

UC Davis

UC Davis Previously Published Works

Title

The Structure of Systematicity in the Brain

Permalink

<https://escholarship.org/uc/item/1pp1q2sd>

Journal

Current Directions in Psychological Science, 31(2)

ISSN

0963-7214

Authors

O'Reilly, Randall C
Ranganath, Charan
Russin, Jacob L

Publication Date

2022-04-01

DOI

10.1177/09637214211049233

Peer reviewed



HHS Public Access

Author manuscript

Curr Dir Psychol Sci. Author manuscript; available in PMC 2022 June 30.

Published in final edited form as:

Curr Dir Psychol Sci. 2022 April ; 31(2): 124–130. doi:10.1177/09637214211049233.

The Structure of Systematicity in the Brain

Randall C. O'Reilly,

Charan Ranganath,

Jacob L. Russin

Departments of Psychology and Computer Science, Center for Neuroscience, University of California, Davis, 1544 Newton Ct, Davis, CA 95618

Abstract

A hallmark of human intelligence is the ability to adapt to new situations, by applying learned rules to new content (systematicity) and thereby enabling an open-ended number of inferences and actions (generativity). Here, we propose that the human brain accomplishes these feats through pathways in the parietal cortex that encode the abstract structure of space, events, and tasks, and pathways in the temporal cortex that encode information about specific people, places, and things (content). Recent neural network models show how the separation of structure and content might emerge through a combination of architectural biases and learning, and these networks show dramatic improvements in the ability to capture systematic, generative behavior. We close by considering how the hippocampal formation may form integrative memories that enable rapid learning of new structure and content representations.

The ability to directly transfer existing knowledge to new situations by applying learned rules to new content (*systematicity*) and thereby generate an open-ended number of different behaviors (*generativity*) is particularly advanced in humans relative to other animals. However, this ability is nevertheless shared in a limited but instructive way with even relatively simple computer programs. For example, it is easy to write a function that takes two arbitrary strings as arguments, and systematically returns the concatenation of those two strings. With just a bit more logic, calling such a function repeatedly can easily construct an open-ended number of sentences. However, learning to do something like this in a neurally-plausible manner is much more challenging, and neural network models continue to be criticized as lacking in these signature human abilities, with much of their recent successes potentially attributable to something closer to rote memorization of increasingly large datasets (Fodor & Pylyshyn, 1988; Lake & Baroni, 2017; Marcus, 2018; O'Reilly et al., 2014; Plaut, McClelland, Seidenberg, & Patterson, 1996). Likewise, empirical cognitive neuroscience is only beginning to uncover the relevant brain areas involved in these abilities (Constantinescu, O'Reilly, & Behrens, 2016; Frankland & Greene, 2020; Park, Miller, Nili, Ranganath, & Boorman, 2020; Summerfield, Luyckx, & Sheahan, 2020).

oreilly@ucdavis.edu .

R. C. O'Reilly is Chief Scientist at eCortex, Inc., which may derive indirect benefit from the work presented here.

The essential trick employed by the simple concatenation function is that arbitrary content can be routed into it and operated upon generically, independent of any details of the content. The cognitive equivalent of this, which is widely recognized as critical for human systematic behavior (Fodor & Pylyshyn, 1988) is a separation of *structure* (i.e., the rule-processing system, akin to the function) from *content*. In language, syntax has traditionally been thought of as an example of content-independent structure, and Chomsky's famous example: "colorless green ideas sleep furiously" demonstrates that people can accurately judge grammatical correctness with arbitrary content. Without adopting traditional abstract symbolic syntactic frameworks, understanding how something like content-independent syntactic structure can be learned and represented via known biological mechanisms is thus a critical step toward advancing our understanding of the neural basis of our signature human abilities.

In this paper, we review cognitive and neural evidence consistent with the separation between structure and content across various cognitive domains, and recent neural network models that demonstrate how such a separation might emerge through a combination of architectural biases and learning, producing significantly greater systematicity. Critically, structure and content also need to interact and be integrated in various ways, and certain brain areas appear to be specialized for this integration, so as is often the case, the brain is likely to apply multiple representational strategies in parallel.

