NeuroImage 155 (2017) 422–436

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Evidence for similar patterns of neural activity elicited by picture- and word-based representations of natural scenes

Manoj Kumar[a,c,*], Kara D. Federmeier[a,b,c], Li Fei-Fei[d], Diane M. Beck[a,b,c]

[a] Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[b] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA
[c] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
[d] Department of Computer Science, Stanford University, Stanford, CA, USA

A R T I C L E I N F O

Keywords:
Natural scenes
Semantics
Pictures
Words
MVPA
fMRI

A B S T R A C T

A long-standing core question in cognitive science is whether different modalities and representation types (pictures, words, sounds, etc.) access a common store of semantic information. Although different input types have been shown to activate a shared network of brain regions, this does not necessitate that there is a common representation, as the neurons in these regions could still differentially process the different modalities. However, multi-voxel pattern analysis can be used to assess whether, e.g., pictures and words evoke a similar pattern of activity, such that the patterns that separate categories in one modality transfer to the other. Prior work using this method has found support for a common code, but has two limitations: they have either only examined disparate categories (e.g. animals vs. tools) that are known to activate different brain regions, raising the possibility that the pattern separation and inferred similarity reflects only large scale differences between the categories or they have been limited to individual object representations. By using natural scene categories, we not only extend the current literature on cross-modal representations beyond objects, but also, because natural scene categories activate a common set of brain regions, we identify a more fine-grained (i.e. higher spatial resolution) common representation. Specifically, we studied picture- and word-based representations of natural scene stimuli from four different categories: beaches, cities, highways, and mountains. Participants passively viewed blocks of either phrases (e.g. "sandy beach") describing scenes or photographs from those same scene categories. To determine whether the phrases and pictures evoke a common code, we asked whether a classifier trained on one stimulus type (e.g. phrase stimuli) would transfer (i.e. cross-decode) to the other stimulus type (e.g. picture stimuli). The analysis revealed cross-decoding in the occipitotemporal, posterior parietal and frontal cortices. This similarity of neural activity patterns across the two input types, for categories that co-activate local brain regions, provides strong evidence of a common semantic code for pictures and words in the brain.

Introduction

Seeing a furry, four legged animal with a wagging tail, hearing the sound of barking, and reading the word “dog” all evoke a (subjectively) common concept in our minds. What neural processes allow this common concept to emerge from processing that is initially modality and stimulus specific? A long-standing question is whether a common concept arises because these different stimuli all ultimately access the same representation – that is, elicit the same pattern of neural activity. In other words, is there a “common code” for semantic information in the brain that can be accessed from multiple modalities and stimulus types?

Before addressing the possibility of a common code, researchers needed to identify areas involved in representing conceptual information. Initially, univariate fMRI methods were used to find candidate brain regions important for conceptual/semantic processing. For example, researchers contrasted activity evoked by real words and pseudowords (which are perceptually like real words but lack learned semantics) or strings of consonants. This literature uncovered a distributed network of brain regions involved in semantic processing, including regions of lateral and ventral temporal cortex anterior to visual associative regions, the angular gyrus, the left inferior frontal gyrus, left dorso-medial prefrontal cortex, left ventro-medial prefrontal cortex, and the posterior cingulate gyrus (see review by Binder et al.

* Corresponding author at: Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 Matthews Avenue, IL, USA.
E-mail address: mkumar9@illinois.edu (M. Kumar).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.037
Received 14 September 2016; Accepted 18 March 2017
Available online 24 March 2017

(2009). Other studies have used pictorial stimuli (Chao et al., 1999; see review by Martin (2007)) and found a similarly distributed network of brain regions. A few studies have used paradigms with both types of stimuli to find brain areas that respond to both pictures and words when these are processed for semantics, getting slightly closer to the search for a common semantic code. Repetition suppression has been shown in the left fusiform region for both pictures and words (Kherif et al., 2010), and regions similar to the semantic network discussed above are activated by both pictures and words (Vandenberghe et al., 1996). Collectively, this work has revealed that there are a number of brain regions that are activated during the semantic analysis of words and pictures, and, based on these studies, some have proposed a model of semantic representation for concrete objects that is distributed across multiple regions, including sensory and motor systems (Martin, 2007; Pulvermüller and Fadiga, 2010).

Although these studies point to a distributed “common store” for semantic information, they are not sufficient to demonstrate the existence of a common semantic code. It is possible that the same brain areas become active when meaning is extracted from multiple input modalities and/or types, but that these brain regions nonetheless process each differently – for example, using different subpopulations of neurons for each stimulus type. Thus, evidence for a common code in a particular region requires not only finding areas of common activation but also showing that different input modalities evoke similar representations, or shared patterns of activity, within those areas. To obtain this kind of evidence, the literature has turned to multi-voxel pattern analysis (MVPA; see review by Kaplan et al. (2015)).

