Rate-Coding Bundle Memory: A Unified Model of Memory and Control for Language Comprehension

Teun van Gils (Teun.vanGils@mpi.nl)
Neurobiology of Language Department, Wundtlaan 1
6525 XD Nijmegen, The Netherlands

Rowan P. Sommers (Rowan.Sommers@uni-osnabrueck.de)
Institute of Cognitive Science, Wachsbleiche 27
49090 Osnabrück, Germany

Abstract

Neurobiological and cognitive models often tackle individual facets of cognition, yet few successfully integrate across multiple domains. For instance, (pre-transformer) connectionist approaches explain a broad range of cognitive phenomena, including graded semantics and context sensitivity, but struggle with symbolic reasoning, compositionality and dynamic variable binding, necessary components of language comprehension (Fodor & Pylyshyn, 1988; Marcus, 1998; Lake et al., 2017; Kazanina & Poeppel, 2023). Conversely, symbolic models excel at these tasks but often lack a biologically inspired explanation of how these symbolic operations might be implemented in the brain (Do & Hasselmo, 2021). We propose a unified model, inspired by (psycho-)linguistic theory (McElree et al., 2003; Seuren, 2009), that integrates these aspects, focusing on the interplay between memory, unification and control (Baggio & Hagoort, 2011). This model, which we call Rate-Coding Bundle Memory (RCBM), is designed to be both biologically plausible and capable of addressing a variety of cognitive tasks, including those that require the incremental integration and differention of entities in linguistic descriptions of scenes. We demonstrate the model's performance on a range of tasks, including controlled storage, memory retrieval, tracking multiple memories, and semantic inference; and show, by lesioning different components of the model, that the various memory and control components are crucial for the model's ability to perform these tasks. Our model improves upon existing working memory models of the prefrontal cortex (Manohar et al., 2019; Fiebig et al., 2020), and shows parallels with properties of grid- and place cells during hippocampal replay (Do & Hasselmo, 2021; Kurth-Nelson et al., 2023; Kazanina & Poeppel, 2023).

Keywords: working memory; symbolic computation; recurrent neural networks; cognitive neuroscience; computational linguistics

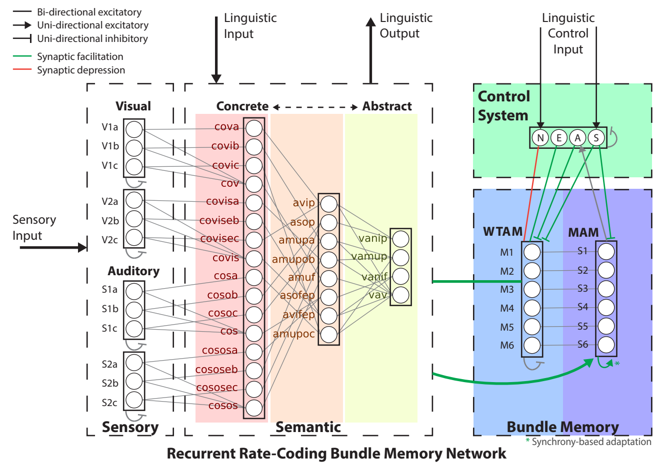

Figure 1: *The RCBM model consists of three main components: a semantic system, a control system, and a memory system. The semantic system (left and center) consists of both visual and auditory sensory feature neurons (left) and conceptual feature neurons (center). Each neuron in the semantic system can directly receive external input through corresponding sensory or linguistic input. Linguistic input consists of multiple sentences in an artificial language, presented word-by-word. Each word is represented by a conceptual feature neuron, and these conceptual feature neurons are organized in increasingly abstract layers, with the rightmost layer representing the most abstract concepts. Each non-sensory semantic neuron can directly output corresponding artificial words for the model to produce a response. The control system (right top) consists of four control neurons: a novelty detection neuron (N), an existing memory detection neuron (E), an ambiguity detection neuron (A), and a memory suppression neuron (S). Special linguistic inputs can directly activate the novelty detection neuron (determiner, e.g.* ***a*** *man v.s.* ***the*** *man) and the suppression neuron (end-of-sentence token). The memory system (right bottom) consists of two memory pools: a Winner-Take-All Memory pool (WTAM) and a Multiple Activation Memory pool (MAM). Each memory pool consists of a set of memory neurons, which are connected to the conceptual feature neurons and have pairwise connections between them. The control system interacts heavily with the memory system, as it is responsible for activating the appropriate memory neurons and suppressing the memory system when necessary.*

