Compositionality of neural representations of logical rules: An fMRI study

elena.benini@psych.rwth-aachen.de
carlo.reverberi@unimib.it

Elena Benini[1], Lilia Del Mauro[2], Gianpaolo Basso[3], Giosuè Baggio[4], Carlo Reverberi[2,5]

[1]RWTH Aachen University, Germany, [2]Department of Psychology, University of Milano-Bicocca, 20126, Milan, Italy, [3]Department of Medicine and Surgery, University of Milano-Bicocca, 20900, Monza, Italy, [4]NTNU, Trondheim, Norway, [5]NeuroMi — Milan Center for Neuroscience, 20126 Milan, Italy

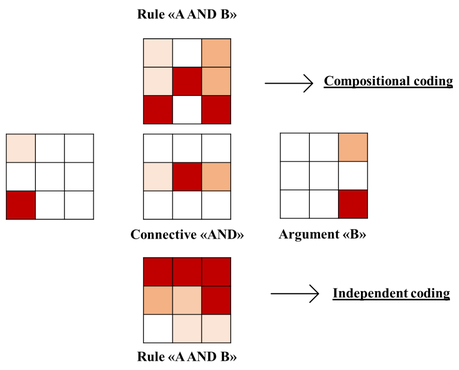
Introduction

Rules are widely used in everyday situations to pursue relevant goals. Rules are often composed by two **propositions** combined by the **elementary logical connectives AND, OR, and IF**.

Rule representation seems to involve a fronto-parietal network (Baggio et al., 2016; Zhang et al., 2013). The neural representation of *compound* rules has been shown to be **compositional** (Reverberi et al., 2012a), with partially segregated brain networks encoding the order of the simple rules composing the compound rule (Reverberi et al., 2012b).

Compositionality

A formulation of the principle of semantic compositionality in linguistic maintains that "at least one meaning of a whole is a function only of the meanings of the parts and of the way they are syntactically combined" (Baggio, 2021, p. 8). The neural representation of a rule is compositional if "representations of parts are recoverable from representations of wholes" (Baggio, 2021, p. 12).



In the middle, activity patterns coding for the elements composing a rule: proposition A, connective AND, and proposition B. In the compositional code (at the top), the pattern coding for the rule is the combination of the patterns of the single elements composing it. In the independent code (bottom), the pattern for the rule is unrelated to the patterns of its composing elements.

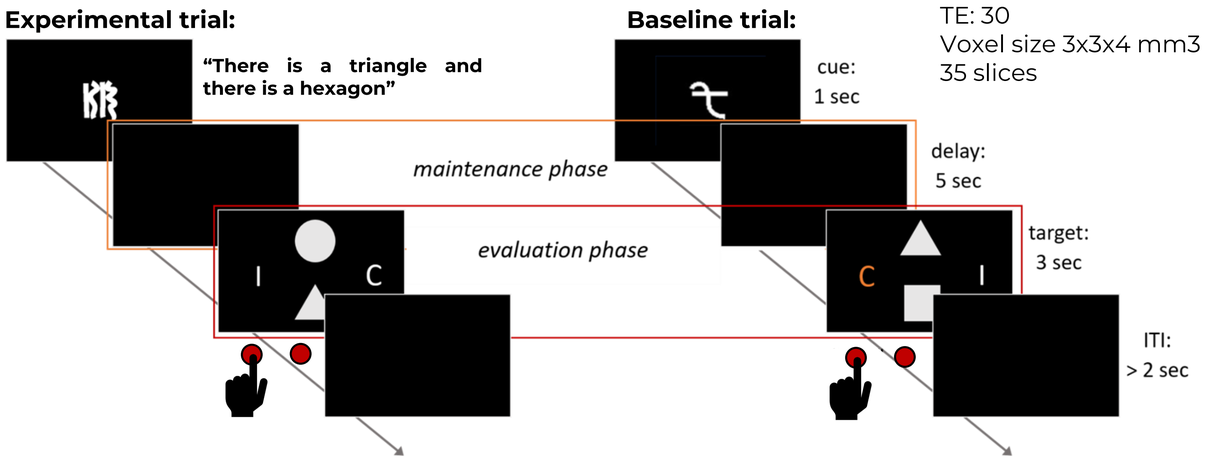
We investigated whether the **neural representation of simple rules formed by two propositions and a logical connective** (e.g., "*There is a triangle* and *there is a hexagon*") contains some features of the neural code of the elements composing the rule (i.e., **is compositional**).
We applied multivariate pattern analysis (MVPA) to functional magnetic resonance imaging (fMRI) data collected while participants maintained in working-memory and evaluated simple rules.