MVPA affords the ability to move beyond the extant univariate-based evidence supporting a common store by asking whether words and pictures evoke similar patterns of activation. For example, one can train a classifier on the pattern of activity from one type of stimulus (e.g., pictures) and attempt to then classify the pattern of activity elicited by a different stimulus type (e.g., words). We refer to this cross-modal training and testing of a classifier as “cross-decoding”. Such studies have been performed using a variety of modalities (words and pictures: Fairhall and Caramazza, 2013; Shinkareva et al., 2011; pictures, written words, spoken words and natural sounds: Simanova et al., 2014). A few other studies have used Representational Similarity Analysis (RSA; Nili et al., 2014), a technique that uses distances between vectors (built from semantic feature lists or from the BOLD signal) to determine similarities between categories, in order to assess similarity in semantic representations across modalities (auditory words and pictures: Devereux et al., 2003; Liuzzi et al., 2015; written words and pictures: Bruffaerts et al., 2013). From these studies, cross-modal effects have been primarily detected in the left hemisphere: in the precuneus (lPrecu), posterior middle temporal gyrus (pMTG), inferior parietal sulcus (lIPS), precentral gyrus (lPCG), fusiform gyrus (lFG) and the inferior temporal gyrus (lITG).

The use of MVPA and RSA methods, then, have provided evidence that within the distributed semantic network, there are commonalities in the patterns of activation that are elicited by similar concepts across different forms of representation. There are two limitations of the extant literature on cross-modal representations utilizing MVPA methods. First, in comparison to MVPA studies in a single modality that explore fine-grained object categories (e.g. Kriegeskorte et al., 2007; Eger, Ashburner, Haynes, Dolan, and Rees, 2008; Borghesani et al., 2016), studies on cross-modal representations have tended to use stimulus sets that varied across important semantic dimensions, such as animacy, size, and function (although see Bruffaerts et al. (2013) for a notable exception). Because of the substantive differences in their functional and motor affordances, some of these categories (e.g., tools and dwellings) activate clearly separate brain structures: e.g., dorsal motor regions in the case of tools versus ventral medial areas, such as the parahippocampal cortex, in the case of dwellings. In these cases, then, successful cross-decoding may reflect representational similarity at a fairly coarse level; that is, successful cross-decoding can reflect the fact that the objects activate very different regions of cortex. A more stringent measure of a common code in the brain would be to show representational similarity across categories that activate common brain. The few studies that have included fine-grained category distinctions (Fairhall and Caramazza, 2013; Bruffaerts et al., 2013) have been limited to individual objects as opposed to large-scale navigable natural scenes, which brings us to the second limitation of this literature. If we are interested in identifying cross-modal representations, such representations should extend beyond the object domain. Here, therefore, we sought to extend the cross-modal literature to natural scene categories, using four outdoor scene categories (beaches, cities, highways and mountains) known to activate very similar regions of cortex (Walther et al., 2009), making cross-decoding in those regions non trivial. Here, cross-decoding of category membership across representation type (pictures and words) must occur in a higher dimensional space (i.e. at a higher resolution) than for stimulus sets containing categories that show markedly different levels of activation across different brain regions. In other words, successful cross-decoding – training classifiers on one modality and testing on another – among these categories would necessarily imply locally similar neural patterns (i.e. within a restricted region of interest) between pictures and words.

Thus, in the present experiment, participants were scanned while they viewed full color photographs of real world scenes and, extending prior work that has mostly used single words (nouns), read two word phrases that described those categories of natural scenes (e.g. 'beautiful seashore'). By varying the specific noun that was used (e.g., seashore, beach, seaside) and pairing these with a range of adjectives (e.g., beautiful, humid, sandy), we provided a richer semantic stimulus while minimizing adaptation effects that might arise through simple repetition of just the category word (e.g. 'beach'). To test for evidence of a common semantic code (here, across pictures and words) and, more generally, to elucidate the semantic network involved in understanding natural scenes, we performed a cross-decoding analysis through the entire brain using a whole brain searchlight (Kriegeskorte et al., 2006). If we can successfully cross-decode from pictures to words and words to pictures, this would show that the category representations accessed from the two modalities is locally similar – thus better supporting the existence of a common code.

Experimental methods

Participants

Nine subjects (5 females and 4 males; two of the subjects were authors on the paper) participated in the study, which was approved by the Institutional Review Board of the University of Illinois. A tenth subject was dropped prior to analysis because his vision in the scanner had been uncorrected. All participants were in good health, with no past history of psychiatric or neurological diseases, and all gave their written, informed consent. The nine included subjects had normal or corrected-to-normal vision.

Visual stimuli and experimental design

Scene stimuli consisted of 64 distinct color images from each of four categories: beaches, cities, highways, and mountains, using images drawn from a similar set as Walther et al. (2009), which were downloaded from the Internet. Photographs were chosen to capture the high variability within each scene category.

Word stimuli consisted of 64 two-word phrases in each of the four categories. The first word in each phrase was an adjective and the second word was a noun (see Appendix A1). Adjectives appropriate to each category were chosen. The adjectives were matched for word length and word log-frequency across all the categories using the celex database (Baayen et al., 1993). We chose three synonyms for the nouns in each category to make the phrases different and more engaging across the trials (beach category: beach, seaside and seashore; city category: city, town and