Supplemental Materials:
Decoding semantic representations from fNIRS signals

Benjamin D. Zinszer,[a,b,*] Laurie Bayet,[b,c,d] Lauren L. Emberson,[e] Rajeev D.S. Raizada,[a,b] Richard N. Aslin[b]
[a]Brain & Cognitive Sciences Department, University of Rochester, Rochester, NY 14627
[b]Rochester Center for Brain Imaging, University of Rochester, Rochester, NY 14627
[c]Boston Children's Hospital, Boston, MA
[d]Harvard Medical School, Boston, MA
[e]Psychology Department, Princeton University

1 Within- and across-category decoding

The eight stimuli were organized into two categories of objects: four body parts (*hand*, *foot*, *mouth*, and *nose*) and four animals (*dog*, *kitty*, *bunny*, and [teddy] *bear*). These broad category-based differences may contribute to decoding accuracy when choosing between items in different categories. For example, *dog* and *foot* might be distinguished not by their specific meanings, but also by the neural response differences for animals in general compared to the responses for body parts in general. To test whether our own decoding accuracy relied on these comparisons, we divided the pairwise decoding trials into within-category (e.g., *dog* vs. *kitty*) and between-category (e.g., *dog* vs. *foot*) and examined the accuracy of each set of trials.

Within Group 1 (analyzed in Experiment 1), the average leave-one-out between-subjects decoding accuracy for within-category comparisons was 0.69, and between-category accuracy was 0.70. This difference was not significant (paired t-test, p=0.76). In Experiment 2, only the lateral array produced significant between-groups decoding. In our follow-up the lateral array (the only case of significant between-groups decoding) reflected a slight between-category advantage: Within-category accuracy = 0.75; between-category accuracy = 0.81, but across only 3 groups, this difference is minor.

In the semantic model-based decoding, within-category accuracy exceeded between-category accuracy. Across the eight subjects in Group 1, between-category accuracy was higher than within-category accuracy in four of the subjects and vice versa in the other four. Average within- and between-category accuracies did not significantly differ, $t(7) = 1.17$, $p=0.28$, although mean within-category was 0.73 and mean between-category was 0.60. This overall pattern held up at the group-level, where average within-category semantic decoding accuracy was 0.75, and between-category was 0.71 (for the combined arrays).

Thus, overall we do not see evidence to suggest that the reported decoding accuracies relied upon between-category (body part vs. animal) comparisons. Both within-category and between-category comparisons were similarly accurate at the group level, and individual participant accuracy was more suggestive of a within-category advantage.

2 Semantic decoding with the GloVe model

To test the generalizability of our semantic model-based decoding, we repeated the procedures with another popular distributional semantic model, the Global Vectors for Word Representation (GloVe: https://nlp.stanford.edu/projects/glove/; Pennington, Socher, & Manning, 2014). We attempted the decoding analyses reported in this manuscript using GloVe's pre-trained vectors with 300 dimensions (sources: Wikipedia 2014 + Gigaword 5; glove.6B.zip).

Systematic comparisons of distributional semantic models have found that performance differences are relatively small between different models (Murphy, Talukdar, & Mitchell, 2012), so we expected decoding accuracy under GloVe to be similar to that of the COMPOSES model. Further, the 8x8 similarity structures representing GloVe and COMPOSES were highly correlated with each other (r=0.75, p<0.001), suggesting that these models contain roughly the same information about the stimuli and, consequently, would compare similarly with the fNIRS data.

Following the procedures of Experiment 1, GloVe performed slightly worse than COMPOSES for individual participants (mean accuracy 0.64, compared to 0.66 for COMPOSES) and the same for group-level accuracy (0.75 for both models). In Experiment 2, the mean decoding accuracy for GloVe across the three groups slightly improved: 0.74 for both arrays (compare with 0.73), 0.55 in the lateral array (vs. 0.49), and 0.77 in the posterior array (vs. 0.73).

3 References

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. http://doi.org/10.3115/v1/D14-1162