Neural and computational dissociations between objects, scenes, and near-scale reachspaces

Emilie L. Josephs (ejosephs@g.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St
Cambridge, MA 02138, USA

Talia Konkle (tkonkle@fas.harvard.edu)
Department of Psychology, Harvard University, 33 Kirkland St
Cambridge, MA 02138, USA

Abstract:

Space-related processing engages a network of brain regions separate from those engaged in object-related processing. This dissociation has largely been explored using images depicting a navigable scale of space compared to singleton objects. However, this scheme does not account easily for near-scale reachable spaces, which are not navigable but typically contain more than one object. To examine how these views are processed in the brain, human participants underwent functional neuroimaging in which brain responses to near-scale "reachspaces" were compared with responses to scenes and objects. We found evidence for three regions that prefer reachspaces to both scenes and objects: one in ventral visual cortex, one in occipito-parietal cortex, and one in superior parietal cortex. Furthermore, we found that both object- and scene-preferring ROIs were substantially driven by reachspaces, although to an intermediate degree. Finally, we provide computational evidence using deep convolutional neural networks that these three scales of space have separable visual features, potentially accounting for some of the differences in neural representation. Taken together, these results show that perceptual processing of reachspaces may require specialized neural circuits, and may also draw on both object- and scene-based processes.

Keywords: reachspace; object; scene; fMRI; CNN.

There is a large-scale neural division of labor between object-processing and scene-processing networks (Grill-Spector, Kourtzi & Kanwisher, 2001; Epstein & Kanwisher, 1998). However, in naturalistic visual experience, we frequently encounter intermediate-scale spaces, like the tops of desks and kitchen counters. One on hand, these views are like scenes in that they contain spatial layout and multiple objects. On the other hand, these views are unlike scenes in that they are not navigable. These observations raise the question of whether near-scale, reachable space is processed in the same brain networks as far-scale navigable space or single, manipulable objects, or whether they drive their own network of regions.

In previous work, we found initial evidence for a region along the ventral surface that preferred reachspaces to both objects and scenes. (Josephs & Konkle, CCN 2017). Here, we sought to replicate this finding, and further ask whether any additional areas could be detected with increased power.

Additionally, differences in activation along ventral visual cortex have been strongly linked to visual feature differences, and are well-predicted by responses in deep convolutional neural networks (e.g. Güçlü & van Gerven, 2015). Thus, we also examined deep net responses to objects, reachspaces, and scenes, as a proof of concept that such visual feature differences exist between these scales of spaces.

Figure 1: Example reachspaces images. Reachspace views consist of spaces within reach, typically consisting of multiple objects on a spatial layout, which are typically experienced while performing a task with one's hands.

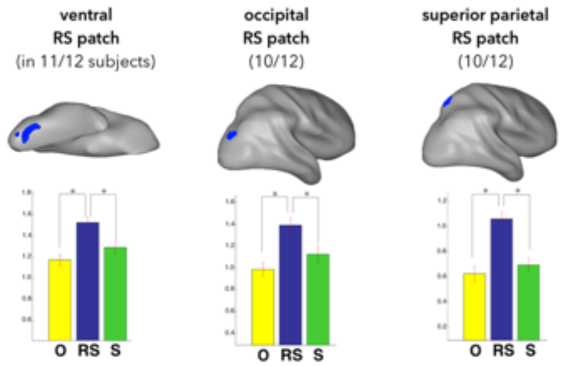

Figure 2: fMRI results, showing three reachspace-preferring ROIs and the activations within them.

Methods

We collected 60 images each of objects, scenes and reachspaces (see Figure 1), with 10 images from each of 6 semantic categories (bar, bathroom, dining room, kitchen, office, and art studio).

Observers (N=12) underwent whole-brain functional neuroimaging in a standard blocked design, with 6s stimulus blocks and intermittent 10s fixation blocks. Participants performed a 1-back repetition detection task. Each run included 2 blocks per condition, and each participant completed 8 runs. In addition to this, participant viewed 2 blocks of localizer runs, which included images of faces, hands, bodies, objects, scenes, multiple objects and white noise.

Results

First, we examined whether there were regions that responded more to reachspaces than both objects and scenes. A conjunction analysis (performed in half of the data) looked for voxels that showed significant activation for both reachspaces > objects and reachspaces > scenes. This analysis revealed three reachspaces-preferring areas (Figure 2). One of the ROIs, hereafter the ventral reachspaces patch (vRSp) matched the ROI previously id entified. The two additional ROIs were found in occipito-parietal and superior parietal cortex, respectively (the occipital reachspace patch, oRSp; and the superior parietal

