January 2011

Commonality of neural representations of words and pictures

Svetlana V. Shinkareva
University of South Carolina - Columbia

Vincente L. Malave
University of California, San Diego

Robert A. Mason
Carnegie Mellon University

Tom M. Mitchell
Carnegie Mellon University

Marcel Adam Just
Carnegie Mellon University, just@cmu.edu

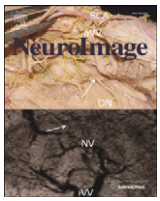
Published In
NeuroImage, 2418-2425.

NeuroImage 54 (2011) 2418–2425

Contents lists available at ScienceDirect

NeuroImage

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



Commonality of neural representations of words and pictures

Svetlana V. Shinkareva [a,*], Vicente L. Malave [b], Robert A. Mason [c], Tom M. Mitchell [d], Marcel Adam Just [c]

[a] Department of Psychology, University of South Carolina, Columbia, SC 29208, USA
[b] Cognitive Science Department, University of California, San Diego, La Jolla, CA 92093, USA
[c] Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[d] Machine Learning Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 22 February 2010
Revised 4 October 2010
Accepted 13 October 2010
Available online 23 October 2010

Keywords:
Brain state
Classification
fMRI
Multi-voxel pattern analysis
Pictures
Words

ABSTRACT

In this work we explore whether the patterns of brain activity associated with thinking about concrete objects are dependent on stimulus presentation format, whether an object is referred to by a written or pictorial form. Multi-voxel pattern analysis methods were applied to brain imaging (fMRI) data to identify the item category associated with brief viewings of each of 10 words (naming 5 tools and 5 dwellings) and, separately, with brief viewings of each of 10 pictures (line drawings) of the objects named by the words. These methods were able to identify the category of the picture the participant was viewing, based on neural activation patterns observed during word-viewing, and identify the category of the word the participant was viewing, based on neural activation patterns observed during picture-viewing, using data from only that participant or only from other participants. These results provide an empirical demonstration of object category identification across stimulus formats and across participants. In addition, we were able to identify the category of the word that the participant was viewing based on the patterns of neural activation generated during word-viewing by that participant or by all other participants. Similarly, we were able to identify with even higher accuracy the category of the picture the participant was viewing, based on the patterns of neural activation demonstrated during picture-viewing by that participant or by all other participants. The brain locations that were important for category identification were similar across participants and were distributed throughout the cortex where various object properties might be neurally represented. These findings indicate consistent triggering of semantic representations using different stimulus formats and suggest the presence of stable, distributed, and identifiable neural states that are common to pictorial and verbal input referring to object categories.

Introduction

The way that concrete objects are represented in the human brain is an important question in cognitive neuroscience. Recently, multi-voxel pattern analysis methods have been applied to fMRI-measured brain activity to associate the brain activity patterns with presented stimuli (see Haynes and Rees (2006), Norman et al. (2006), O'Toole et al. (2007) and Pereira et al. (2009) for reviews of this approach). This approach has the potential to be particularly useful in determining how semantic information about objects is represented in the cerebral cortex. Using multi-voxel pattern analysis, previous studies succeeded in identifying the cognitive states associated with viewing categories of visually depicted objects (Carlson et al., 2003; Cox and Savoy, 2003; Hanson and Halchenko, 2007; Hanson et al., 2004; Haxby et al., 2001; O'Toole et al., 2005; Polyn et al., 2005; Shinkareva et al., 2008), objects presented in the combined word (noun) and picture form (Mitchell et al., 2008), or objects referred to by a written word (Just et al., 2010). In this work we explore whether the patterns of brain activity associated with thinking about concrete objects are dependent on stimulus presentation format, whether an object is referred to by a written or pictorial form.

Multivariate pattern analysis has been successfully used in other cross-modal or cross-task classification applications. For example, to distinguish between activation patterns during mental addition and subtraction, after training the classifier on data from separate experiment requiring saccades to the right or left (Knops et al., 2009), training the classifier on stimuli from sensory domain to separate stimuli in the motor domain, thus illustrating that fMRI signal is similar when perceiving and performing actions (Etzel et al., 2008), or to decode different individual numbers across symbolic and non-symbolic number formats (Eger et al., 2009).

Several studies have postulated that much of the semantic representation of objects is common between written and pictorial stimulus formats, with little functional differentiation (Bright et al., 2004; Chee et al., 2000; Gates and Yoon, 2005; Vandenberghe et al., 1996). We hypothesized that the patterns of brain activity associated

* Corresponding author. Fax: +1 803 777 9558.
E-mail address: shinkareva@sc.edu (S.V. Shinkareva).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.10.042

with thinking of an object when it is referred to by a written word and when it is depicted by a line drawing are similar. Thus it should be possible to use the brain activity patterns extracted during picture-viewing to identify the semantic category of the stimulus during word-reading, and vice versa.

