COMMENTARY

Why the neural ingredients for a language of thought are not like spatial cells (commentary on Kazanina & Poeppel, 2023)

Sander van Bree[1,2,3]

[1]Centre for Cognitive Neuroimaging, School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK

[2]Department of Medicine, Justus Liebig University, Giessen, Germany

[3]Vision and Computational Cognition Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Correspondence
Sander van Bree, Centre for Cognitive Neuroimaging, School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK.
Email: sandervanbree@gmail.com

Edited by: John Foxe

KEYWORDS: cognitive neuroscience, compositionality, Language of Thought, place cells, spatial navigation

1 | INTRODUCTION

The Language of Thought (LoT) refers to a body of hypotheses about how the mind works (Fodor, 1975). The central idea common across weak and strong accounts (Chalmers, 2023) is that thought partially or wholly consists in combining primitive representations stored in a repertoire, where the meaning of a complex thought is determined by the order of inserted constituents. With this architecture, a mind can produce arbitrarily many thoughts by drawing from its repertoire, with such thoughts carrying an internal structure that can be used to parse novel combinations and generate deductive inferences. In a recent article, Kazanina and Poeppel (2023) posit that spatial cell types in and around the hippocampal formation offer the right kind of neural ingredients for LoT, suggesting a linking hypothesis from mind to brain. The authors submit that place cells, border cells, object cells, and other cell types offer representations that flexibly combine to generate complex expressions about space, and this is taken to show that the key ingredients for LoT are on stock. For example, border cells might implement the function Border(X), which returns TRUE if there is a border of some sort at location X. Then, in keeping with LoT, this might be composed with other predicates, such as Object(X) and Place(X), respectively realized by object and place cells. As a result, animals construct full-fledged conceptions about the spatial layout of an environment, with such complex ideas displaying LoT properties like abstraction, role-filler independence, and vectorial stability. Role-filler independence is particularly central to this critique, and it refers to the idea that predicates or *roles* such as Border(X) have a separate representational format from the *fillers* or contents that enter into them (e.g., {X,Y} coordinates in 2D Euclidean space). In this piece, I argue against the ambitious and constructive suggestion that spatial cell types are the neural ingredient for LoT on the basis of four arguments.

2 | REMAPPING

As outlined in the target article, spatial cells show all sorts of remapping behaviour, which means they display systematic alteration in spiking across cognitive scenarios, spatial environments, or moments—in a word,

contexts. The phenomenon of remapping is relevant for evaluating the idea that spatial cell types underpin LoT because LoT posits the existence of a stable repertoire of representations that are available for general reuse, and its neural ingredients must accommodate that. In this section, I demonstrate that some forms of remapping streamline with the linking hypothesis in question, while others are problematic.

First, sometimes spatial cell types remap without changing their receptive field, such as when place cells respond less vigorously to the same coordinates in a new environment (Leutgeb et al., 2005). Second, sometimes cells remap and change their receptive field to accommodate an environmental change. For example, a border cell might curve its receptive field after an animal relocates from a square to a circular enclosure (Solstad et al., 2008). Third, sometimes spatial cells spontaneously remap and acquire a new and statistically independent receptive field (O'Keefe & Conway, 1978). Finally, sometimes spatial cell types abandon their field and stop firing for locations altogether (Knierim, 2002). The latter two forms are instances of global remapping (Kazanina & Poeppel, 2023; Latuske et al., 2018).

Let us square each of these examples with the notion that spatial cells offer ingredients for an LoT. First, when a spatial cell changes only its firing rate, one might say the *role* (e.g., Place) and the *filler* (X, coordinates) persist across contexts—no representations changed, only the vigour of the response. Second, when a spatial cell adapts its field to include for example a longer border in a second context, or a curved one, we might say that the *role* Border(X) is maintained and that the *filler X* that inserts coordinates gets modified slightly to accommodate wall extension or bending. Third, when a place cell acquires an orthogonal location, we might conclude it still carries its predicate or *role* (e.g., Place), but it has become bound to a new filler (e.g., a set of coordinates in the North-East). Finally, when a spatial firing field extinguishes altogether, neither *roles* nor *fillers* are maintained.

The first and second form of remapping are unproblematic, and as argued by Kazanina and Poeppel (2023), the third form of remapping is also compatible with the linking hypothesis because it points to the existence of variation-tolerating coding formats suited for an abstract and open-ended concept like border or object. That much makes sense, but the last form of remapping—receptive field extinction—poses a deep problem for the idea that spatial cell types offer the neural ingredients for LoT. To see why, consider the following argument:

1. [*predicate*] cell α encodes *predicate*(x) in context 1, but does not do so in context 2.

2. A neural LoT ingredient for *predicate*(x) can be used to build LoT expressions with *predicate*(x) across contexts.
3. Therefore, [*predicate*] cell α is not a neural LoT ingredient for *predicate*(x).