Here, we present the Rate-Coding Bundle Memory (RCBM) model, which builds on work by Manohar et al. (2019), whose memory model aligns with multiple empirical findings about working memory storage and retrieval. Their model dynamically binds different semantic features together into *conjunctive neurons*, which can subsequently be retrieved through a winner-take-all mechanism. Although this approach maps well onto neural correlates of memory retention, it still faces difficulties in tasks such as discourse incrementation (accumulating context as new information arrives), novelty detection (separating new entities from familiar ones), and the *problem of two* (differentiating similar referents without conflation). These limitations highlight the need for a control mechanism – one that can orchestrate read-write processes and dynamically suppress or enhance the retrieval of specific memories.

We redesigned this rate-coding model for a more linguistic task, and extended it to include a control system (Figure 1). We refer to our model as the Rate-Coding Bundle Memory (RCBM) model. The main tasks of the control system include the identification of novel and existing stimuli and of detecting ambiguity in the input, in order to aid the focused storage, modification, and retrieval of the appropriate items. By presenting artificial words to the network one by one, the network identifies entities that are being referred to and is able to e.g. distinguish between two similar referents (Figure 2).

We systematically tested this approach with a diverse set of tasks wherein the network had to perform tasks across

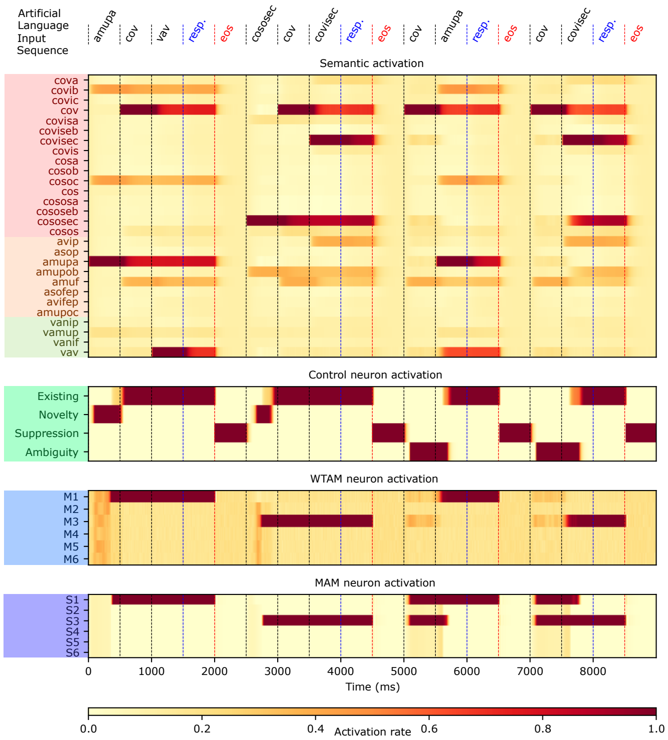

Figure 2: *An example trial of the problem of two test, presented as a sequence of words in an artificial language (top). In the first two sentences, the model is introduced to two entities with an overlapping feature, e.g. a blue square and a purple square. The next two sentences each start with this overlapping feature (e.g. square) and then introduce a distinguishing feature (e.g. either blue or purple), requiring the model to temporarily withold its response until the distinguishing feature is presented. The model stores the two entities in M1/S1 (sentence 1) and M3/S3 (sentence 2) in the WTAM and MAM pools, respectively (bottom); when presented with the overlapping feature (sentence 3/4), the ambiguity detection neuron is activated (middle), supressing the winner-takes-all process until disambiguating information is presented.*

4 categories: controlled storage, memory retrieval, multiple memories, and semantic association. These tasks encompass challenges required for the comprehension of written language, aiming to build up a coherent semantic representation of the presented information over time. Performance was near-perfect on all tasks (Figure 3, top row). Lesioning different components of our model lead to measurable reductions in performance (Figure 3), indicating that the control and memory systems are crucial for the model's ability to perform these tasks.

