A proposed neuroimaging experiment investigating compositional representations in relational perception

Atlas Kazemian and Kelsey Han

May 10, 2022

1 Introduction

Symbolic systems, including natural languages, are able to make "infinite use of finite means" (Von Humboldt, cited in [Chomsky et al., 2006], p. 15) through complex combinations of simple parts. In other words, they are compositional. Similarly, the brain is capable of forming complex, novel ideas from known, simpler concepts. This compositional nature of thought has led to the proposition of the Language of Thought (LoT) hypothesis ([Fodor, 1975]). LoT describes a symbolic, abstract representational system in the brain that is shared across different input modalities (language and vision), and enables the formation of infinite thoughts through the combination of existing concept representations. Although compositionality has been extensively studied in the domain of natural language, there has been little focus on exploring its implications in biological vision, including perception and cognition. Here, we explore compositionality in the domain of vision by considering a simple case of perceptual relations. We start by developing a formal definition of compositionality, and build upon pre-existing ideas to propose a neuroimaging experiment. In our analysis, we exploit the power of Tensor Product Representations to develop two possible frameworks for testing the compositinality of neural representations. Lastly, we discuss the implications of our work and possible future directions.

2 Formal Definition of Compositionality

In this section, a formal definition of compositionality is proposed by considering a set of inputs, outputs, their corresponding structures and functions applied to them. Starting with the input and output, let:

$$i = s(c_1, c_2, c_3), o = s'(C_1, C_2, C_3) \tag{1}$$

where i is an input of constituents c_i within a structure s, and o is the output with constituents C_i within a structure s'. Then, a function f is compositional if:

$$f(s, c_1, c_2, c_3) = (s', C_1, C_2, C_3), C_1 = F_1(c_1), C_2 = F_2(c_2), C_3 = F_3(c_3) \tag{2}$$

For some functions F_1, F_2, F_3 operating on individual constituents. In other words, a function is compositional if its inputs and outputs can be subdivided into constituents, and the function of the whole is built up of functions of the parts. Given this definition, we set out to describe mental representations as compositional structures over which cognitive functions apply.

3 Cognition as Computation over Compositional Structures

We use the theoretical constructs of systematicity and productivity from [Fodor and Pylyshyn, 1988] to describe compositionality in cognition. To elaborate, mental representations are systematic since the brain's ability to compute $F_1(c_1)$, $F_2(c_2)$, and $F_3(c_3)$ extends to its ability to compute $F_2(c_1)$, $F_1(c_3)$, etc. Mental processes are productive, meaning finite knowledge of language, logic, or calculus results in the brain's ability to handle infinitely many inputs, and form infinitely many representations despite its limited storage capacity.

Mental representations are patterns of activation over groups of neurons. Therefore, they can be described as numerical activation vectors. In order to make sense of the world, the representations for simple, individual concepts need to combine to form more complex and meaningful thoughts. For numerical activation vectors, this can be achieved using Tensor Product Representations (TPRs) ([Smolensky, 1990]). Encoding mental representations as activation vectors allows us to exploit the power of TPRs to model compositional structures in the brain.

4 Mental Representations as Tensor Product Representations

Constructing Tensor Product Representations requires that structures be viewed as a number of roles (possibly unbounded in certain contexts) which are bound to particular fillers for particular instances of the structure ([Smolensky, 1990]). For instance, a string can be viewed as possessing an infinite set of roles $\{r_1, r_2, r_3, ...\}$ where i is the role of the ith element. Then, a string like $'abc'$ involves bindings $\{a/r_1, b/r_2, c/r_3, ...\}$, where the roles r_i for $i > 3$ are all unbound. Similarly, roles can be temporal , spatial, relational, etc depending on the context. For instance, in speech processing, roles can be a set of time points and the fillers the amount of energy in a particular frequency band.