Structure and Content in Cognitive Models

The FINST (fingers of instantiation) framework of Pylyshyn (1989) provides an early, simple model for how structure can be represented independent of specific content and also maps well onto cognitive neuroscience data reviewed in the next section. The key idea is that abstract, content-independent, pointer-like "fingers" can index a small set (up to about 4) of different visual locations at a time. These content-independent pointers were originally proposed to serve as place-holders for encoding relationships among items in a scene (e.g., INSIDE(a,b)).

The core ideas in the FINST framework can be extended to represent any kind of abstract structure that specifies the relationships between different indexed elements. For example, to describe the action of *giving*, there are distinct functional roles (giver, recipient, item), and each of these could be represented using a separate FINST-like thematic role pointer, instead of using the semantic content specific to a particular situation (e.g., "Radhika gave Charan a gift"). Because *giving* is encoded in terms of content-independent roles indexed by FINST-like pointers, these representations can automatically generalize to new contents, and additional inferences could be made about the properties and implications of the structural relationship (e.g., the giver no longer has the item, and may have some expectation of reciprocity depending on the nature of the transaction and relationship). Encoding sufficiently rich structural knowledge to do so requires much more than FINST's, but encoding this knowledge in a relatively content-independent manner would clearly support systematic reasoning.

In short, FINST-like indexes provide a plausible attention-based neural mechanism for role-filler variable slots in the context of classical symbolic representational frameworks. At a cognitive level, the kinds of elaborated structural representations and processes that have been studied include a range of different levels of complexity, from verb / action based elements as in the *giving* example (Boylan, Trueswell, & Thompson-Schill, 2015) to more elaborated *schemas* or *scripts* describing longer sequences of events (e.g., the schema associated with a kid's birthday party), as we discuss below. The ability to transfer structural knowledge across content domains is also central to established models of analogical reasoning (Gentner & Holyoak, 1997) (e.g., the relational structure of *orbiting* can be transferred from the planetary to atomic scale).

Structure Representation in the Human Brain

There is a well-established distinction between spatial and object processing in the human brain, which can be reframed as one example of how the brain separates structure and content via distinct, but interacting pathways. Visual (and auditory) networks in the brain route sensory input into distinct dorsal and ventral stream pathways (Ungerleider & Mishkin, 1982), with the ventral visual pathway extending from early visual cortex to inferotemporal (IT) cortex characterized as the *What* pathway, specialized for object or scene recognition (i.e., visual *content*). The dorsal pathway through the parietal lobe is specialized for spatial *Where* processing, based on extensive evidence that this pathway represents spatial and relational information in a relatively content-independent manner.

The potential for the dorsal stream pathway to support systematic structure-sensitive processing was already well developed by Pylyshyn (1989), and has been incorporated into psycholinguistic theories (Frankland & Greene, 2020; Landau & Jackendoff, 1993). Furthermore, there is substantial evidence that posterior parietal cortex encodes nonspatial structural information as well. First, it is clear that the parietal lobe plays a critical role in the sensory guidance of action performance (Orban, Van Essen, & Vanduffel, 2004), as captured in the *What vs. How* ventral / dorsal framework (Goodale & Milner, 1992). In humans, this action coding extends to the representation of verb-based argument structure (e.g., as in the *giving* example above) in the inferior parietal area (angular gyrus), supporting the core of structure at the sentence level (Binder & Desai, 2011; Boylan et al., 2015; Palmer, Bonial, & Hwang, 2016). More abstract structural and linguistic concepts are thought to build directly upon these parietal action, space and time foundations (Frankland & Greene, 2020; Landau & Jackendoff, 1993), including event representations in higher-order parietal areas as discussed below.

It is well established that parietal spatial and action representations anticipate the effects of eye, head, or body movements (Cavanagh, Hunt, Afraz, & Rolfs, 2010) — this suggests that acquisition of structure in the parietal lobe could be based on *predictive learning* (O'Reilly, Russin, Zolfaghar, & Rohrlich, 2021). Specifically, learning driven by the difference between a predicted sensory outcome of an action and the subsequent sensory input can drive improved predictions and shape the formation of more abstract structural representations to more efficiently generate these predictions. While it is fairly straightforward to learn to predict specific sensory outcomes from motor actions, an

important area of current research is to determine the extent to which more abstract, structural representations can emerge, capturing the consistent, generalizable relationships that hold across a large number of such actions (Summerfield et al., 2020).