Methods and procedures

Experimental paradigm

Participants viewed words and line drawings of concrete nouns from two semantic categories (*tools* and *dwellings*). There were five exemplars per category: *drill, hammer, screwdriver, pliers,* and *saw*; and *apartment, castle, house, hut* and *igloo*. All stimuli were presented in white against a black background.

Brain activation data for viewing words and for viewing pictures were collected in two separate functional imaging acquisitions during the same scanning session, separated by an experiment from a different study. The order of the two acquisitions was balanced across participants. In each acquisition for words and pictures, there were 6 iterations of presentations of the 10 stimuli, each time in a different random order, for a total of 60 presentations. Participants silently read the words or viewed the line drawings. They were instructed to consistently think about the same properties of the object upon each presentation to encourage repeated neural activation representative of multiple attributes of the object. To ensure that each participant had a consistent set of properties to think about, they individually generated and wrote a set of properties related to each exemplar (such as *cold*, *knights*, and *stone* for *castle*), presented once as a word and once as a picture, in a session prior to the scanning session. The intention of property generation was to foster the retrieval and assessment of salient properties of the object; however, nothing was done to elicit consistency across the two stimulus formats or across participants.

Each stimulus was presented for 3 s, followed by a 7 s rest period, during which the participants were instructed to fixate on an X displayed in the center of the screen. There were six additional presentations of a fixation, 21 s each, distributed across the session to provide a baseline measure of activation. A schematic representation of the paradigm is shown in Fig. 1.

fMRI procedure

Functional images were acquired on a Siemens Allegra 3.0T scanner (Siemens, Erlangen, Germany) at the Brain Imaging Research Center of Carnegie Mellon University and the University of Pittsburgh, using a gradient echo EPI pulse sequence with TR = 1000 ms, TE = 30 ms and a 60° flip angle. Seventeen 5-mm thick oblique-axial slices were imaged with a gap of 1 mm between slices. The acquisition matrix was 64×64 with $3.125 \times 3.125 \times 5$ mm voxels.

fMRI data processing and analysis

Data processing and statistical analysis were performed using Statistical Parametric Mapping software (Wellcome Department of Cognitive Neurology, London, UK). The data were corrected for slice timing, motion, and linear trend and were temporally smoothed with a high-pass filter using a 190 s cutoff. The data were prepared for pattern classification methods by spatial normalization into MNI space using the 12-parameter affine transformation and resampled to $3 \times 3 \times 6$ mm^3 voxels. Voxels outside the brain or absent from at least one participant were excluded from further analysis. The percent signal change (PSC) relative to the fixation condition was computed at each voxel for each stimulus presentation. The mean of the four images acquired within a 4 s window, offset 4 s from the stimulus onset (to account for the delay in hemodynamic response), provided

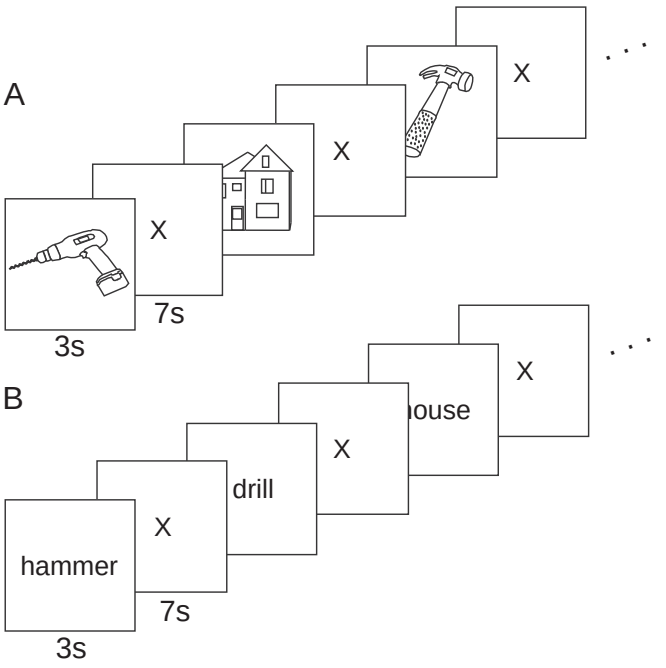


Fig. 1. Schematic representation of the experimental paradigm for the (A) pictures and (B) words experiments. All stimuli were presented in white against a black background.

the main input measure for the classifiers. The mean PSC data for each word or picture presentation were further transformed to have mean zero and variance one, to equate between participants' variation in activation elicited by exemplars.