To construct the TPR, let S represent the set of all filler/role bindings for a structure, then we can write:

$$S = \{f/r | f \in F, r \in R\} \tag{3}$$

Where F and R represent the set of all fillers and roles respectively. Then, the set can be represented as a tensor product representation:

$$\{f_i/r_i\}_i = \mathbf{f}_i \otimes \mathbf{r}_i \tag{4}$$

Where $\mathbf{f}_i$ and $\mathbf{r}_i$ are the activation vectors of the ith role and filler taken from vector spaces of all roles and fillers. Mental representations can be modeled using TPRs by binding the right set of roles and fillers. For instance, the word "*achievable*" may be formed using two $filler$:role bindings, where the roles are the activation vectors encoding the mental representations of positions (left and right) in the word, and the fillers are the activation vectors encoding the mental representations of the words ("*achieve*", and "*able*"). This results in the two binding pairs: *Achieve*:left and *able*:right. We can then form a TPR by taking the sum of the tensor products that bind each role to its corresponding filler(Figure 1 top). TRPs can also be used to bind representations to more complex thoughts. For instance, for the simple phrase "*there is a pen inside the bag*", the roles could be *inside*, $reference$, and *target*, and the fillers the mental representations of the objects pen and cup. The $filler$:role bindings for this example are shown in Figure 1 (bottom).

5 Proposed experiment

5.1 Stimuli

The proposed experiment involves passive viewing of simple scene images inside an fMRI scanner. Each scene is composed of a pair of naturalistic objects arranged in a given spatial configuration. The scenes are varied systematically by changing the object identities and the spatial relations between a pair of objects. The object exemplars are drawn from one of 4 categories, CylinderA, CylinderB, CubeA, CubeB and spatial relations are one of 4 types $r_{left\_of}, r_{right\_of}, r_{in\_front\_of}, r_{behind}$. For example, a scene image can be a CylinderA object appearing on the left side of a CubeA object, or a CylinderB placed in front of a CubeB object. In each scene, one object serves as the landmark, whereas the other object is assigned as the target object. Spatial positions at which objects appear are randomized within a set of predetermined coordinates to keep retinal locations dissociated from the landmark/target roles. (Figure 2). In summary, there are a total of 32 possible conditions, with 4 object categories serving as the target or landmark for each of the 4 relational rules.

5.2 Experiment Design

Each trial consists of a fixation followed by a scene presentation. During fixation, subjects are instructed to gaze at a cross appearing at the center of the null background and maintain the fixation during the subsequent stimuli presentations. Each fixation is presented for 500 ms, followed by a scene

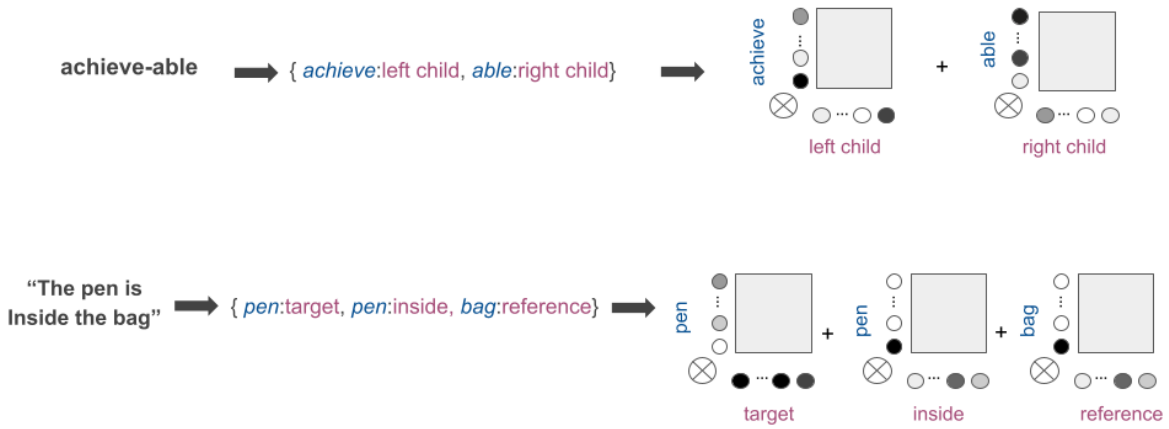

Figure 1: (Top) The Tensor Product Representation of a compound word. (Bottom) The Tensor Product Representation of a phrase relating two objects.

