Title: The representational division of labor in Helmholtzian brain science

Presenting Author: Alex B Kiefer

Author(s): Alex B Kiefer, CUNY Graduate Center

Abstract:

Hierarchical predictive coding models of perception suppose that perceptual content and consciousness supervene on neuronal representations comprising the latent variables of a generative model of sensory input, encoded in the brain’s “top-down” synaptic connections, while bottom-up processing, driven by a feedforward model, provides an error signal used to update the generative model.

Despite the interest these theories have generated, they are not the only game in town. David Heeger has recently proposed an alternative model, for example, according to which feedforward networks carry the representation, as in classical Marr-style models, and the top-down generative model provides feedback.

Common to both approaches, and to many other contemporary theories, is a set of commitments that cluster around what I call the (emerging) “Helmoltzian consensus” in cognitive neuroscience and machine learning: perceptual representation learning and inference depend essentially on the interplay between generative and traditional feedforward models, however precisely that interplay is cashed out. The first contribution of my paper is to make this broadly shared background commitment explicit.

The second contribution is to argue that, while there is certainly much to choose from among various theories of this type, in terms of formal details and specific neuronal mechanisms for local computation, arguments about the representational division of labor in Helmholtzian cognitive architectures, e.g. assigning a representational function to the generative model and a corrective function to the feedforward model, are unlikely to be fruitful as currently framed.

I argue that it is not only premature but likely mistaken to associate the contents of consciousness with either the generative or feedforward models. Rather, I hypothesize, conscious experience (whatever one’s precise theory of consciousness) is likely to rest on a computational base that appeals essentially to both generative and feedforward networks and their interaction.