Pattern classification methods

Classifiers were trained to identify cognitive states associated with viewing words or pictures from the pattern of brain activity (mean PSC) elicited by the same or different stimuli formats. Classifiers (described below) were functions f of the form: f: *mean_PSC* $\rightarrow Y_j$, $j = \{1, 2\}$, where Y_j were the two categories (*tools*, *dwellings*) and where *mean_PSC* was a vector of mean PSC voxel activations, as described above. To evaluate classification performance, trials were divided into training and test sets. Prior to classification, relevant features (voxels) were extracted (selected as described below) to reduce the dimensionality of the data, using data from only the training set for this selection. A classifier was built from the training set, using the selected features. Classification performance was then evaluated by applying the classifier to the left-out test set. Our previous exploration with different data sets has indicated that several feature selection methods and classifiers produce comparable results. Here we employed a single feature selection method and a single classifier for all participants, chosen for simplicity. Thus the choice of feature selection method and a classifier was not optimized for this data set. Other related methods and parameter values may produce comparable outcomes.

Feature selection

Feature selection first identified the voxels whose responses were the most stable over presentations and then selected from among the stable voxels those that best discriminated among items within the training set. In the first step, the 400 most stable voxels were identified, where voxel stability was computed as the average pairwise correlation between 10-object vectors across the five training presentations. In the second step, these 400 voxels were assessed for how discriminating they were, by training a logistic

regression classifier to discriminate among categories based only on the training set (Shinkareva et al., 2008). From among the 400 voxels selected for stability, a subset of 120 voxels (henceforth, diagnostic voxels) was selected based on having the highest (absolute value) regression weights in the logistic regression based on the training data. Previous studies with different data sets indicated reliable classification accuracies for feature sets of 10–400 voxels, varying in size depending on participant. In this work we used the same number (a typical number of features generating high classification accuracies based on other data sets) of voxels for all participants.

Classification

The Gaussian Naïve Bayes (GNB) pooled variance classifier was used (Mitchell, 1997). It is a generative classifier that models the joint distribution of a class Y (either *tools* or *dwellings*) and attributes $\times$ (voxels), and assumes the attributes $X_1,\ldots,X_n$ are conditionally independent given Y. The classification rule is:

$$Y \leftarrow \arg\max_{y_j} P\left(Y = y_j\right) \prod_i^n P\left(X_i | Y = y_j\right), j = 1, 2.$$

Classification results were evaluated using k-fold cross-validation, where one presentation of all exemplars per class was left out for each fold. Hence the identification accuracy was always based only on test data that were disjoint from the training set. For a two-class classification problem with equally frequent classes the chance level is 0.5. The observed accuracy was then compared to the binomial distribution; if the observed accuracy had a p-value of at most 0.05, then the result was considered significant (Pereira et al., 2009).

Cross-participant analysis

Data from all but one participant were used to train a classifier to identify the data from the left-out participant. This process was repeated reiteratively, leaving out each of the participants. Feature selection was conducted by combining the data of all participants except the one left out. A discriminating set of 120 diagnostic voxels was selected on the basis of logistic regression weights, using data only from the training set.

Analyses of a single brain region at a time

To determine the relative information value of various brain regions in identifying object categories, single brain regions that consistently contained voxels used in identification of object categories across participants, without making the assumption that their voxel patterns are spatially matched in a common space, were identified. The full list of anatomical regions was defined using the Anatomical Automatic Labeling (AAL) system (Tzourio-Mazoyer et al., 2002). In addition to existing AAL regions, two parietal regions, left and right IPS, were defined, and three temporal regions, the superior, middle, and inferior temporal gyri, were segmented into anterior, middle, and posterior portions based on planes F and D from the Rademacher scheme, resulting in a total of 71 regions (Rademacher et al., 1992). Since the number of voxels within each anatomical region is considerably smaller compared to the whole brain, within each participant a cross-validated accuracy based on each individual region for that participant was computed using a logistic regression classifier with L2 regularization using all the voxels from that region (although all of the voxels in the region were included in the analysis, the logistic regression assigns greater weights to those voxels that contribute most to the classification). The mean classification accuracy was computed for each region across participants and compared to a binomial distribution. The obtained p-values (computed using a normal approximation) were then compared to the level of significance $\alpha = 0.05$, using the Bonferroni correction to account for multiple comparisons, which is appropriate for a map at an anatomical region level. This analysis was done for category identification across and within stimulus formats, word and picture.