image presented for 2500 ms. Trials within the same run involve different exemplars of the same spatial relation and object types to ensure high signal-to-noise for each condition. Prior to each run, subjects are shown the target objects and asked to locate the target relative to the landmark while maintaining the fixation. Within each run, subjects are asked to locate the target stimuli relative to the landmark stimuli and perform a one-back test in which they press a button when two subsequent scenes are identical. Runs are repeated for as many times as needed (e.g., 8 times).

5.3 Frameworks

5.3.1 Predictive modeling approach

In this framework (Figure 3), model tensor product representations constructed from independent sets of vectors are compared to the neural responses obtained from the experiment. Here, the TPR for a scene is represented by the linear sum of the outer products between filler and role embedding vectors, where the embeddings come from a vector space VF of choice for fillers and a vector space VR for roles. The types of fillers and roles are determined by the experimenter's hypothesis. For example, one could presuppose different object entities to be fillers, while each of the relations and the target/landmark assignment are roles. The choice of vector spaces from which one obtains role and filler embeddings is an empirical question. They may be pre-trained embeddings such as the representations of hidden units in a pre-trained neural network or even taken from existing neural data. The crucial thing is that the neural representation in the brain during scene viewing is approximated by the TPR, where the response is decomposable into the representations of its constituent parts. The model TPR can then be compared on the representations in different brain regions of interest (e.g., high-level areas in visual cortex or the fronto-parietal network) to adjudicate between different brain regions that may employ TPRs during the perception of a compositional scene.

5.3.2 Role-Unbinding with TPGN

Another approach is to build a tensor product generation neural network to investigate the TPRs in the brain (Figure 4). In this approach, the embedding spaces for filler and role vectors are not determined prior to the experiment but instead are derived by decomposing the neural responses obtained during the experiment. Once we have the fMRI scans for each scene image, the neural response for the image

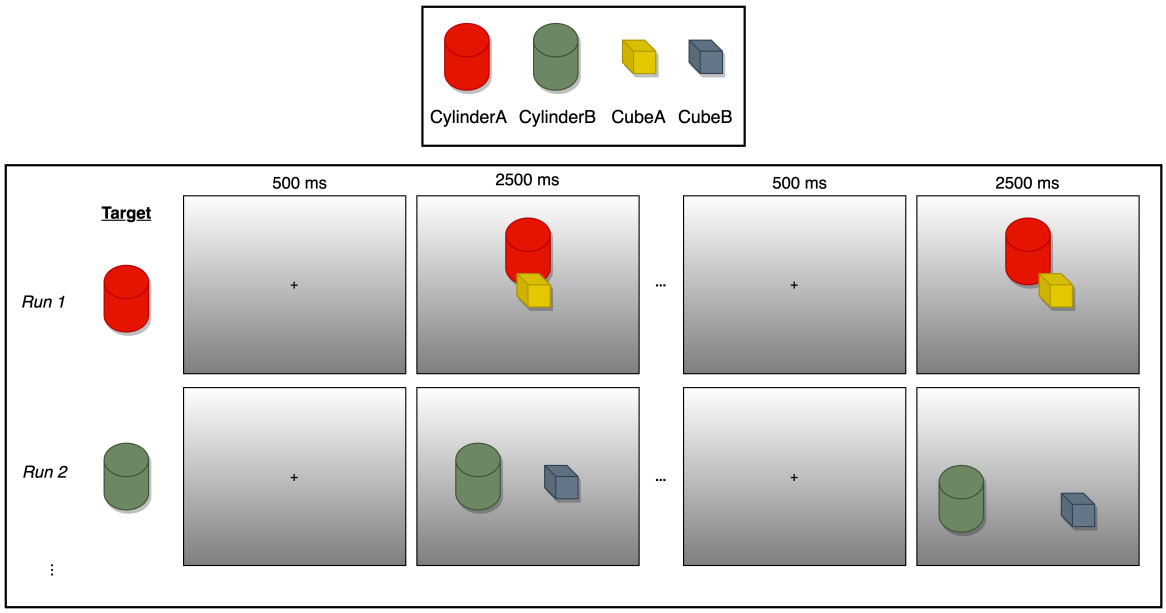

Figure 2: (Top) Stimuli, (Bottom) A sample experiment design.