Even if the dorsal-ventral streams separately encode structure and content in the brain, fundamental questions remain about how these streams interact at the level of detailed neural mechanisms, to support systematic cognitive function. A recent neural network model provides a useful example for how FINST-like attentional pointers can operate on newly-learned content information, in the context of separate structure and content representations (Russin, Jo, O'Reilly, & Bengio, 2020). Unlike many existing models that have relied upon various hand-coded mechanisms to directly emulate programming-language like variable-binding functionality in neural hardware, this model learned entirely via error-driven learning, with only very broad, biologically plausible architectural constraints between two processing pathways.

One such pathway had full access to the temporal ordering of words within a sentence, while the other was only able to process the single current word at any given time (Figure 1). Furthermore, the temporal-order sensitive pathway could only influence the network output via attentional modulation of the other pathway (similar to the FINST-like attentional pointers). Distinctions such as these could plausibly derive from evolved differences in the initial wiring of the neural architecture, providing affordances upon which subsequent learning operates.

With these constraints in place, the network learned syntax-like structural representations in the temporal-order pathway, and semantic-like content information in the other pathway, and in so doing, exhibited significantly more systematic behavior on challenging out-of-domain generalization tests than unstructured models (Lake & Baroni, 2017). Specifically, when the model was trained on examples generated from a simple phrase structure grammar, the syntax-like pathway was able to direct attention to the proper sequence of content items based on commands like “turn left twice,” and generalize this to a novel verb, e.g., “jump twice.” The ability of attention-like mechanisms to drive this FINST-style indexing into novel content items is consistent with the central importance of such mechanisms in the neural basis of structure / content dynamics. Furthermore, this model demonstrates how learned, distributed neural representations can take on a syntax-like role without traditional explicit symbolic grammar trees. However, more work is needed to understand the nature of structure representations in the brain supporting the more complex and sophisticated forms of human cognition, compared to these simplified models.

Interestingly, a similar distinction between structure and content pathways was achieved in a fairly different way using a variable-binding system based on an explicit slot-filler (key-value) lookup table mechanism (Webb, Sinha, & Cohen, 2021). In this model, the structure processing pathway learned to control this lookup table independent of the specific content stored there, while the content pathway learned representations of individual items. This architectural structure / content separation enabled the model to exhibit much more systematic behavior than comparison models without such a separation. Other related architectures are discussed in Akyürek and Andreas (2021).

While these models provide important initial demonstrations of how a separation between structure and content can support systematicity, it is critical to appreciate that many real-world situations also require strong interactions between the two. This has perhaps been most extensively studied in the case of pronouncing English words: as every English-as-a-second-language learner knows too well, the number of exceptions and sub-regularities are mind-spinning, and monolithic neural network pathways may provide the most appropriate mechanism for learning these mappings (Plaut et al., 1996). Likewise, the difference in parsing “Fruit flies like a banana” vs. “Time flies like an arrow” requires an interaction between syntax and semantics. In the visual domain, these interactions can be illustrated in the case of viewing a scene of a kitchen, where incoming content information may activate structural representations of typical kitchens, which then guide visual attention and semantic processing (Hayes & Henderson, 2019). In other words, human cognition likely represents a combination of separated and interacting streams (Franklin, Norman, Ranganath, Zacks, & Gershman, 2020; O’Reilly et al., 2014).

Binding of Structure and Content in the Medial Temporal Lobes

So far, we have used the classic model of the dorsal and ventral streams as one example of the separation of structure and content, but structure exists across multiple scales. For example, structural knowledge is required to parse and comprehend individual sentences, but comprehending a sentence within the context of a larger text passage often requires a mental representation of the structure of events depicted in that passage. Recent fMRI research has uncovered a candidate set of brain regions that may represent the abstract structure of events — the *Posterior Medial* (PM) network — and regions that may represent the characteristics of the specific people, places and things that serve as event content — the *Anterior Temporal* (AT) network (Figure 2). The PM and AT networks can be considered as anatomically higher up from the classic dorsal (PM) and ventral (AT) streams (Kravitz, Saleem, Baker, & Mishkin, 2011; Ranganath & Ritchey, 2012). Although these networks are distinct, the hippocampus is in a key position to integrate information across the two networks. AT network regions project predominantly to the lateral entorhinal cortex (LEC), and PM network regions mostly project to the medial entorhinal cortex (MEC). The hippocampus proper then binds information from these streams together into unitary episodic memories (Ranganath & Ritchey, 2012). Considerable evidence suggests that the PM and AT networks differentially represent information about structure and content, at higher, more complex levels and over longer spatiotemporal scales than the more posterior dorsal and ventral streams. For example, the PM network encodes the structure of events and spatial contexts and participates in the reconstruction of past events and the simulation of future events (Ranganath & Ritchey, 2012), and is also central for discourse comprehension (Martín-Loeches, Casado, Hernández-Tamames, & Álvarez-Linera, 2008).