Participants

Twelve right-handed adults (eight female) from the Carnegie Mellon University community participated and gave written informed consent approved by the University of Pittsburgh and Carnegie Mellon University Institutional Review Boards. Six additional participants were excluded from the analysis due to head motion greater than 2.5 mm.

Results

Category identification across stimulus formats within participants

When a classifier was trained for each participant on word data to determine if it was possible to identify object categories based on brain activation data evoked by picture stimuli, the highest classification accuracy obtained for a single participant was 0.97 (that is, correct category identification for 58 out of 60 picture presentations). Reliable ($p<0.05$) classification accuracies were reached for 11 out of 12 participants (filled bars in Fig. 2). The mean classification accuracy for categories of pictures when trained on data generated by words was 0.75 (SD = 0.14).

When a classifier was trained for each participant on picture data to identify the object categories based on brain activation data evoked by word stimuli, the highest classification accuracy obtained for a single participant was 0.97 (that is, correct category identification for 58 out of 60 word presentations). Reliable ($p<0.05$) classification accuracies were reached for six out of 12 participants (filled bars in Fig. 3). The mean classification accuracy for word categories when trained on pictures was 0.66 (SD = 0.17).

In summary, classifiers trained on data from one stimulus format were able to successfully identify object categories in the imaging data from the other stimulus format. The identification was reliable for a

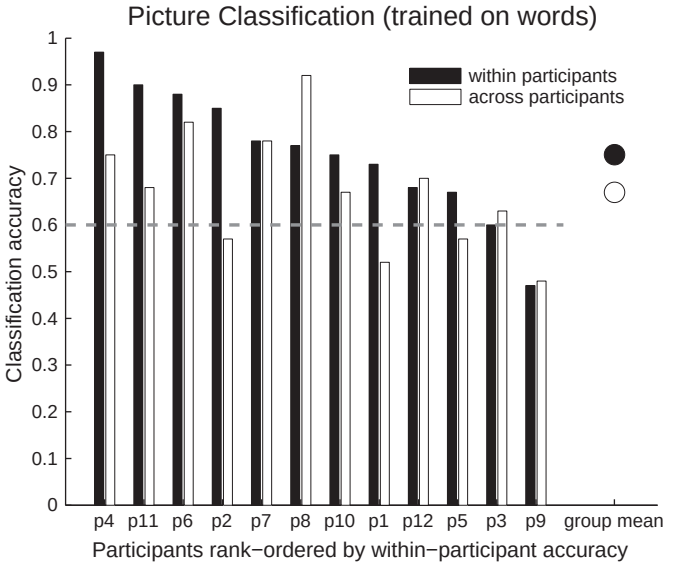


Fig. 2. Classification accuracies across stimulus formats when training on words to identify the category of viewed pictures. Reliable ($p<0.05$) accuracies for identification of category of viewed pictures (filled bars) were reached for 11 participants when training on word data, and reliable ($p<0.05$) accuracies for identification of category of viewed pictures when training on the union of word data from the other participants (unfilled bars) were reached for 8 out of 12 participants. The dashed line indicates the $\alpha = 0.05$ level of significance.

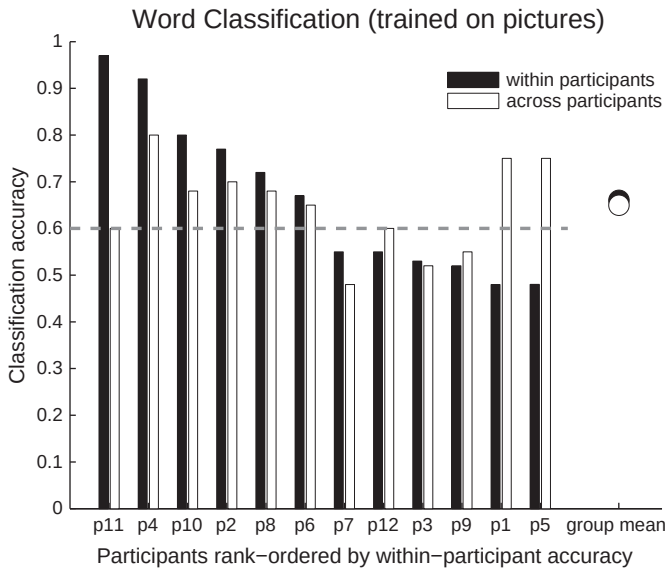


Fig. 3. Classification accuracies across stimulus formats when training on pictures to identify the category of viewed words. Reliable ($p<0.05$) accuracies for identification of category of viewed words (filled bars) were reached for 6 participants when training on pictures data, and reliable ($p<0.05$) accuracies for identification of category of viewed words when training on the union of picture data from the other participants (unfilled bars) were reached for 9 out of 12 participants. The dashed line indicates the $\alpha = 0.05$ level of significance.

greater number of participants when word-generated activation was used to classify picture-generated activation. These findings were obtained when the classification was conducted within each participant.