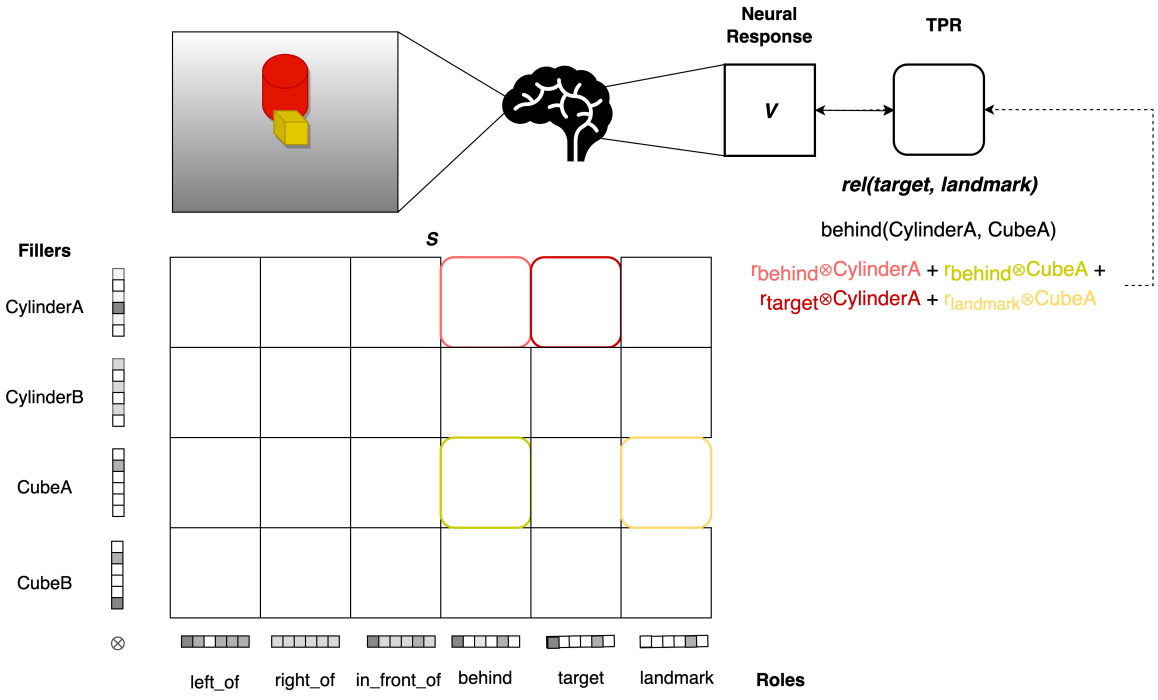

Figure 3: Predictive modeling approach with an example TPR.

