The Structure of Cognition Across Computational Cognitive Neuroscience

Richard Gao (rigao@ucsd.edu)
Department of Cognitive Science, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093, USA

Dylan Christiano (dchristiano@g.ucla.edu)
Psychiatry and Biobehavioral Science, University of California, Los Angeles
760 Westwood Plaza, Los Angeles, CA, 90024

Thomas Donoghue (tdonoghue@ucsd.edu)
Department of Cognitive Science, University of California, San Diego

Bradley Voytek (bvoytek@ucsd.edu)
Department of Cognitive Science, Halıcıoğlu Data Science Institute, and Neurosciences Graduate Program
University of California, San Diego

Abstract

Computational Cognitive Neuroscience aims to characterize the neural computations underlying behavior. To do so, we must integrate our understanding of cognition across its different subfields: cognitive science, computational neuroscience, cognitive neuroscience, and machine learning. One key challenge is evaluating whether the structure of cognitive processes – their definitions and interrelations – in each subfield is similar. If not, how different are they and how can we measure and ameliorate those differences? To answer these questions, we mined scientific abstracts from conferences representative of subfields to learn field-specific word embeddings of cognitive concepts using Word2Vec. Vector representations are then used to generate hierarchical and 2D visualizations, forming empirical cognitive ontologies for each subfield. We find that robust ontologies, such as clusters representing language-related concepts, are automatically generated from each corpus. While differences between corpora are evident, exploratory analysis with word vectors can perform similarity queries, as well as more complex algebraic queries, e.g., "working memory" without "memory" retrieves "attention". These results demonstrate the utility of automated text-mining and natural language processing in serving as a hypothesis-generating procedure to populate manually-maintained ontologies in cognitive science, as well as suggesting potentially overlooked research opportunities across subfields.

Keywords: ontology, cognitive processes, text-mining, neuroinformatics, meta-analysis

Introduction

The goal of Computational Cognitive Neuroscience (CCN), to quote the conference website directly, "*is to develop computationally defined models of brain information processing [...] that will ultimately have to perform feats of intelligence such as perception, internal modeling and memory of the environment, decision-making, planning, action, and motor control under naturalistic conditions.*" Therefore, CCN represents the intersection of cognitive science, computational neuroscience, cognitive neuroscience, and artificial intelligence in investigating the cognitive (or computational) processes of intelligent systems.

At first glance, this proposed merger appears straightforward (though technically challenging), as the subfields would only need to combine the knowledge they have independently gathered on cognitive processes such as "memory". However, this assumes that these disparate fields all mean (approximately) the same thing when they refer to the term "memory". In practice, however, it is unclear whether "memory" means the same thing to a cognitive scientist as it does to a computational neuroscientist. This merging problem is therefore not simply a task of connecting labelled pieces of information from different fields, but necessarily involves actively mapping between terms and concepts across disciplines, and creating conceptual alignment across the terminology. Here we ask whether, aside from surveying individual scientists, we can ask such questions about conceptual alignment between subfield in an empirical and quantitative way.

The Structure of Cognition - Cognitive Ontology

In general, scientific models can be thought of as relationships between concepts that make up a framework for understanding the physical world – an ontology. These concepts are often born from folk intuition, and are iteratively refined through empirical testing. As a classic example, the intuition that everyday objects are made up of elementary substances evolved from earth, water, air, fire, and aether to the model of atomic elements we know today.

Similarly, cognitive processes can be thought of as abstractions to overlapping aspects of different behaviors. An important milestone for CCN is to construct and refine these conceptual models, as well as filling out the relationship between them. Hence, one way to define meaning is to examine the relationship between concepts, e.g., to ask: where is "perception" embedded within the entire space of cognitive processes?

In this work, we define an "empirical cognitive ontology" as the set of cognitive processes and their relationships as they exist in current scientific literature. Specifically, if processes X and Y are often studied and communicated in conjunction – such as is the case for attention and working memory – then they are "close" to each other within the latent space of cognitive processes. Importantly, this definition does not speak to the existence of some Platonic structure of cognition, only of what exists within scientific literatures.