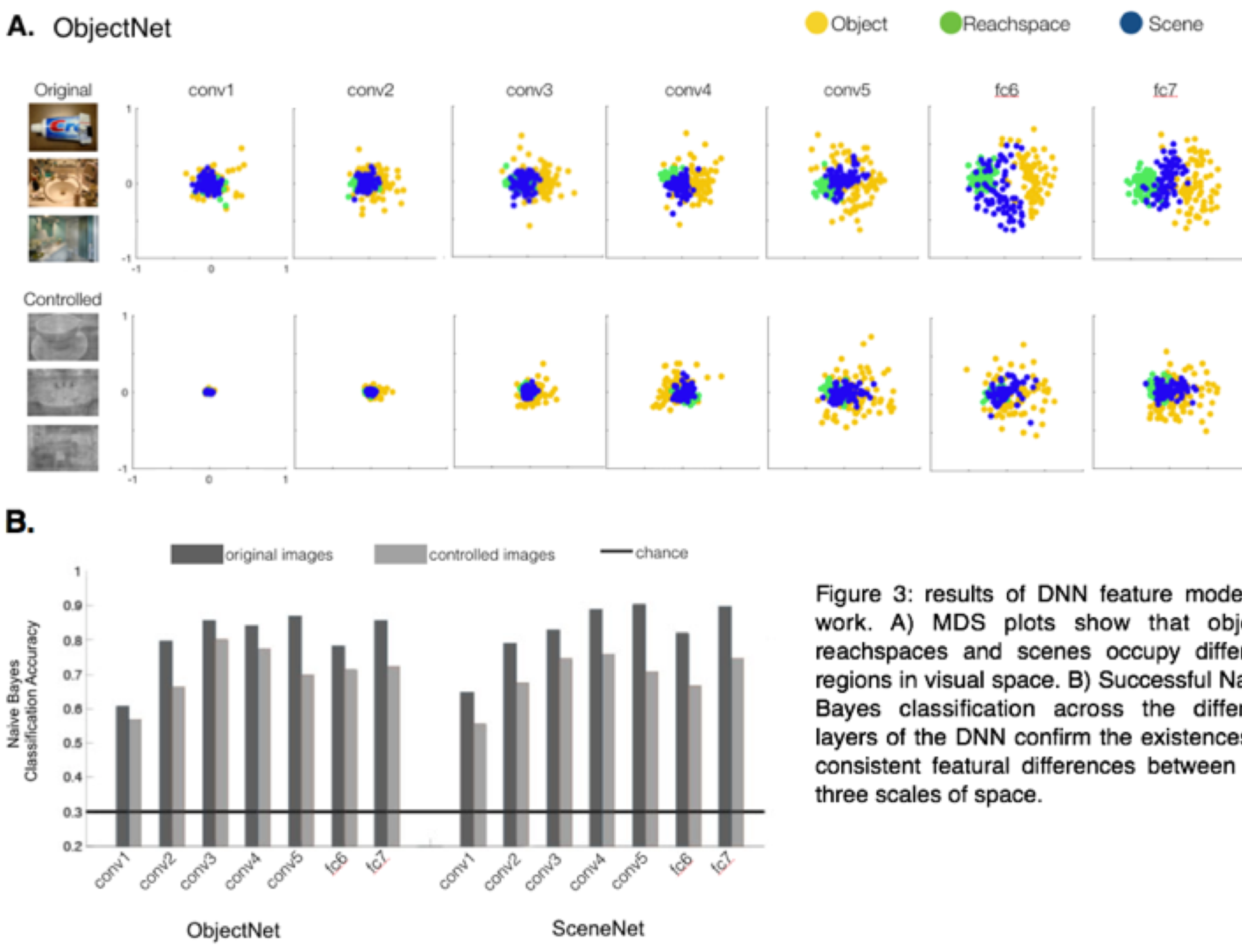

Figure 3: results of DNN feature modeling work. A) MDS plots show that object, reachspaces and scenes occupy different regions in visual space. B) Successful Naïve Bayes classification across the different layers of the DNN confirm the existences of consistent featural differences between the three scales of space.

reachspace patch, spRSp). Activations within these ROIs extracted from the remaining half of the data bore out this preference: reachspaces activated these areas significantly more than both scenes and objects. Additional analyses showed that this preference was general across semantic category, and that these regions were not driven more by animate categories like hands, faces, or bodies. Taken together, these results indicate that RSs preferentially activate certain regions of ventral and dorsal cortex, and suggest that near-space perception may require specialized neural computations from both object and scene-specific processing.

Additionally, we examined the responses to reachspaces in known object- and scene-processing brain regions. Reachspaces elicited substantial activity in both networks, recruiting LOC, pFs, PPA, OPA and RSC, but to a lesser degree than the preferred stimulus type. This suggests that reachspaces may be recruiting both object- and scene-based computations, in addition to the more specialized computations hypothesized above.

Next, we examined whether deep neural networks would also show a dissociation between objects, reachspaces, and scenes. Deep net activations were extracted from two models, one AlexNet architecture trained to do object classification, and the other trained to do scene classification (Zhou et al., 2014). To visualize the structure of these layer representations we used multidimensional scaling (Figure 3), which show that objects, scenes and reachspaces naturally dissociate in intermediate to later layers. This pattern held both on raw images and on the same images equated on low-level features such as luminance, contrast and spatial frequency. These results were quantified using Naive Bayes classification, confirming that objects, reachspaces and scenes are naturally distinguishable on the basis of their visual features, in both object-trained and scene-trained networks. This modeling work raises the possibility the visual feature differences among these stimuli may in part underlie differences in their neural representation.

Discussion

Images of near-scale reachspaces preferentially activate three regions in ventral and dorsal cortex, more than both objects and full-scale scenes. This cortical division of labor suggests that near-space processing may require specialized representations relative to full-scale scenes and isolated objects. The computational modeling results support the possibility that these neural activation differences may be related to differences in the visual feature turning. Taken together, this work shows a division between the neural representations of far-scale, navigationally-relevant space and near-scale, reachable space.

References

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision research, 41(10-11), 1409-1422.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature, 392(6676), 598.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005-10014

B. Zhou, A. Lapedriza, J. Xiao, A. T Recognition using Places Database.” Advances in Neural Information Processing Systems 27 (NIPS), 2014.