Our results show that our proposed control mechanism can be used for robust one-shot learning, ambiguity resolution, and compositional semantic integration, all within a framework that remains biologically plausible (which is not clear for transformers). We propose that anatomically plausible regions for

Figure 3: *The RCBM model performs well on all tests, while lesions to the model lead to systematic reductions in performance. Each bar represents the proportion of trials (right y-axis) that the model successfully (blue) or unsuccessfully (red) completed for each test. The tests are displayed on the x-axis (top) and grouped into four categories. The different lesions are displayed on the y-axis (left), with each lesion representing a different part of the model that was turned off. Per test, there were up to 4 trial types that were each repeated 25 times, for a total of 12000 trials across all tests and lesions.*

memory integration such as the hippocampus and dorsolateral prefrontal cortex could be directed by these control signals, gating which memories are retrieved, updated or suppressed at any moment, enabling the understanding of linguistic discourse. With RCBM, we thus offer a single model capable of addressing multiple recognized challenges in cognitive neuroscience, and we hope that this work pushes forward the idea of a unified account wherein connectionist and symbolic aspects coexist naturally, rather than being "duct-taped" together at the algorithmic level.

Acknowledgments

This research was funded by the Max Planck Society for the Advancement of Science (mpg.de/en). We thank Peter Hagoort and Markus Ostarek for their support and insights leading to this work, and Sanjay Manohar for helpfully sharing his insights into the original model and feedback on our work.

References

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the n400. , *26*(9), 1338–1367. (ISBN: 0169-0965) doi: 10/bx8w4g

Do, Q., & Hasselmo, M. E. (2021). Neural circuits and symbolic processing. , *186*, 107552. Retrieved 2025-04-07, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10121157/ doi: 10.1016/j.nlm.2021.107552

Fiebig, F., Herman, P., & Lansner, A. (2020). An indexing theory for working memory based on fast hebbian plasticity. , *7*(2), ENEURO.0374–19.2020. Retrieved 2020-10-28, from http://eneuro.org/lookup/doi/10.1523/ENEURO.0374-19.2020 doi: 10.1523/ENEURO.0374-19.2020

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

Kazanina, N., & Poeppel, D. (2023). The neural ingredients for a language of thought are available. , *27*(11), 996–1007. Retrieved 2023-11-29, from https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(23)00193-6 (Publisher: Elsevier) doi: 10.1016/j.tics.2023.07.012

Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., ... Schwartenbeck, P. (2023). Replay and compositional computation. , *111*(4), 454–469. Retrieved 2025-04-07, from https://linkinghub.elsevier.com/retrieve/pii/S0896627322011254 doi: 10.1016/j.neuron.2022.12.028

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. , *40*, e253. Retrieved 2025-04-07, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/building-machines-that-learn-and-think-like-people/A9535B1D745A0377E16C590E14B94993 doi: 10.1017/S0140525X16001837

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. , *101*, 1–12. Retrieved 2023-11-12, from https://www.sciencedirect.com/science/article/pii/S0149763419300624 doi: 10.1016/j.neubiorev.2019.03.017

Marcus, G. F. (1998). Rethinking eliminative connectionism. , *37*(3), 243–282. Retrieved 2020-08-30, from https://linkinghub.elsevier.com/retrieve/pii/S0010028598906946 doi: 10.1006/cogp.1998.0694

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. , *48*(1), 67–91. Retrieved 2022-01-19, from https://linkinghub.elsevier.com/retrieve/pii/S0749596X02005156 doi: 10/fk5ppj

Seuren, P. A. M. (2009). *The logic of language: Language from within volume II*. Oxford University Press.