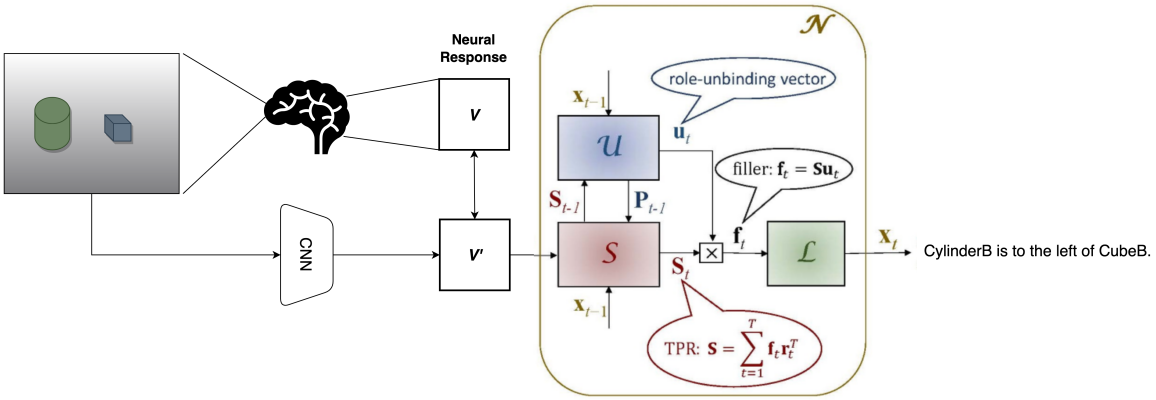

Figure 4: Role-unbinding approach with a CNN encoder and TPGN (figure adapted from [Huang et al., 2017]).

is regarded as a TPR, and the goal is to train TPGN, or a TPR-capable generation network, that learns the role-unbinding vector representations to decompose the TPR ([Huang et al., 2017]). The low SNR of neural responses may result in overfitting, which can be mediated by using an CNN encoder fit with neural responses.

6 Implications

In this work, we seek to investigate the compositional nature of canonical perceptual relations by testing one possible instantiation of compositional structures upon which the brain computes, the TPRs. The existence of certain neural codes such as grid-like mapping suggests that the neural infrastructure for compositional structures may support TPRs in the brain ([Frankland and Greene, 2020], [Whittington et al., 2020]). However, the extent to which these representations are utilized for various tasks is largely unknown. While there has been empirical evidence for the presence of TPRs in the brain, previous works are limited to either low-level processes such as modulation of receptive field activity by retinal positions and spatial transformations in the parietal cortex ([Andersen et al., 1985], [Pouget and Sejnowski, 1997]) or a simple conceptual combination ([Baron and Osherson, 2011]).

The question of whether perceiving spatial relations involves compositional representations is an interesting unexplored frontier as it sits at the interaction of low-level perceptual processing and high-level reasoning. On the one hand, the task seems to naturally lend itself to compositional structures and involve sufficiently high-level cognitive processing that easily interfaces with linguistic representations. On the other hand, the relational meaning is extracted from a short presentation in an almost involuntary and effortless fashion, a signature of automatic visual processing ([Hafri and Firestone, 2021]).

If the neural responses during relational perception are successfully modeled by TPRs, it may suggest that TPRs are an ubiquitous form of representation utilized by the brain to support efficient readout for downstream compositional tasks. The next steps would be investigating to what extent these compositional representations apply beyond physical relations. For example, one fruitful direction would be to apply the method outlined here to uncover compositional representations during event and social perception.

References

[Andersen et al., 1985] Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230(4724):456–458.

[Baron and Osherson, 2011] Baron, S. G. and Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage*, 55(4):1847–1852.

[Chomsky et al., 2006] Chomsky, N. et al. (2006). *Language and mind*. Cambridge University Press.

[Fodor, 1975] Fodor, J. A. (1975). *The language of thought*, volume 5. Harvard university press.

[Fodor and Pylyshyn, 1988] Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71.

[Frankland and Greene, 2020] Frankland, S. M. and Greene, J. D. (2020). Concepts and compositionality: in search of the brain's language of thought. *Annual review of psychology*, 71:273–303.

[Hafri and Firestone, 2021] Hafri, A. and Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences*, 25(6):475–492.

[Huang et al., 2017] Huang, Q., Smolensky, P., He, X., Deng, L., and Wu, D. (2017). Tensor product generation networks for deep nlp modeling. *arXiv preprint arXiv:1709.09118*.

[Pouget and Sejnowski, 1997] Pouget, A. and Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of cognitive neuroscience*, 9(2):222–237.

[Smolensky, 1990] Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1-2):159–216.

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