We can think of the hippocampus, PM, and AT networks as complementary learning systems (Russin, Zolfaghar, Park, Boorman, & O’Reilly, 2021), where the PM and AT networks are specialized for slow, integrative learning leading to highly generalizable representations, complementing the rapid episodic memory supported by the hippocampus. By capturing states of PM and AT network regions at critical moments, the hippocampus is in a unique position to rapidly form memories that combine information about structure and content

that is encountered during a specific period of time (i.e., an *episode*). When encountering a similar item or situation later on, hippocampal pattern completion can lead to reinstatement of previous states in the PM and AT networks, thereby facilitating the generation of a mental model of the current situation in a new or ambiguous context (Franklin et al., 2020).

Although they embody principles that are largely consistent with the ones proposed here, other computational models (Stachenfeld, Botvinick, & Gershman, 2017; Whittington et al., 2020) have placed more emphasis on the role of the hexagonally-tiled *grid cells* in the MEC, which are thought to represent the topological structure of spatial environments, and hippocampal *place cells*, which are thought to encode specific locations within an environment. These models suggest that the MEC and hippocampus can represent any kind of task in a 2D state-space, consistent with fMRI studies showing MEC activity patterns characteristic of grid cells during the performance of novel tasks (Constantinescu et al., 2016). Notably, however, these and other studies of humans suggest that grid cells and grid cell-like fMRI activity patterns are evident throughout the PM network and in the medial prefrontal cortex (Jacobs et al., 2013).

Further work is needed to disentangle the unique contributions of the MEC relative to earlier parietal-lobe pathways in structure representation. One possibility is that the use of structural knowledge to guide flexible behavior is driven by the PM network, and the MEC plays an important role in the use of episodic memory to support rapid structure learning by the PM network. For instance, it is possible that the MEC grid cell network might perform an initial step of *pattern separation* on inputs from the PM network, moving them apart in representational space and thus making them easier to learn without interference (Frankland et al, in prep). This is consistent with results from lesion studies suggesting that hippocampal or MEC lesions can significantly impair spatial learning, while generally sparing navigation in familiar environments (Hales et al., 2014; Kolarik et al., 2016).

Conclusions

There is good reason to believe that the architecture of the neocortex is optimized to support systematic and generative behavior by bifurcating sensory information into separate structure and content pathways along the dorsal and ventral streams. The hierarchically organized pathways through the parietal lobe generate different levels of representations that encode structural relationships reliably associated with different actions and events. Such representations capture the general “logic” of actions and events: how things move, transform, and interact over space and time, in a way that can be readily applied to novel content. Significant work remains to be done in order to establish the nature of structure representations. By understanding the evolution of learning and dynamics across interacting brain networks, we can make significant progress toward understanding the nature of the representations that give rise to our uniquely human intelligence.

Supported by:

We thank Jon Cohen, Alex Petrov, Erie Boorman, Ignacio Saez, Taylor Webb, Stephen Frankland, James Antony, John Rohrlich, Mary Zolfaghar, and members of the CCN, DML, and Cohen labs for discussion and comments.

ONR grants N00014-20-1-2578, N00014-19-1-2684/ N00014-18-1-2116, N00014-18-C-2067, N00014-17-1-2961, N00014-15-1-0033.

ONR grants N00014-20-1-2578, N00014-19-1-2684/ N00014-18-1-2116, N00014-18-C-2067, N00014-17-1-2961, N00014-15-1-0033.