Previous Works on Cognitive Ontology-Mapping

The problem of mapping cognitive ontologies has been previously investigated. Notably, Poldrack and colleagues (2011) started a monumental effort in charting the ontological space of cognitive processes, as well as their related experimental tasks and disease correlates, aptly named the Cognitive Atlas. These authors hand-crafted hundreds of cognition-relevant terms and their relations with each other, and invited researchers to contribute to documenting new relations. While quality-controlled, curating these processes by hand is ultimately subjective, relies on extensive manual effort, and must match the speed at which new evidence linking old processes is published.

Recent efforts have leveraged more sophisticated and automated computational techniques towards a similar goal, on empirical data and meta-analysis of literature. Eisenberg et al. (2018) surveyed over 500 participants with a battery of psychological tests related to self-regulation and found latent factors relating to a smaller number of internal cognitive processes. Yarkoni et al. (2011) created Neurosynth as a meta-analysis of fMRI studies, providing voxel-level identification of the neural support of cognition. Text-mining has also been applied to article abstracts to find a small number of clusters representative of cognitive "latents" (Alhazmi et al., 2018), or to find associations between neuroscientific concepts, as well as gaps between topics that, statistically, should be more strongly related than they are (Voytek & Voytek, 2012).

Comparing Multiple Ontologies Across CCN

While the above efforts towards creating a cognitive ontology through combining data at a larger scale have been fruitful, they all come from the perspective of cognitive neuroscience and neuroimaging. Many cognitive processes, however, do not share the same practical definition across computational neuroscience and cognitive neuroscience, even if they are called the same thing. For example, "memory" within cognitive science may in fact be more associated with "perception", but to "sleep" in computational neuroscience, from the perspective of existing literature.

In this work, we seek to quantify how different the empirical ontologies are across the different areas of CCN. The importance of evaluating the various empirical structure of cognition is two-fold: first, because scientific findings are published at an ever-increasing rate over the last few decades, automated consolidation of these findings into a condensed ontology that agrees with human curation would serve as a valuable educational tool. Second, by examining the different ontologies extracted from different subdisciplines, we can more efficiently foster productive collaboration by identifying differences between ontologies, as well as avoid the potential cross-talk of referring to entirely different concepts using the same name.

Data & Methods

Text Data from Literature

We collected conference proceedings from Cognitive Science Society (COGSCI), Cognitive Neuroscience Society (CNS), Computational and Systems Neuroscience (COSYNE), and Neural Information Processing Systems (NEURIPS) to represent literature from the various subfields of CCN. Text was either extracted through crawling conference websites directly or manually downloaded and converted from pdf documents. Each corpus consisted of all the accepted abstracts from 2008-2018, ranging between 4800 to 7000 documents (specific years vary for each conference due to formatting idiosyncrasies).

Vector Representation of Concepts and Arithmetic

We trained a separate Word2Vec model using sentence-level representation of each corpus, resulting in a 100-dimensional vector for each unique vocabulary in the corpus. All subsequent analyses were restricted to a subset of 805 cognitive terms that were collected from the "Concepts" page from the Cognitive Atlas. These were used as the main search terms below, and will thus be referred to as "cognitive terms". Vector algebra can be performed on individual word vectors, as well as linear combinations of word vectors, to query for similar and dissimilar concepts.

Automated Creation of Cognitive Ontologies

Using their vector representation, we perform exploratory analysis using dimensionality reduction (t-SNE & UMAP) and hierarchical clustering for visualization of the top-100 most common cognitive terms in each corpus. All data, figures, and code can be found: https://github.com/voytekresearch/IdentityCrisis

