How are the statistics of object co-occurrence represented in human visual cortex?

Michael F. Bonner (michafra@sas.upenn.edu)
Department of Psychology, University of Pennsylvania, 3710 Hamilton Walk, Philadelphia, PA 19104, USA

Russell A. Epstein (epstein@psych.upenn.edu)
Department of Psychology, University of Pennsylvania, 3710 Hamilton Walk, Philadelphia, PA 19104, USA

Abstract:

We set out to identify cortical representations of contextual associations between objects. First we used a large database of densely labeled images to model the natural statistics of object co-occurrence. We then applied this co-occurrence model to fMRI data collected while subjects viewed images of isolated, individual objects. Using voxelwise encoding models, we tested whether object co-occurrence statistics could be used to predict the stimulus-evoked responses of visual cortex. We found that multiple regions of high-level visual cortex reliably encode the statistical associations between objects in their natural contexts, with prominent effects in scene-selective visual regions.

Keywords: fMRI; vision; scenes; natural statistics; co-occurrence; encoding model; objects; memory

Many objects have a natural place in the world—a location or context where they can typically be found. For example, tea kettles and stoves are often found in kitchens, while fire hydrants and traffic lights are found on city sidewalks. This type of contextual knowledge can help people identify where they are in the world (e.g., "I'm in the kitchen") and what other objects they might encounter. A key issue for characterizing how contextual knowledge is represented in the human brain is understanding the statistical structure of object co-occurrence in the natural environment. Here we used a technique adapted from computational linguistics to identify the latent statistical structure of object co-occurrence in a set of 22,000 natural images in which every object was manually labeled. This quantitative model generates a set of multidimensional embeddings that capture information about the contexts in which objects typically occur. We used these embeddings in an fMRI experiment to examine the cortical representation of contextual knowledge during natural object perception.

In the scanner, subjects viewed a large set of objects, shown individually with no scene background, while they performed a perceptual oddball-detection task by responding whenever the stimulus was a visually scrambled object. The stimuli were from 81 object categories (e.g., bicycle, microwave, tree), and there were 10 objects in each category, making for a total of 810 unique images. We used an encoding-modeling approach to identify linear mappings between the contextual embeddings of the objects and their associated cortical responses, and we assessed the performance of these models by determining how well they could predict the fMRI responses in a set of held-out data. Specifically, we applied a 9-fold cross-validation procedure. On each fold of this procedure, a set of voxelwise encoding models were estimated using the images from 72 object categories and then a set of predicted fMRI responses were generated for the stimuli from the other 9 categories. The prediction accuracy of each voxelwise model was quantified by calculating the correlation between the actual and the predicted fMRI responses across all object categories.

Our preliminary findings suggest that high-level visual cortex routinely encodes contextual associations when subjects view individual, isolated objects, even when this information is not present in the immediate perceptual scene. Specifically, we found that the fMRI responses to objects from held-out categories could be reliably predicted by their statistical associations to other commonly co-occurring objects, even when these co-occurring objects were not present in the stimulus. These representations were most prominent in scene-selective regions, which have previously been implicated in the processing of spatial environments and navigational landmarks. Indeed, we observed greater prediction accuracy in scene-selective regions than in object-selective regions, despite the fact that object-selective cortex exhibits a greater overall response to images of individual objects. Altogether, these findings suggest that a key function of high-level visual cortex, and scene-selective regions in particular,

is to link the immediate perceptual stimulus to statistical representations of other stimuli that tend to co-occur with it.