References

- Akyürek E, & Andreas J (2021, June). Lexicon Learning for Few-Shot Neural Sequence Modeling. arXiv:2106.03993 [cs]. Retrieved 2021-08-04, from <http://arxiv.org/abs/2106.03993>
- Binder JR, & Desai RH (2011, November). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15(11), 527–536. doi: 10.1016/j.tics.2011.10.001 [PubMed: 22001867]
- Boylan C, Trueswell JC, & Thompson-Schill SL (2015, November). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, 78, 130–141. doi: 10.1016/j.neuropsychologia.2015.10.007 [PubMed: 26454087]
- Cavanagh P, Hunt AR, Afraz A, & Rolfs M (2010, April). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, 14(4), 147–153. doi: 10.1016/j.tics.2010.01.007 [PubMed: 20189870]
- Constantinescu AO, O'Reilly JX, & Behrens TEJ (2016, June). Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292), 1464–1468. doi: 10.1126/science.aaf0941 [PubMed: 27313047]
- Fodor JA, & Pylyshyn ZW (1988, April). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1–2), 3–71. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2450716> [PubMed: 2450716]
- Frankland SM, & Greene JD (2020, January). Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annual Review of Psychology*, 71(1), 273–303. doi: 10.1146/annurev-psych-122216-011829
- Franklin NT, Norman KA, Ranganath C, Zacks JM, & Gershman SJ (2020, April). Structured Event Memory: A neuro-symbolic model of event cognition. *Psychological Review*, 127(3), 327–361. doi: 10.1037/rev0000177 [PubMed: 32223284]
- Gentner D, & Holyoak KJ (1997, January). Reasoning and learning by analogy: Introduction. *American Psychologist*, 52, 32–34. [PubMed: 9017930]
- Goodale MA, & Milner AD (1992, January). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25. [PubMed: 1374953]
- Hales JB, Schlesiger MI, Leutgeb JK, Squire LR, Leutgeb S, & Clark RE (2014, November). Medial Entorhinal Cortex Lesions Only Partially Disrupt Hippocampal Place Cells and Hippocampus-Dependent Place Memory. *Cell Reports*, 9(3), 893–901. doi: 10.1016/j.celrep.2014.10.009 [PubMed: 25437546]
- Hayes TR, & Henderson JM (2019, October). Scene semantics involuntarily guide attention during visual search. *Psychonomic Bulletin & Review*, 26(5), 1683–1689. doi: 10.3758/s13423-019-01642 [PubMed: 31342407]
- Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei X-X, ... Kahana MJ. (2013, September). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, 16(9), 1188–1190. doi: 10.1038/nn.3466 [PubMed: 23912946]
- Kolarik BS, Shahlaie K, Hassan A, Borders AA, Kaufman KC, Gurkoff G, ... Ekstrom AD. (2016, January). Impairments in precision, rather than spatial strategy, characterize performance on the virtual Morris Water Maze: A case study. *Neuropsychologia*, 80, 90–101. doi: 10.1016/j.neuropsychologia.2015.11.013 [PubMed: 26593960]
- Kravitz DJ, Saleem KS, Baker CI, & Mishkin M (2011, April). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230. doi: 10.1038/nrn3008 [PubMed: 21415848] Integrated review of three parietal lobe pathways supporting looking, reaching, and navigating.