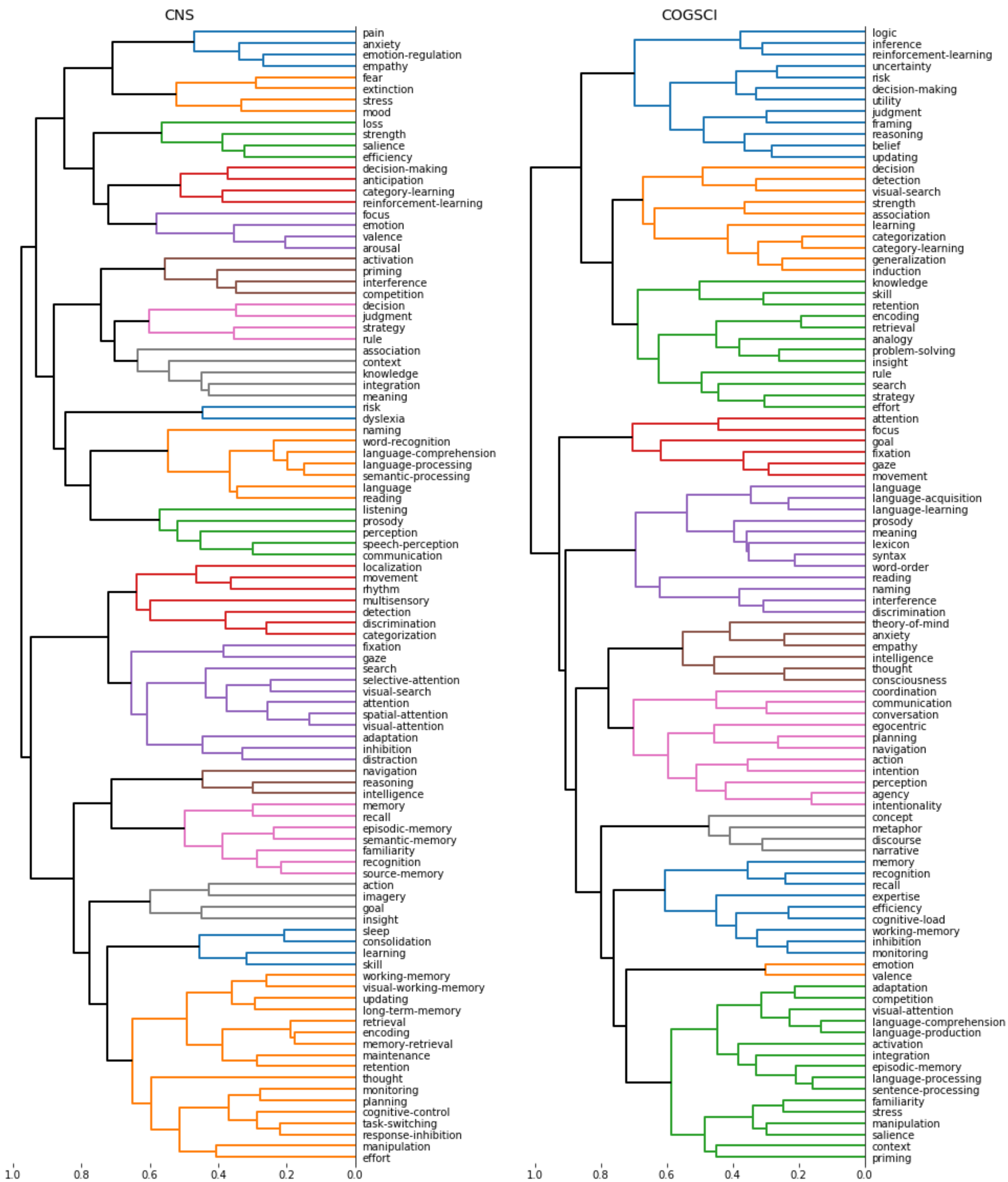

Figure 1: hierarchical clustering results for CNS and COGSCI word embeddings for the top 100 most frequent cognitive terms. Note that, for example, both ontologies contain a self-contained language cluster (left-middle, orange & green; right-middle, purple), while memory-related concepts ("working memory", "maintenance", etc) are clustered near the bottom for COGSCI, but are more spread out for CNS, indicating higher similarity (or co-occurrences) of these concepts in literature.

Results

Concept Similarity and Algebraic Queries

Using the vector representation of cognitive concepts, we can perform similarity and dissimilarity queries. Given the vector for a query term, we can find other vectors with the smallest (similar) and largest (dissimilary) cosine angles. Table 1 shows the 5 most dissimilar concepts to "attention" in each corpus. These terms roughly represent the concepts that co-occur the least with the query term. Interestingly, "risk" and "decision" come up in both CNS and NeurIPS. This presents an untapped opportunity to jointly investigate attentional and decision mechanisms in biological and artificial agents.

Table 1: Top 5 most dissimilar concepts to '**attention**'. Note the occurrence of '**risk**' and '**decision**' in the CNS and NeurIPS corpora, highlighting a potential gap in research linking attention and decision-making.