Category identification across stimulus formats across participants

Data from one stimulus format from all but one participant were used to identify the object category of the other stimulus format presented to the left-out participant. To determine if it was possible to identify the object category of viewed pictures in the left-out participant using brain activation data from other participants, a classifier was trained on the combined word data from all but one participant. This procedure was repeated for each of the participants. The highest classification accuracy obtained for a single participant using this approach was 0.92 (unfilled bars in Fig. 2). Reliable ($p<0.05$) category classification accuracies were achieved in eight participants. The mean classification accuracy was 0.67 (SD = 0.13). Thus it is possible to identify the category of a pictured object for a participant based on the word-generated activation of other people, while maintaining accuracy comparable to classification based on the participant's own data.

The attempt to classify categories also succeeded in the opposite stimulus format comparison, when classifiers were trained on picture activation and then used to identify word categories. A classifier was trained on the union of picture data from all but one participant to determine if it was possible to identify the object category of viewed words in the left-out participant. This procedure was repeated for each of the participants. The highest accuracy obtained for a single participant was 0.80 (unfilled bars in Fig. 3). Reliable ($p<0.05$) category classification accuracies were achieved in nine participants. The mean classification accuracy was 0.65 (SD = 0.10). Thus it was possible to identify the category of an object whose name a participant was reading based solely on the neural signature derived from a set of other participants' activations generated during picture viewing.

To ensure that there was no effect of acquisition order (viewing words or pictures first), a supplementary analysis made cross-participants cross-stimulus format identifications based on only the first of the two conditions (viewing words or pictures first) performed by each participant. The absence of a reliable difference (p-value = 0.6) between the two mean identification accuracies, computed across participants for words and pictures, indicated that the order of presentation of the two stimulus formats did not affect the main conclusions.

These findings indicate that a high degree of commonality exists in the neural activation patterns elicited by the stimulus categories, both across stimulus formats and across participants. This similarity in patterns of brain activation enabled the cognitive state identification that was successfully achieved by the classifiers. A secondary finding is that the classification accuracy for identifying the object category of a picture after training on word data was generally higher than the accuracy for identifying the object category of a word after training on picture data.

Category identification across stimulus formats based on single brain regions

To determine whether cognitive states elicited by input from one stimulus format can be identified based on the activation in a single brain region stimulated by input from the other stimulus format, classifiers were trained using voxels from only one anatomical region (such as left inferior parietal lobule (IPL)) at a time. The accuracies obtained in this ancillary analysis were surprisingly high. For example, in the case of one participant whose classification accuracy was 0.97 for category identification of pictures based on word-generated activation throughout the cortex, the single-region accuracy was 0.83 for the left IPL and 0.78 for the left precentral gyrus. For the same participant, the classification accuracy was 0.92 for category identification of words based on picture-generated activation throughout the cortex, while the single-region accuracy was 0.90 for the left IPL and 0.77 for the left precentral gyrus. The regions that generated reliable accuracies across participants in this single-region category identification across stimulus formats were the fusiform gyrus and precuneus, paracentral lobule and superior parietal lobule (SPL) bilaterally. In the left hemisphere, the superior extrastriate (SES), inferior extrastriate cortex (IES), intraparietal sulcus (IPS), IPL, supplementary motor area (SMA), posterior cingulate, postcentral and precentral gyri, and posterior superior and inferior temporal gyri produced reliable accuracies for cross-participant identification in both classification directions. In addition, the right IPL generated reliable accuracies across participants when training on words to identify picture categories. The ability to identify the object category based on activation generated from the other stimulus format in selected anatomical regions indicates the presence of discriminating semantic information in these regions.

Category identification within word stimulus format only

Although the main focus of this paper is the identification of brain activation patterns across stimulus formats, it is also interesting to examine identification of brain activation patterns for each of the stimulus formats. Of particular interest is the within-word data, which have seldom been examined in this way; a majority of previous attempts at classification of semantic categories have used data generated by picture stimuli. For each participant, a classifier was trained on the activation generated by a subset of the word data, and then tested on an independent subset on the ability to identify which category of word the participant was viewing. The highest classification accuracy in a single participant was 0.93 (correct category identification on 56 out of 60 word presentations). Reliable ($p<0.05$) classification accuracies for identifying the category of the word that a participant was viewing were reached for 11 participants (filled bars in Fig. 4). The mean classification accuracy for word-category identification across the 12 participants was 0.74 (SD = 0.13).