- Lake BM, & Baroni M (2017, October). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. arXiv:1711.00350 [cs]. Retrieved 2019-09-23, from <http://arxiv.org/abs/1711.00350>
- Landau B, & Jackendoff R (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2), 217–265. doi: 10.1017/S0140525X00029733
- Marcus G (2018, January). Deep Learning: A Critical Appraisal. arXiv:1801.00631 [cs, stat]. Retrieved 2021-04-22, from <http://arxiv.org/abs/1801.00631>
- Martín-Loeches M, Casado P, Hernández-Tamames JA, & Álvarez-Linera J (2008, June). Brain activation in discourse comprehension: A 3t fMRI study. *NeuroImage*, 41(2), 614–622. doi: 10.1016/j.neuroimage.2008.02.047 [PubMed: 18394923]
- Orban GA, Van Essen D, & Vanduffel W (2004, January). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8(7), 315–324. [PubMed: 15242691]
- O'Reilly RC, Petrov AA, Cohen JD, Lebiere CJ, Herd SA, & Kriete T (2014). How Limited Systematicity Emerges: A Computational Cognitive Neuroscience Approach. In Calvo IP & Symons J (Eds.), *The architecture of cognition: Rethinking Fodor and Pylyshyn's Systematicity Challenge*. Cambridge, MA: MIT Press.
- O'Reilly RC, Russin JL, Zolfaghar M, & Rohrlach J (2021). Deep predictive learning in neocortex and pulvinar. *Journal of Cognitive Neuroscience*, 33(6), 1158–1196. [PubMed: 34428793]
- Palmer M, Bonial C, & Hwang JD (2016, November). VerbNet: Capturing English verb behavior, meaning and usage. In Chipman SEF (Ed.), *The Oxford Handbook of Cognitive Science* (pp. 315–336). Oxford University Press.
- Park SA, Miller DS, Nili H, Ranganath C, & Boorman ED (2020, July). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*. doi: 10.1016/j.neuron.2020.06.030
- Plaut DC, McClelland JL, Seidenberg MS, & Patterson K (1996, July). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological review*, 103, 56–115. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8650300> [PubMed: 8650300]
- Pylyshyn Z (1989, June). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32(1), 65–97. doi: 10.1016/0010-0277(89)90014-0 [PubMed: 2752706] An original proposal for a structure / content separation mechanism based on attention, that we think may still have currency in neural network models of the brain.
- Ranganath C, & Ritchey M (2012, October). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, 13(10), 713–726. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22992647> [PubMed: 22992647]
- Russin JL, Jo J, O'Reilly RC, & Bengio Y (2020). Systematicity in a Recurrent Neural Network by Factorizing Syntax and Semantics. In *Proceedings for the 42nd Annual Meeting of the Cognitive Science Society* (p. 7).
- Russin JL, Zolfaghar M, Park SA, Boorman E, & O'Reilly RC (2021). Complementary Structure-Learning Neural Networks for Relational Reasoning. In *Proceedings for the 43rd Annual Meeting of the Cognitive Science Society* (p. 7).
- Stachenfeld KL, Botvinick MM, & Gershman SJ (2017, November). The hippocampus as a predictive map. *Nature Neuroscience*, 20(11), 1643–1653. doi: 10.1038/nn.4650 [PubMed: 28967910]
- Summerfield C, Luyckx F, & Sheahan H (2020, January). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, 184, 101717. doi: 10.1016/j.pneurobio.2019.101717 [PubMed: 31669186] Recent synthetic review of a wide range of data and theory supporting the idea that the parietal lobe represents content-independent structure.
- Ungerleider LG, & Mishkin M (1982, January). Two Cortical Visual Systems. In Ingle DJ, Goodale MA, & Mansfield RJW (Eds.), *The Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Webb TW, Sinha I, & Cohen JD (2021, March). Emergent Symbols through Binding in External Memory. *ICLR 2021: Proceedings of the International Conference on Learning Representations*. Retrieved 2021-04-09, from <http://arxiv.org/abs/2012.14601>
- Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, & Behrens TEJ (2020, November). The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 0(0). doi: 10.1016/

j.cell.2020.10.024Computational model implementing structure / content separation in the hippocampal system, with nice illustrations of different forms of structure and how they can support systematic behavior in different domains.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

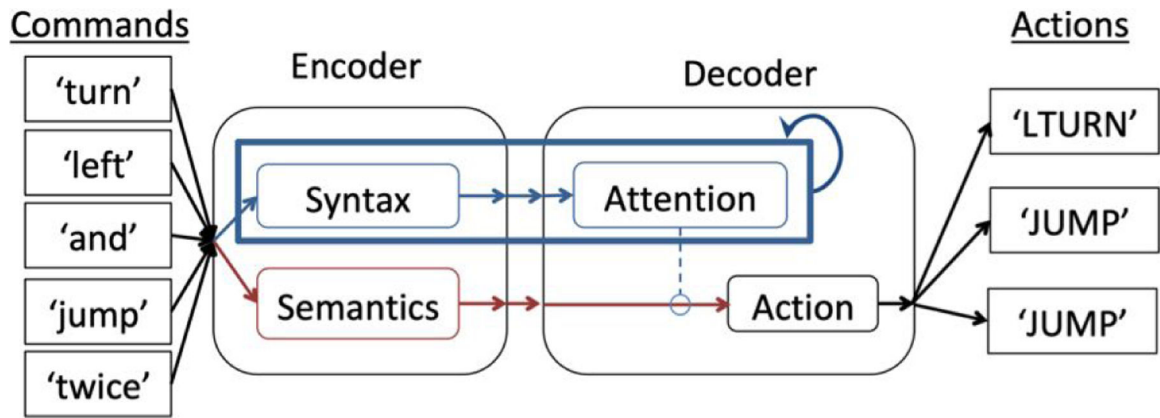


Figure 1:

Diagram depicting a neural network model (Russin et al., 2020) with separate pathways for learning structure (syntax) and content (semantics). The structure pathway was recurrent (as shown by the blue self-connection) and had full access to the temporal ordering of the words in the instructions. However, it could only influence the output actions taken at each time step through an attention mechanism (indicated by the blue dotted line and open circle). The character of the representations in these different pathways facilitated systematic generalization, but was not built in, and emerged due to the architectural constraints imposed by the separation.

