# UC Merced

## Proceedings of the Annual Meeting of the Cognitive Science Society

### Title

Additive Analogies Reveal Compositional Structure in Neural Network Weights

### Permalink

https://escholarship.org/uc/item/60w3r1p6

### Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

### Authors

Tenenbaum, Abi
McCoy, R. Thomas

### Publication Date

2025

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

# Additive Analogies Reveal Compositional Structure in Neural Network Weights

**Abi Tenenbaum**
Yale University, New Haven, Connecticut, United States

**R. Thomas McCoy**
Yale University, New Haven, Connecticut, United States

**Abstract**

A central question in cognitive science is how to reconcile connectionist and symbolic models of the mind (e.g., Fodor & Pylyshyn 1988, Smolensky & Legendre 2006). Attempts have been made to bridge these competing schools of thought by showing how compositional structure can emerge in continuous vector representations (e.g., Manning et al. 2020). A key example is Mikolov et al. (2013), who demonstrated that word embeddings learned by a neural network encode semantic structure: subtracting the vector "man" from "king" and adding "woman" approximates "queen" (i.e., king - man + woman ≈ queen). Our work moves up one level of abstraction, from representations to functions. We analyze whether entire networks display emergent compositional structure by treating a trained network as a single vector (obtained by concatenating the network's parameters) encoding its function. We show that these parameter vectors can be recomposed through simple additive analogies to create networks with new functions.



In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).