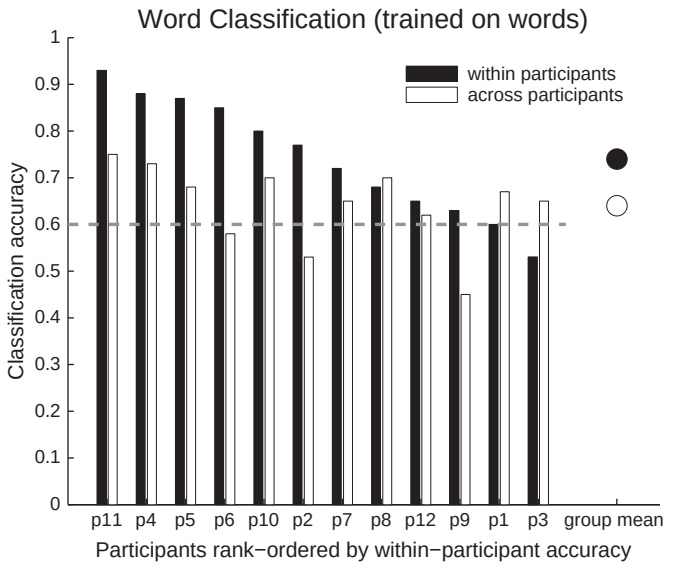


Fig. 4. Classification accuracies for identifying the category of viewed words. Reliable ($p<0.05$) accuracies for identification of category of viewed words (filled bars) were reached for 11 participants, and reliable ($p<0.05$) accuracies for identification of categories of viewed words when training on the union of data from the other participants (unfilled bars) were reached for 9 out of 12 participants. The dashed line indicates the $\alpha = 0.05$ level of significance.

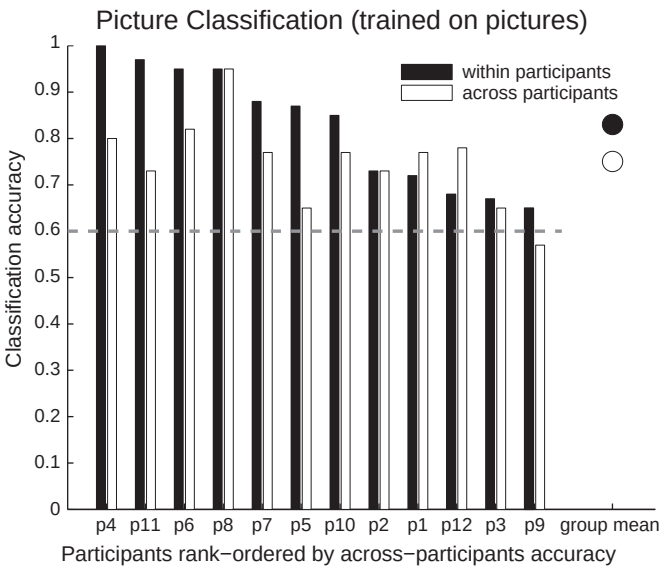


Fig. 5. Classification accuracies for identifying the category of viewed pictures. Reliable ($p<0.05$) accuracies for identification of category of viewed pictures (filled bars) were reached for all participants, and reliable ($p<0.05$) accuracies for identification of category of viewed picture when training on the union of data from the other participants (unfilled bars) were reached for 11 out of 12 participants. The dashed line indicates the $\alpha = 0.05$ level of significance.

The classification of word data was also performed across participants. A classifier was trained on data from 11 of the 12 participants to determine if it was possible to identify word categories in the left-out 12th participant; this procedure was repeated for all participants. The highest classification accuracy obtained was 0.75. Word categories being viewed were reliably ($p<0.05$) identified for 9 out of 12 participants (unfilled bars in Fig. 4). The mean classification accuracy was 0.64 (SD = 0.09). Thus the category of a noun that a person reads and thinks about can be identified solely on the basis of activation patterns obtained from other individuals.

Category identification within picture stimulus format only

For each participant, a classifier was trained to decode which picture category the participant was viewing. Accuracies of at least 0.95 (correct category identification in at least 57 out of 60 picture presentations) were obtained in four of the participants (filled bars in Fig. 5). Reliable ($p<0.05$) classification accuracies were reached for all 12 participants. The mean classification accuracy for picture-category identification across the 12 participants was 0.83 (SD = 0.13).

