The inevitability and superfluousness of cell types in spatial cognition

Xiaoliang Luo (xiao.luo.17@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, London, WC1H 0AP, UK

Robert M. Mok* (rob.mok@rhul.ac.uk)
Department of Psychology, Royal Holloway, University of London,
Royal Holloway, University of London, Egham, TW20 0EX, UK

Bradley C. Love* (b.love@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, London, WC1H 0AP, UK,
The Alan Turing Institute, 96 Euston Rd, London, NW1 2DB, UK

**co-senior author*

Abstract

Discoveries of functional cell types, exemplified by the cataloging of spatial cells in the hippocampal formation, are heralded as scientific breakthroughs. We question whether the identification of cell types based on human intuitions has scientific merit and suggest that "spatial cells" may arise in non-spatial computations of sufficient complexity. We show that deep neural networks (DNNs) for object recognition, which lack spatial grounding, contain numerous units resembling place, border, and head-direction cells. Strikingly, even untrained DNNs with randomized weights contained such units and support decoding of spatial information. Moreover, when these "spatial" units are excluded, spatial information can be decoded from the remaining DNN units, which highlights the superfluousness of cell types to spatial cognition. Now that large-scale simulations are feasible, the complexity of the brain should be respected and intuitive notions of cell type, which can be misleading and arise in any complex network, should be relegated to history.

Keywords: spatial cells; hippocampus; deep neural networks

Introduction

Spatial cells in the hippocampal formation are traditionally regarded to form a cognitive map that facilitates our spatial abilities. However, criteria for classifying these cells, including place, head-direction, and border cells (Moser, Kropff, & Moser, 2008) were subjectively determined by firing patterns that piqued neuroscientists' interest, and most cells in empirical data do not perfectly match idealized "cell types". Computational models have demonstrated that spatial firing patterns can emerge from factors unrelated to space, such as a constraint for sparseness (Franzius, Sprekeler, & Wiskott, 2007). Here, we question the presumed privileged role of spatial cells and propose that they may be inevitable by-products of general computational mechanisms rather than the cornerstone of spatial cognition. To test this, we explored the role of spatial cell-like representations in spatial cognition in deep neural networks (DNNs) optimized for object recognition as an example of general information-processing systems without spatial grounding. We simulate a free-foraging agent in a virtual environment where DNNs process first-person visual scenes, and analyzed the network units' activity and assessed their spatial knowledge. Specifically, we trained linear regression models to decode spatial variables relevant to navigation, including location, heading direction, and distance to border, across DNN layers. We found that spatial variables can be decoded from both trained and untrained DNNs (random weights) at multiple levels of processing. Furthermore, many DNN units fit the criteria for place, head-direction, and border cells. Notably, lesioning these units had minimal impact on decoding performance, suggesting individual spatial-cell types do not play a privileged role, whereby spatial knowledge is distributed across the population. Our work shows how general computational systems like the brain can appear to have domain-specific representations (spatial) even when they are general, and our subjective top-down perceptions of importance (cell types) must be rigorously tested and potentially revised to make progress to understanding the driving neural mechanisms for cognition.

Method

A three-dimensional virtual space is created to resemble a realistic laboratory environment with a variety of visual features. An agent moves randomly in a two-dimensional plane within the three-dimensional laboratory, like how an animal explores an enclosure, and processes first-person views of the environment (Fig. 1A). We define spatial knowledge of the agent with four values. The agent's location is denoted by the Cartesian coordinates t_x, t_y. The agent's heading direction is denoted by the angle t_r. The distance between the agent and the nearest wall is t_b (Fig. 1B). Individual views are processed by DNNs non-sequentially. We train linear regression models on various layers in these networks to assess spatial knowledge over levels of processing (Fig. 1C).

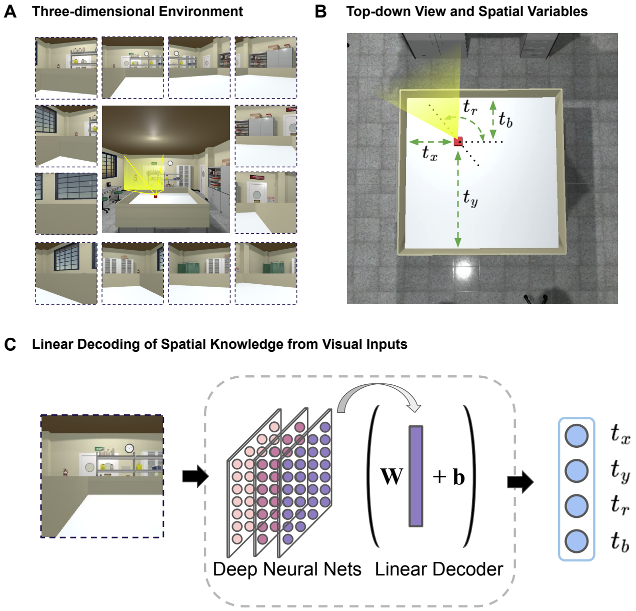
A Three-dimensional Environment
B Top-down View and Spatial Variables
t_r
t_b
t_x
t_y
C Linear Decoding of Spatial Knowledge from Visual Inputs
W
+ b
Deep Neural Nets
Linear Decoder
t_x
t_y
t_r
t_b