Methodology

32 healthy volunteers

Philips Igenia 3 T
TR: 2000 ms
TE: 30
Voxel size 3x3x4 mm3
35 slices

Experimental trial:



Baseline trial:

The trials start with a cue prompting participants to retrieve the associated rule. During the *maintenance phase,* participants maintain such a rule in working-memory. In the *evaluation phase,* they judge whether a target scenario composed of two shapes is compatible or incompatible with the rule. In *baseline* trials, participants perform the same motor action (i.e., a keypress) but the rule is "Press the key to the side of the orange letter".

Experimental conditions. 3 logical connectives (AND, OF, IF) * 2 proposition pairs (there is a triangle-there is a hexagon; square-circle) * 2 orders (T-H, S-C and H-T, C-S) + 1 baseline + 1 memory trials + 1 catch trials = 15 conditions.

FMRI protocol. 6 runs, 64 trials each (4 repetitions of each rule + 4 baseline trials + 6 memory trials + 6 catch trials).

Decoding Analyses

Multivariate Pattern Analysis
BOLD signal in each 2-seconds time bin (with a FIR model) is predicted for each voxel as a function of the current connective and/or propositions and their order. In a whole-brain searchlight analysis (Eztel et al., 2013), spheres of 2-voxels radius of beta coefficients are fed to a Support Vector Machine (SVM) classifier that learns to distinguish such patterns based on the experimental condition, for example, the logical connective.

Decoding rules representations versus compositionality of rules elements
SVM classifying accuracy can tested on new beta patterns derived from trials with…
… the same rules ("leave-one-run-out" cross validation) → Reveals **neural representations of rules**
… on different rules → Reveals **compositionality of the rule element**

Example of analysis to determine compositionality of representations of logical connectives

Test set	AND rules with triangle and hexagon	IF… THEN rules with triangle and hexagon
Train set	AND rules with square and circle	IF… THEN rules with square and circle

Investigated neural representation and compositionality of (1) **logical connectives**, (2) **propositions and** (3) **propositions order.**

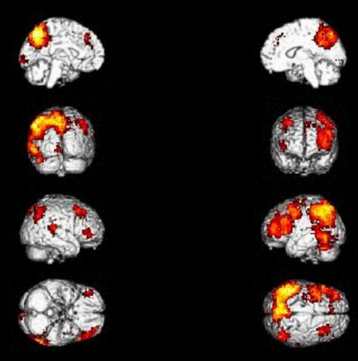
Each analysis run on **maintenance phase** and on **evaluation phase.**

Results

No mask, no family-wise error correction, alpha < 0.005 at the voxel level, cluster > 30 voxels

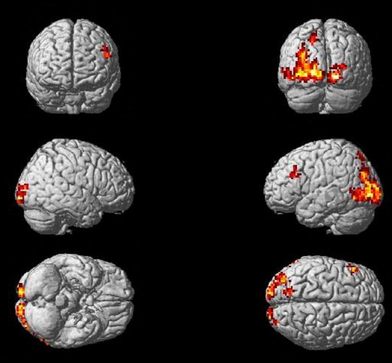
Maintainance phase

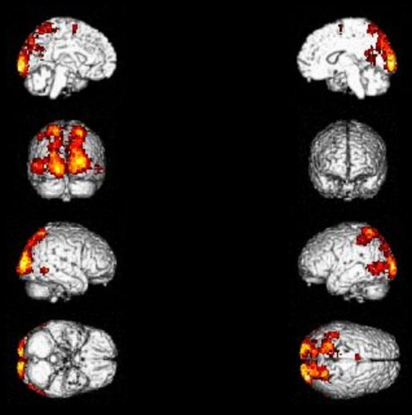
Neural representation of rules elements



Clusters encoding **representations of the different connectives**: left inferior parietal cortex (BA40), temporal middle gyrus (BA21), right angular cortex (BA39), right frontal middle gyrus (BA9), right temporal superior cortex (BA22), right inferior frontal gyrus-pars triangularis (BA45), left inferior occipital cortex (BA17).

