

## **UC Merced**

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

#### **Title**

Measuring representational similarity across neural networks

#### **Permalink**

<https://escholarship.org/uc/item/5675t040>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

#### **Authors**

Lu, Qihong

Ramadge, Peter

Norman, Kenneth

et al.

#### **Publication Date**

2018

# Measuring representational similarity across neural networks

**Qihong Lu**

Princeton University, Princeton, New Jersey, United States

**Peter Ramadge**

Princeton University, Princeton, New Jersey, United States

**Kenneth Norman**

Princeton University, Princeton, New Jersey, United States

**Uri Hasson**

Princeton University, Princeton, New Jersey, United States

## Abstract

Shared structure in neural responses across people can be obscured because these neural responses sit on different "coordinate systems"; hyperalignment can recover this shared structure by placing different people's brain responses into a common functional space (Chen et al., 2015; Haxby et al., 2011). Here, we apply this framework to understand the hidden representations of neural networks. Different neural networks can represent the same input-output mapping using very different weights. We show that hyperalignment can construct a shared representational space that recovers shared representation structure across neural networks. We formally connect representational similarity analysis and hyperalignment and use simulations to demonstrate the robustness of hyperalignment against several types of transformations that preserve the representation geometry of the network. We also empirically tested our method on some supervised learning benchmarks (CIFAR10, MNIST) for both standard and convolutional networks.