For cross-participant identification of the category of the presented picture, the highest accuracy obtained for one of the participants was 0.95 (correct identification of the category in 57 out of 60 object presentations) (unfilled bars in Fig. 5). The classifier achieved reliable ($p<0.05$) accuracy in 11 out of 12 participants. The mean accuracy across participants was 0.75 (SD = 0.10). Within stimulus formats, picture activations were classified more accurately than word activations, a result reminiscent of the finding that picture categories are easier to identify after training on words than vice versa.

Locations of voxels used in category identification within stimulus formats

The locations of diagnostic voxels for within word and picture category identification were distributed across the cortex. The similarity of the locations of these diagnostic voxels across participants and across the two stimulus formats is illustrated in Fig. 6, which indicates the locations of the voxels used by within-stimulus format classifiers in three of the participants (the participants with the highest classification accuracies when using a classifier trained on word data to identify categories of pictures).

Single regions that contained information consistently identifying the object category for words and for pictures, as well as across the two stimulus formats in either direction, were located primarily in the left hemisphere (first column in Table 1). In addition, some regions only supported category identification for either word or picture data (Table 1). The anatomically defined regions of interest (based on the AAL scheme) differ greatly in volume. The number of voxels within each anatomically defined region that contributed to the classification differed much less, with there generally being more such voxels in more posterior (visual) regions and fewer such voxels in more anterior association areas.

Discussion and conclusions

Successful identification of an object category presented in one stimulus format based on neural activation patterns elicited by a different stimulus format implies consistent triggering of semantic representations using different stimulus formats. Thus the neural activity captured by the classifiers reflects semantic property representation, and not just perceptual features associated with stimulus formats. It is the first demonstration of the ability to identify a word category on the basis of activation generated by picture stimuli, and vice versa.

The ability to use neural activation patterns to identify object categories across the stimulus formats suggests a commonality of the neural basis of these two types of semantic knowledge. In addition to the commonality of neural activity across words and pictures, there was also evidence of format-specific neural activation. Some of the more anterior regions (e.g., left pars opercularis and left pars triangularis) contained adequate information to identify the category of either words or pictures but did not support identification across stimulus formats. These findings indicate that some small but information-laden part of

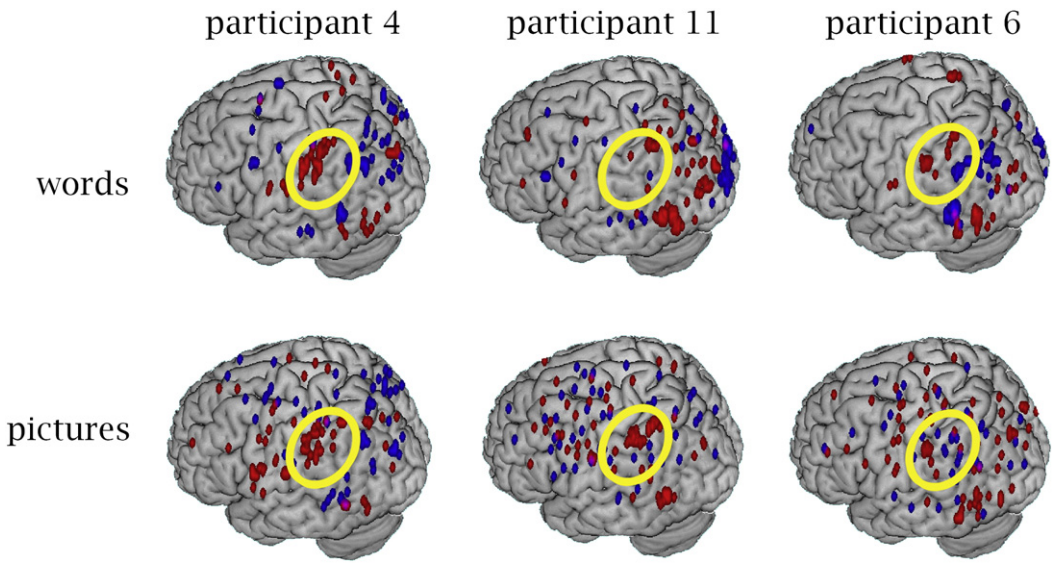


Fig. 6. Similarity across participants and across the two stimulus formats in the locations of diagnostic voxels for word (upper panel) and picture (lower panel) category identification. Diagnostic voxels (union across folds) used by the classifier are shown on the three-dimensional rendering of the T1 MNI single-subject brain (red for the *tools* category and blue for the *dwellings* category). Ellipses highlight the commonality of voxel locations for object identification in LIPL across participants and across the two stimulus formats.