Clusters encoding **representations of the different propositions**: right calcarine sulcus (BA18), left middle occipital cortex (BA19), left inferior frontal gyrus-pars triangularis (BA44), right occipital middle gyrus (BA18), left superior parietal lobule (BA7).



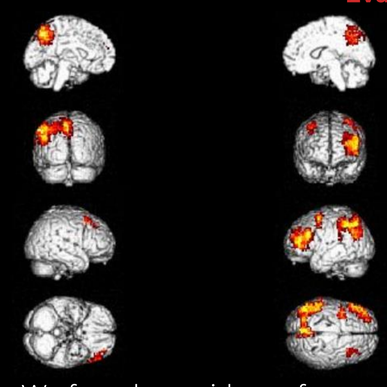


Clusters encoding **representations of the different propositions orders**: left inferior occipital gyrus (BA18), left supplementary motor area (BA6), right inferior temporal gyrus (BA37).

We did not find clusters encoding the neural representations of propositions and propositions order during *evaluation* phase.

Evaluation phase

Compositionality



Clusters encoding the **different logical connectives independenlty from propositions and propositions order**: left middle frontal area (BA46), left precuneos, left precentral gyrus (BA6); right middle frontal gyrus (BA8).

We found no evidence for compositionality of connectives representations during *maintenance*, neither evidence for compositionality of propositions and proposition orders.

Discussion

Connectives might be represented compositionally, especially when their meaning is accessed, namely, during the evaluation phase.
→ Accessing the meaning of the connective is necessary to access rule meaning.
→ At least at certain stages, neural representations of logical connectives is compositional. There exists a compositional processing path from sentences to their meaning (Baggio, 2021).

Rules elements (propositions and their order) seem not to be represented compositionally.
→ Meaning accessed in combination with connective or our measures are not sensitive enough.

Nonetheless, neural patterns of logical rules can be distinguished based on rules elements, especially during superficial representation, namely during maintenance, when the rule is cued and probably retrieved but not evaluated.
Interestingly, different areas encode different elements (rules, propositions, propositions order).

We found evidence that the neural representations of logical connectives such as AND, OR and IF-THEN as they appear in simple rules such as "If A then B" is compositional.

Literature Selection

Baggio, G. (2021). Compositionality in a parallel architecture for language processing. *Cognitive Science*, 45. https://doi.org/10.1111/cogs.12949
Baggio, G., Cherubini, P., Pischedda, D., Blumenthal, A., Haynes, J. D., & Reverberi, C. (2016). Multiple neural representations of elementary logical connectives. *NeuroImage*, 135, 300–310. https://doi.org/10.1016/j.neuroimage.2016.04.061
Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041
Reverberi, C., Görgen, K., & Haynes, J.-D. (2012a). Compositionality of Rule Representations in Human Prefrontal Cortex. Cerebral Cortex, 22(6), 1237–1246. https://doi.org/:10.1093/cercor/bhr200
Reverberi, C., Görgen, K., & Haynes, J.-D. (2012b). Distributed Representations of Rule Identity and Rule Order in Human Frontal Cortex and Striatum. *The Journal of Neuroscience,* 32(48), 17420–17430. https://doi.org/10.1523/JNEUROSCI.2344-12.2012
Zhang, J., Kriegeskorte, N., Carlin, J. D., & Rowe, J. B. (2013).Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions. *The Journal of Neuroscience*, 33(29), 11852–11862. https://doi.org/10.1523/JNEUROSCI.5193-12.2013