Figure 1: **(A-C)** Assessing spatial knowledge in non-spatial perception systems using a linear decoding approach in a virtual environment.

Results

To demonstrate how spatial knowledge can arise from complex computational systems irrespective of their architectural variations or training states, we conducted an investigation involving deep convolutional neural networks including (DCNN) including VGG-16 (Simonyan & Zisserman, 2015) and ResNet-50 (He, Zhang, Ren, & Sun, 2016) and Vision Transformers (ViT; (Dosovitskiy et al., 2020)). These models were evaluated both in their pre-trained form, trained on real images, and in an untrained state with randomly initialized parameters. For clarity, we present results on the penultimate layer representation and uniformly sampled 30% of all locations and their views for training the decoder and tested on the left out locations and views. Our results, as depicted in Fig. 2A, reveal that all models exhibit lower errors than the established baselines (decoding randomly or decoding the center). We also observed the same patterns of results across different model layers. Notably, this phenomenon persisted even in the case of untrained networks, especially in an untrained ViT—comprising nothing more than a hierarchy of fully-connected layers (i.e., self-attention) and non-linear operations.

How is it possible that non-spatial perception models contain such substantial spatial information about the external environment? The common view in the field assumes that cells exhibiting spatial firing profiles play a pivotal role in shaping spatial cognition, as they intuitively seem useful for spatial tasks and navigation. Are spatial cells responsible for the effective decoding of spatial information in our perception model?

To determine if the model's spatial knowledge is primarily supported by spatially-tuned units like those found in the brain, we classified every hidden unit in each model based on criteria used to identify spatial cells in neuroscience for place cells (Tanni, De Cothi, & Barry, 2022), head-direction cells (Banino et al., 2018), and border cells (Banino et al., 2018), respectively. We found many units in the non-spatial models that satisfied the criteria of place, head-direction, and border cells and show mixed selectivity. We selected example units from one model (VGG-16) and plotted their spatial activation patterns in the two-dimensional virtual space (irrespective of direction), and their direction selectivity in polar plots (Fig., 2B-E).

Further, we test whether spatial cells have a special role in spatial cognition by performing a systematic lesion analysis. First, we scored and ranked each unit classified as place, head-direction, and border units. We then re-trained linear decoders without the top n units of a specific cell type and evaluated the model's spatial knowledge. We repeated this procedure with a progressively higher lesion ratio. We show that lesioning spatial units in the models that scored highest on each of the corresponding criteria (place field activity, number of fields, directional tuning, and border tuning), had minimal effect on decoding performance even with a large proportion of the highest ranked spatial units being lesioned (Fig. 2F, top). We also randomly lesioned the same number of spatial units of each criterion (Fig. 2F, bottom) and observed a similar pattern of results across four lesion scenarios, meaning that the impact of lesioning highly-tuned spatial units was no less detrimental to spatial cognition than lesioning a random selection of units.

Conclusion

"Spatial cells" are traditionally viewed to play a pivotal role in spatial cognition. Here, we demonstrate that even deep neural networks of object recognition possess unit representations that satisfy the criteria for place, head-direction and border cells, and our lesion analysis showed these "spatial cells" do not play a special role in spatial cognition. In contrast to the traditional view of a dedicated spatial neural system, we propose that spatial cell-like representations inevitably arise in non-spatial, general information processing systems like the brain. Our findings indicate that the widely utilized top-down approach in neuroscience for identifying interpretable cells warrants re-examination, as it can often produce results that simply align with preconceived expectations without advancing the field.

