Discovering computational principles in models and brains

Christian Brodbeck
University of Connecticut, Storrs, Connecticut, United States

Sahil Luthra
University of Connecticut, Storrs, Connecticut, United States

Phoebe Gaston
University of Connecticut, Storrs, Connecticut, United States

James Magnuson
University of Connecticut, Storrs, Connecticut, United States

Abstract

A growing toolbox is emerging for linking neuroimaging data to computations supporting human cognition. In representational similarity analysis (RSA), for example, activity patterns over voxels are compared in response to similar vs. dissimilar stimuli. Resulting similarity matrices are compared to similarity matrices based on theoretical principles or computational models. Similarly, complex EEG or MEG time series can be compared to information-theoretic variables, such as stimulus entropy or surprisal, allowing inferences about the sensitivity of neural responses to aspects of signal information over time. A challenge is determining to what degree such analyses can identify hallmarks of specific computations rather than computationally non-specific resonance with inputs. Here, we apply RSA and information-theoretic analyses to one well-characterized model: TRACE. We consider whether these analyses identify known principles underlying TRACE, and whether TRACE exhibits sensitivity to information-theoretic variables similar to that observed in human brains.