the wheel.

Behavior and training

Mice ran through the corridor with no specific task (n = 4 animals, 65 sessions recording cortical activity; n = 3 animals, 19 sessions recording activity of LGN boutons). Prior to recording sessions, mice were placed in the virtual environment, typically for 3 days and for up to one week, until they were able to run for at least 80% of the time within a single session. 2 out of 7 mice ran without rewards. 5 out of 7 mice were motivated to run with rewards, by receiving ~2.5 µl of water (4 mice) or of 10% sucrose (1 mouse) with the use of a solenoid valve (161T010; Neptune Research, USA). One animal received rewards at random positions along the corridor. The other 4 mice received rewards at the end of the corridor.

Pupil tracking

We tracked the eye of the animal using an infrared camera (DMK 21BU04.H, Imaging Source) and custom software, as previously described[4].

Pre-processing of imaging data

Image registration in x and y, cell detection and spike deconvolution were performed with Suite2p[43].

For the LGN boutons data, we additionally used the method described in Ref. [33] to align image frames in z-direction (along depth). By using a stack of closely spaced planes (1.8 µm inter-plane distance), we were able to detect small boutons across multiple planes, which could have otherwise moved outside a given plane due to brain movement in the z-direction. In brief, for each imaging stack, the algorithm estimates the optimal stack shift that maximizes the similarity of each plane to their corresponding target image (with target images across planes having been

aligned to each other in x and y direction). After, assigning the shifted planes to their corresponding target image, a moving average across 2 to 3 neighboring planes is applied, resulting in a smooth image, and consequently in smooth calcium traces from boutons sampled from multiple, closely spaced planes.
All ROIs (cell bodies or boutons) were detected from the aligned frames and were manually curated with the Suite2p GUI, as described in Ref [4].

Analysis of responses in Virtual Reality

To obtain response profiles as a function of position along the corridor, we first smoothed the deconvolved traces in time with a 250ms Gaussian window and considered only time points with running speeds greater than 1 cm/s. We then discretized the position of the animal in 1 cm bins (100 bins in total) and estimated the spike count and occupancy map for each neuron. Both maps were smoothed in space with a fixed Gaussian window of 5 cm. Finally, each neuron's response profile was defined as the ratio of the smoothed spike count map over the smoothed occupancy map. We assessed the reliability of the response profiles based on a measure of variance explained and selected those with variance explained higher than 5%. To cross-validate the response profile patterns in Virtual Reality, we divided each session's trials in two halves (odd vs even) and obtained a pair of response profiles for each neuron. Odd trials were used as the train set, to determine the position at which cells preferred to fire maximally. Odd trials were subsequently excluded from further analysis.

The same splitting into odd and even trials was used to estimate each cell's spatial modulation index (SMI). For each cell, the position of the peak response was measured from the response profile averaged across odd trials ('preferred position'). We then obtained the response, $R_{p.}$, at the preferred position and the visually-identical position 40 cm away ('non-preferred position': R_n), using the response profile averaged across even trials. Cells with maximal response close to the start or end of the corridor (0-15 cm or 85-100 cm) were excluded, because their preferred position fell outside the visually matching segments. SMI was defined as:

$$SMI = \frac{R_{p.} - R_n}{R_{p.} + R_{n.}}$$

Therefore, a response with two equal peaks would have $SMI = 0$, whereas a response with one peak would have $SMI = 1$.

To cross-validate the response profile patterns and to estimate SMIs in passive viewing, we used the same odd trials from Virtual Reality as a train set. Based on those we obtained response profile patterns and SMIs from all trials during passive viewing. To isolate periods when the animal was stationary during passive viewing, we considered only time-points when the speed of the animal was less than 5cm/s. Response profiles during stationary viewing were estimated only if the animal was stationary in at least 10 trials within a session. To isolate periods when the animal was running during passive viewing, we considered only time-points when the speed of the animal was higher than 1 cm/s. Response profiles in running during replay were estimated only if the animal was stationary in at least 10 trials within a session.

Response reliability was defined as the cross-validated fraction of variance in firing rate explained by the response profile. Response profiles were estimated based on 80% of the data (shuffled trials) and tested on the remaining 20%. (5-fold cross-validation). For each estimate, reliability was defined as:

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

where $y(t)$ is the actual, smoothed firing rate of the neuron at time t, $y'(t)$ is the predicted firing rate for the same time bin based on the neuron's response profile and μ is the mean firing rate of the training data. Only neurons with a reliability greater than 5% were considered for further analysis.