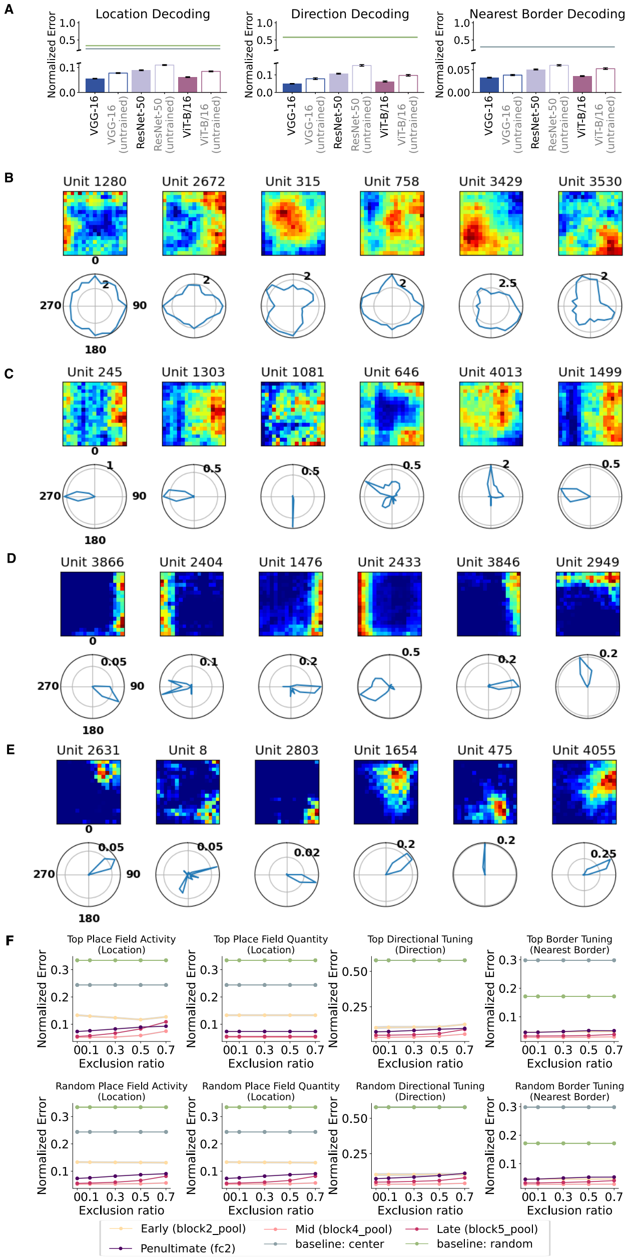

Figure 2: **(A)** Spatial knowledge can be decoded across trained and untrained deep neural networks irrespective of architectures; **(B)** Place cells; **(C)** Head-direction cells; **(D)** Border cells; **(E)** Place-direction cells; **(F)** Decoding of spatial knowledge is robust to spatial cell lesions.

Acknowledgments

This work was supported by ESRC (ES/W007347/1), Wellcome Trust (WT106931MA), and a Royal Society Wolfson Fellowship (18302) to B.C.L., and the Medical Research Council UK (MC_UU_00030/7) and a Leverhulme Trust Early Career Fellowship (Leverhulme Trust, Isaac Newton Trust: SUAI/053 G100773, SUAI/056 G105620, ECF-2019-110) to R.M.M.

References

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., ... Kumaran, D. (2018, May). Vector-based navigation using grid-like representations in artificial agents. *Nature*, *557*(7705), 429–433. Retrieved 2023-01-05, from `http://www.nature.com/articles/s41586-018-0102-6` doi: 10.1038/s41586-018-0102-6

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2020, 10). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR*, *abs/2010.1*. Retrieved from `https://arxiv.org/abs/2010.11929v2` doi: 10.48550/arxiv.2010.11929

Franzius, M., Sprekeler, H., & Wiskott, L. (2007, August). Slowness and Sparseness Lead to Place, Head-Direction, and Spatial-View Cells. *PLoS Computational Biology*, *3*(8), e166. Retrieved 2023-01-04, from `https://dx.plos.org/10.1371/journal.pcbi.0030166` doi: 10.1371/journal.pcbi.0030166

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee computer society conference on computer vision and pattern recognition* (Vol. 2016-Decem, pp. 770–778). IEEE Computer Society. doi: 10.1109/CVPR.2016.90

Moser, E. I., Kropff, E., & Moser, M.-B. (2008, July). Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annual Review of Neuroscience*, *31*(1), 69–89. Retrieved 2023-08-31, from `https://www.annualreviews.org/doi/10.1146/annurev.neuro.31.061307.090723` doi: 10.1146/annurev.neuro.31.061307.090723

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In Yoshua Bengio and Yann LeCun (Ed.), *International conference on learning representations.* Retrieved from `http://arxiv.org/abs/1409.1556`

Tanni, S., De Cothi, W., & Barry, C. (2022, August). State transitions in the statistically stable place cell population correspond to rate of perceptual change. *Current Biology*, *32*(16), 3505–3514.e7. Retrieved 2023-07-10, from `https://linkinghub.elsevier.com/retrieve/pii/S0960982222010089` doi: 10.1016/j.cub.2022.06.046