Here, the word "*predicate*" can be substituted with any truth-evaluable function assumed to be coded by cell types. This includes place cells, border cells, object cells and the spatial concepts these encode, but it also generalizes to non-spatial contexts such as concept cells and the iconic or lexical information they represent. Insofar these cells remap to lose their predicate, they cannot be "ontologically sufficient" for the representations and operations in LoT because LoT posits concepts that combine into complex constructions across multifaceted situations, with field-abandoning cells definitionally unequipped to realize that.

3 | WHAT INSTANTIATES THE FILLERS?

A second criticism is that the linking hypothesis in question addresses only predicates or *roles*, not the deeply symbolic *fillers* they parse. Namely, in the target piece, border cells, landmark cells, place cells, head-direction cells, and object cells are taken to offer a representational template by which a specific location (or directional angle) can be truth-evaluated. What is missing however is an account of how specific spatial locations themselves—*fillers*—are neurally realized, and how they get inserted into a predicate. Where is the representation for *North-East*, or more precisely a number-like {*horizontal, vertical*} coordinate? And what are the neural ingredients for role-filler binding? Until the X in Place(X) as well as its insertion into various roles is given a how-possibly account, no testable linking hypothesis synthesizes.

The problem of fillers is uniquely challenging because unlike the abstract, general, and pluralistic nature of spatial predicates that may well be within the scope of scruffy synaptic codes, the spatial locations that enter into such predicates are *variables* that invite neat symbolic coding formats (Gallistel, 1989). While Kazanina and Poeppel (2023) implicitly acknowledge the need for neural accounts towards such formats and very briefly touch on some, the cells they posit as LoT ingredients figure as realizers of *truth-evaluable claims about* symbolic fillers, not as ingredients for *the fillers themselves*. This is a problem because the neural ingredients for LoT must account for both roles and fillers, especially the symbolic representational demands of the latter.

—WILEY—

4 | LACK OF COMPOSITIONALITY

A third criticism of the linking hypothesis is that it is unclear where compositionality—a bedrock of LoT—fits into the story. Take the classical example of "John loves Mary." This expression is made up of concepts, and their ordering matters a great deal. If the constituents are shuffled around such that "Mary loves John," a new romance is told. In contrast, spatial concepts such as Place(X), Object(X), and Border(X) appear to configure into complex wholes via non-compositional principles. Concretely, it is unclear why the expression {Border (West), Object (South), Place (East)} should establish a different spatial environment than any other permutation, such as {Place (East), Object (South), Border (West)}.

Crucially, this point is not about how multiple spatial concepts combine, such as when Place and Object concepts are added into the same expression using conjunctively coding cells. Nor is it about the feasibility of logical connections, such as when cells activate when certain logical criteria are met (e.g., Border (West) AND Object (South)), which as discussed in the target article may be supported by theta-modulated cell types.

Ways to encode multiple representations and logically condition them are useful for a neural account of LoT—certainly such properties cohere better with LoT than purely associationist formulations. But the theory is about much more than that. LoT posits a processing architecture in which complex expressions are formed and determined by the meaning and relative ordering of predicates and symbols entered into it. Yet in the target model, constructions with spatial cells derive their meaning through the summing of spatial features, some conditioned upon each other, some like that. What is missing is the embedding of such features in a syntactical mould that does half the work for meaning derivation. In short, unlike words in Fodor's mental language, spatial concepts in the current formulation configure non-compositionally into a cognitive map, and as such the account falls short of meeting a necessary desideratum.

There are three possibilities. First, I am wrong and spatial cell types do implement LoT via a compositional architecture hitherto undescribed. Second, spatial cell types do not implement LoT but there is some other neural realization for a LoT of spatial cognition. Third, LoT simply does not apply to the function of spatial cognition (or even cognition in general), meaning no possible neural candidate can ultimately be made to work.

5 | ARE POPULATIONS A COUNTER TO RECEPTIVE FIELD LOSS?

One might object to the first argument of receptive field extinction by maintaining that the level of neural implementation for spatial predicates is not that of individual cells, but large assemblies of spatial cells. Under this lens, it is true that individual spatial cells can lose their identity, but this does not change the fact that spatial cell types considered holistically are the neural ingredients for spatial LoT predicates.

This defense requires further substantiation—at first glance, it appears to kick the can down the road. Populations of spatial cells drift in makeup; some cells lose their identity across contexts (Muller & Kubie, 1987), some change their predicate (Jeffery, 1998), and some non-spatial cells acquire a spatial receptive field (Roth et al., 2012). Such dynamism and representational turnover points to the existence of an unknown allocator mechanism which flexibly allocates roles to neurons, and which binds architecturally independent fillers stored elsewhere to them. In my view, *this* is where the neural ingredients of LoT reside.