Analysis of responses to drifting gratings

The deconvolved single-cell traces obtained from 4 planes were linearly interpolated to match the sampling rate of the imaging session (30 Hz). Single-cell activity was then baseline-corrected by subtracting the mean activity and averaged across the time-window when the stimulus was on (4s). To select for responsive cells, we compared the mean activity across repeats during the stimulation period to the mean activity during blank trials. We selected cells whose maximum activity during at least one stimulus type was two standard deviations higher than the mean activity during blank periods. We defined as the cell's preferred spatial and temporal frequency the frequency at which the cell fired maximally.

Reliability of responses to drifting gratings was estimated based on predictions obtained from ridge regression with a fixed ridge regression coefficient, λ = 0.01. For each spatial and temporal frequency pair, we defined a sparse column vector in time, whose values were zero except from the time points when the corresponding stimulus appeared on the screen. This vector was then shifted backward by 1 s and forward by 4 s in steps of 0.033 s, resulting in a stimulus matrix with 151 columns. Stimulus matrices for each frequency pair and two blank conditions were pooled together, yielding the design matrix, with dimensions: (# of frames) × (151 shifts × 22 stimulus conditions, including blanks). We used this matrix to predict firing rates using five-fold cross-validation, and to estimate response reliability from expression (1).

Comparison of responses to Virtual Reality and to gratings

To compare responses across conditions (Virtual Reality and drifting gratings) we estimated a selectivity index as:

$$selectivity\ index = \frac{R_{Gratings} - R_{VR}}{R_{Gratings} + R_{VR}}$$

Where R_{VR} is the peak amplitude of the response profile in Virtual Reality and $R_{Gratings}$ is the mean response to the 20 gratings (4 spatial, 5 temporal frequencies).

To assess the robustness of this index, we computed it separately for odd and even trials.

General linear models

To assess the joint contribution of all visual and behavioral factors in Virtual Reality we fitted the V1 responses to three multilinear regression models similar to Ref. [4]. The models had the form: $\hat{y} = X\hat{\beta}$, where X is an T-by-M matrix with T time points and M predictors and $\hat{y}$ is the predicted calcium trace (T-by-1 array). Optimal coefficient estimates $\hat{\beta}$ (M-by-1 array) that minimize 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.

The simplest model, the *visual* model, relied only on 'trial onset' (first 10cm in the maze), 'trial offset' (last 10 cm in the maze) and the repetition of visual scenes within the visually matching segments (from 10 to 90 cm in the maze). The basis functions for all predictors were square functions with width of 2 cm and height equal to unity. To model the repetition of visual scenes, a predictor within the visually matching segments comprised of two square functions placed 40 cm apart. Thus, the visual model's design matrix had 30 predictors plus a constant: 5 predictors for trial onset, 5 predictors for trial offset and 20 predictors within the visually matching segments.

The second model, the *non-spatial* model was used to assess the influence of all the behavioral factors we measured: running speed, reward events, pupil size and the horizontal and vertical pupil position. These factors were added as predictors to the design matrix of the *visual* model, as follows: running speed was shifted backward and forward in time twice, in 500ms steps, thus contributing 5 continuous predictors; pupil size and horizontal and vertical pupil position contributed 1 continuous predictor each; each reward event contributed one binary predictor at the time of the reward. The continuous predictors of running speed and pupil size were normalized between 0 and 1, whereas pupil position was normalized between -1 and 1 to account for movements in opposite directions.

The third model, the *spatial* model, allowed for an independent scaling of the two visually-matching segments in the maze. For each predictor within the visually-matching segments, the two square functions were allowed to vary their height independently. The height of the two square functions was parameterized by a parameter α, such that the two functions had unit norm. An α = 0.5 corresponded to a purely visual representation with SMI close to 0, while $\alpha = 1$ or $\alpha = 0$ would correspond to a response only in the first or second segment, and an SMI close to 1.

To choose the best model, 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, we performed multiple ridge regression fits, searching for the optimal value of α using a step size of 0.1, for each λ.

The predictions obtained in the time domain were subsequently processed similarly as the original deconvolved traces, to obtain predicted response profiles and SMIs.