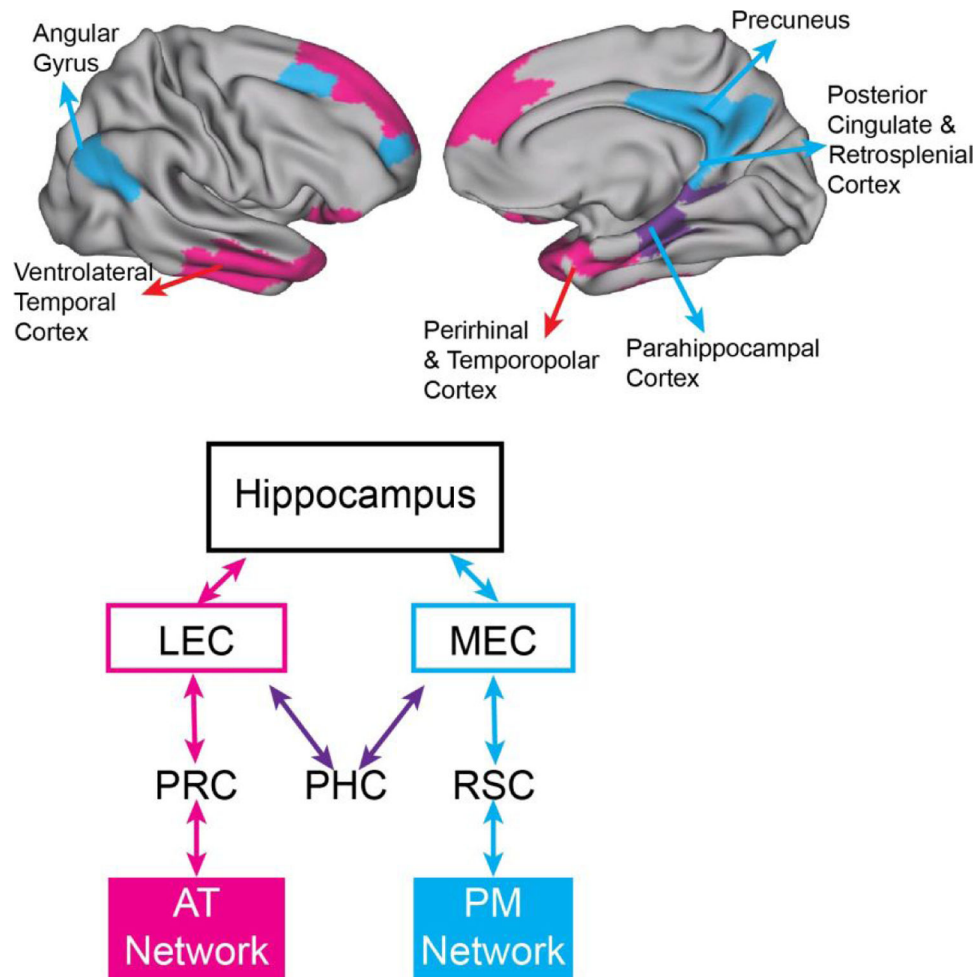


Figure 2: Two cortico-hippocampal networks. Top: Surface renderings depict the neocortical regions in the Posterior Medial (PM) Network (cyan), implicated in schema representation, and Anterior Temporal (AT) Network (magenta), implicated in representations of people and things. Bottom: A diagram schematically depicts the pathways by which information from the AT and PM Networks is directed to parallel pathways in the medial temporal lobes before fully converging in the hippocampus. PRC = Perirhinal cortex; PHC = Parahippocampal cortex; RSC = Retrosplenial cortex; LEC = Lateral entorhinal cortex; MEC = Medial entorhinal cortex.