	COGSCI		CNS		COSYNE		NEURIPS
Term	Similarity	Term	Similarity	Term	Similarity	Term	Similarity
logic	0.095934	risk	0.12498	olfactory-perception	0.1151	risk	0.037504
word-order	0.17964	past-tense	0.14287	generalization	0.14257	regret	0.083393
strength	0.232	morphology	0.16206	error-signal	0.15329	utility	0.12122
dream	0.23473	decision	0.1678	reading	0.16827	loss	0.17387
reasoning	0.24044	naming	0.17019	morphology	0.17152	decision	0.21896

Similarity queries can also be performed with linear combinations of vectors. Word vectors preserve semantic relationships through algebraic manipulation, with the canonical example being "king" – "man = "queen" – "woman" in a general corpus. When we query for "working memory" alone, the most similar terms are other memory-related concepts (not shown). However, we can search for concepts similar to working memory outside a shared context with memory by subtracting the vector for "memory" from "working memory" (Table 2).

Table 2: Top 5 similar concepts to **working memory – memory**. Note the prevalence of attention-related concepts, indicating that when working memory is studied independent of "general" memory, it's usually in conjunction with attention.

	COGSCI		CNS		COSYNE		NEURIPS
Term	Similarity	Term	Similarity	Term	Similarity	Term	Similarity
working-memory	0.47509	working-memory	0.45336	working-memory	0.44793	working-memory	0.50255
monitoring	0.35663	sustained-attention	0.39058	decision-making	0.24372	decision-making	0.50106
spatial-ability	0.3414	hyperactivity	0.32376	spatial-attention	0.21647	skill	0.44838
mental-arithmetic	0.31707	cognitive-control	0.31681	proprioception	0.20605	navigation	0.43163
anxiety	0.31031	inhibition	0.29918	listening	0.20407	causal-inference	0.41289

The most related concepts in each corpus appear to be those related to decision-making and attention, potentially reflecting a relationship between these short-timescale processes.

Hierarchical and 2-D Cognitive Ontology

Figure 1 above shows examples of hierarchical clustering for Cognitive Neuroscience and Cognitive Science. First, we note that sensible clusters appear for each corpus. For example, both corpora have a cluster of relatively well-defined language-related terms, indicating that research on language within cognitive neuroscience and cognitive science are relatively self-contained, i.e., do not involve simultaneous investigation of other processes. We also note corpus-specific differences, such as a lack of memory-related cluster for COGSCI, as it appears in CNS. On the other hand, "consciousness" and "theory of mind" make up a cluster in Cognitive Science, indicating the presence of research investigating "higher-level" cognitive processes in ways that do not exist in Cognitive Neuroscience. Due to space constraint, we did not include 2-D visualizations using UMAP and t-SNE; they can be found in the online repository:
https://github.com/voytekresearch/IdentityCrisis/figures/

Conclusion

In summary, we find that 1) vector algebra and cosine similarity can be directly applied to query for related and unrelated concepts; 2) sensible ontologies can be automatically extracted; and 3) we observe similarities and differences between the empirical ontologies of different subfields. These results demonstrate the utility of automated text-mining and semantic analysis in serving as a hypothesis-generating procedure to further populate manually-maintained ontologies in cognitive science, such as the Cognitive Atlas, as well as in suggesting potentially overlooked opportunities across subfields.

Acknowledgement

We would like to thank the Voytek Lab for their discussion and feedback. R.G. is supported by NSERC PGS-D; B.V. is supported by the Whitehall Foundation (2017-12-73) and the National Science Foundation (BCS-1736028).

References

Alhazmi *et al.* (2018) Semantically defined subdomains of functional neuroimaging literature and their corresponding brain regions. *Human Brain Mapping.*

Eisenberg *et al.* (2018) Uncovering mental structures through data-drive ontology. Preprint *http://dx.doi.org/10.31234/osf.io/fvqej*

Poldrack, R. A. *et al.* (2011) The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Front. Neuroinform*.

Voytek, J. B. & Voytek, B. (2012) Automated cognome construction and semi-automated hypothesis generation. *Journal of Neuroscience Methods* **208,** 92–100.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* **8,** 665–670.