

Influence of verbal labels on concept formation and perception in a deep unsupervised neural network model

OBJECTIVES/RESEARCH QUESTION

Whether language influences perception and thought remains a subject of intense debate. Would the presence or absence of a linguistic label facilitate or hinder the acquisition of new concepts? We here address this question in a neurocomputational model.

METHODS

We used a computational brain model of fronto-occipital (extrasyllvian) and fronto-temporal (perisyllvian) cortex including spiking neurons. With Hebbian learning, the network was trained to associate word forms (phonological patterns, or “labels”) in perisyllvian areas with semantic grounding information (sensory-motor patterns, or “percepts”) in extrasyllvian areas. To study the effects of labels on the network’s ability to spontaneously develop distinct semantic representations from the multiple perceptual instances of a concept, we modelled each to-be-learned concept as a triplet of partly overlapping percepts and trained the model under two conditions: each instance of a perceptual triplet (patterns in extrasyllvian areas) was repeatedly paired with patterns in perisyllvian areas consisting of either (1) a corresponding word form (label condition), or (2) white noise (no-label condition).

To quantify the emergence of neuronal representations for the conceptually-related percepts, we measured the dissimilarity (Euclidean distance) of neuronal activation vectors during perceptual stimulation. Category learning performance was measured as the difference between within- and between-concept dissimilarity values (DissimDiff) of perceptual activation patterns.

RESULTS

The presence or absence of a linguistic label had a significant main effect on category learning ($F=2476$, $p<0.0001$, DissimDiff with labels $m=0.92$, $SD=0.32$; no-labels $m=0.36$, $SD=0.21$). DissimDiff values were also significantly larger in areas most important for semantic processing, so-called semantic-hubs, than in sensorimotor areas (main effect of centrality, $F=2535$, $p<0.0001$). Finally, a significant interaction between centrality and label type ($F=711$, $p<0.0001$) revealed that the label-related learning advantage was most pronounced in semantic hubs.

CONCLUSION

These results suggest that providing a referential verbal label during the acquisition of a new concept significantly improves the cortex’ ability to develop distinct semantic-category representations from partly-overlapping (and non-overlapping) perceptual instances. Crucially, this effect is most pronounced in higher-order semantic-hub areas of the network. In sum, our results provide the first neurocomputational evidence for a “Whorfian” effect of language on perception and concept formation.