Then how might spatial cell allocation occur? Kazanina and Poeppel (2023) drive at the heart of this when they briefly touch on two proposals. One explanation is that noisy inputs cascade through nonlinear synaptic weight matrices, possibly with grid cells exerting a scaffolding function to establish spatial fields from incoming activity (Fyhn et al., 2007). Alternatively, a hypothesis more conducive to LoT's symbolic form is that thermodynamically stable units store primitive contents, after which an unknown allocator mechanism allocates online representations in a principled manner (Gallistel & King, 2009). Here, RNA-based computational architectures are one possible candidate (Akhlaghpour, 2022; Gershman, 2023).

6 | CONCLUSION

Regardless of how spatial cells acquire their identity, the purpose of this contribution is to show that spatial cells themselves are not plausible neural ingredients for LoT. First, the receptive fields of spatial cells sometimes disappear, while the theory at hand demands a reusable repertoire of concepts. Second, spatial cells do not explain the symbolic fillers that enter into expressions—leaving open the trickiest part of the story. Third, LoT ingredients must per definition enable a compositional semantics, but the target model builds mental expressions via

non-compositionally combinable features. Fourth, spatial cell *populations* do not offer an ontologically sufficient system for LoT because representational turnover points to the existence of further necessary mechanisms. With that said, while spatial cells do not offer an inventory, lexicon, or the ingredients for LoT, they might reflect one way in which these things are brought into action. Joining the metaphor, if spatial cells figure in the implementation of LoT at all, they do so as cooked neural dishes, not ingredients.

Distilling these arguments, three prescriptions emerge. First, we need to distinguish concepts like *allocator* and *allocated*, and *ingredients* and *dishes* in our pursuits. Second, it may pay off to collectively hedge our bets (Kitcher, 1993) by exploring novel or under-resourced hypotheses for symbol storage and role-filler binding that could accommodate LoT desiderata. Third, any linking hypothesis must meet a theory's mental level tenets. In the case of LoT, this includes the extended reuse of primitive symbols to construct thoughts over time, and a compositional architecture that unlocks computational properties absent in associationist ones.

CONFLICT OF INTEREST STATEMENT
The author declares no competing interests.

PEER REVIEW
The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16329.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no datasets were generated or analysed.

KEYWORDS
cognitive neuroscience, compositionality, Language of

Thought, place cells, spatial navigation

ORCID
Sander van Bree https://orcid.org/0000-0003-4894-5938

REFERENCES

Akhlaghpour, H. (2022). An RNA-based theory of natural universal computation. *Journal of Theoretical Biology*, *537*, 110984. https://doi.org/10.1016/j.jtbi.2021.110984

Chalmers, D. J. (2023). The computational and the representational language-of-thought hypotheses. *Behavioral and Brain Sciences*, *46*, e269. https://doi.org/10.1017/s0140525x23001796

Fodor, J. A. (1975). *The language of thought*. Harvard university press.

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*, 190–194. https://doi.org/10.1038/nature05601

Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology*, *40*(1), 155–189. https://doi.org/10.1146/annurev.ps.40.020189.001103

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. Wiley-Blackwell. https://doi.org/10.1002/9781444310498

Gershman, S. J. (2023). The molecular memory code and synaptic plasticity: A synthesis. *Biosystems*, *224*, 104825. https://doi.org/10.1016/j.biosystems.2022.104825

Jeffery, K. J. (1998). Learning of landmark stability and instability by hippocampal place cells. *Neuropharmacology*, *37*(4), 677–687. https://doi.org/10.1016/S0028-3908(98)00053-7

Kazanina, N., & Poeppel, D. (2023). The neural ingredients for a language of thought are available. *Trends in Cognitive Sciences*, *27*(11), 996–1007. https://doi.org/10.1016/j.tics.2023.07.012

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford University Press.

Knierim, J. J. (2002). Dynamic interactions between local surface cues, distal landmarks, and intrinsic circuitry in hippocampal place cells. *The Journal of Neuroscience*, *22*(14), 6254–6264. https://doi.org/10.1523/JNEUROSCI.22-14-06254.2002

Latuske, P., Kornienko, O., Kohler, L., & Allen, K. (2018). Hippocampal remapping and its entorhinal origin. *Frontiers in Behavioral Neuroscience*, *11*, 253. https://doi.org/10.3389/fnbeh.2017.00253

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M.-B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, *309*(5734), 619–623. https://doi.org/10.1126/science.1114037

Muller, R. U., & 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. https://doi.org/10.1523/JNEUROSCI.07-07-01951.1987

O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in the freely moving rat: Why they fire where they fire. *Experimental Brain Research*, *31*(4), 573–590. https://doi.org/10.1007/BF00239813

Roth, E. D., Yu, X., Rao, G., & Knierim, J. J. (2012). Functional differences in the backward shifts of CA1 and CA3 place fields in novel and familiar environments. *PLoS ONE*, *7*(4), e36035. https://doi.org/10.1371/journal.pone.0036035

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*(5909), 1865–1868. https://doi.org/10.1126/science.1166466