the neural representation of objects is specific to the format of stimulus presentation. Format-specific representation is consistent with patterns of performance in patients with disorders of semantic memory, such as an asymmetry in the *retrieval* of words and pictures in Pick's disease (Hodges and Patterson, 1997), despite of considerable commonality of the neural representation. Brain pathology can selectively damage some biological infrastructure of the retrieval process yet apparently leave the neural representation intact. Thus neural representations are likely to be separable from at least some of the neural structures that are used in retrieving a word that name an object. Taken together, these results are consistent with the proposal of a common semantic representation for the two stimulus formats accompanied by some degree of format-specific differentiation (Bright et al., 2004; Chee et al., 2000; Gallagher et al., 2000; Gates and Yoon, 2005; Vandenberghe et al., 1996).

We propose that words and pictures can give rise to a common neural representation. However, we note that the commonality of neural representations across stimulus formats may also be influenced by the commonality of the task instructions in the two conditions. Regardless of the instructions, some of the cross-modal similarity in neural activity may have followed from automatic processing of the familiar stimuli, uninfluenced by instructions; determining how much of the commonality is due to automatic processing versus strategic recall of object properties is an issue that can be addressed in future research. Furthermore, despite the object properties for the two stimulus formats being generated by the participants in the same session prior to the scanning, these properties differed to some degree between word and picture stimulus formats for most of the participants (mean Dice coefficient across participants was 0.68, SD = 0.23). Even if the deliberate recall is helping the classification, our results still show the semantic encoding is not influenced much by the stimulus form and that our methods detect the locations and patterns of the similar encodings. Future studies may shed further light on this issue by presenting the stimuli in a way that does not permit strategic retrieval schemes (such as presenting items for very short durations) or by measuring EEG or MEG responses that might permit a temporal separation between automatically and strategically activated brain regions. Although it is obvious that there should be a commonality between what neural events are

Table 1
Anatomical regions, selected based on data from 12 participants, that contain information sufficient to decode semantic category. L indicates left hemisphere and R indicates right hemisphere.

Areas/Function	Words, pictures and Cross-format (bidirectional)	Word-specific	Picture-specific
Language	L posterior superior temporal, posterior middle temporal gyri	L pars opercularis L mid-superior temporal gyrus	L pars triangularis
Motor/somatosensory	L postcentral gyrus L precentral gyrus L supplementary motor area		
Visual-spatial	L inferior parietal, supramarginal, angular gyri L intraparietal sulcus L/R superior parietal gyrus, precuneus, paracentral lobule		R intraparietal sulcus
Executive		L middle frontal gyrus R superior frontal gyrus	
Visual	L inferior occipital, lingual gyri L posterior inferior temporal gyrus L cuneus, superior occipital, middle occipital gyri L/R fusiform gyrus		L calcarine fissure L/R cerebellum R inferior occipital, lingual gyri R posterior inferior temporal gyrus
Other	L posterior cingulum		

evoked by a word and a corresponding picture, particularly when the task instructions are the same in the two cases, it is quite another thing to demonstrate the ability to classify thoughts across modalities using the common neural representation.

Semantic organization

The set of brain areas that contain category-specific information about tools and dwellings can be classified into areas that support category identification across stimulus formats and areas that support either word or picture category identification.[1]

Independent of stimulus format, object categories can be accurately decoded from several regions, located primarily in the left hemisphere and distributed throughout the cortex. The ability to identify an object's category based on data from another stimulus format suggests the presence of semantic, non-perceptual content in the object's representation in those regions. The regions that contain sufficient information for category identification across stimulus formats include areas that have been associated with encoding the representation of manipulable objects and objects denoting shelter (Just et al., 2010). These regions include areas in the ventral-visual pathway that have previously been shown to be activated by both word and picture stimuli; these areas are hypothesized to relate to abstract form representations of objects (Bright et al., 2004; Chao et al., 1999; Devlin et al., 2005; Perani et al., 1999; Pietrini et al., 2004). In addition, areas associated with higher-level visuo-spatial processing, language, motor and somatosensory functions contained sufficient information for category identification across stimulus formats. For example, the LIPL, which has been implicated in manipulation knowledge, was previously shown to activate in response to both word and picture stimuli (Boronat et al., 2005).

The common neural representation of words and pictures indicates that they share a feature-based, distributed, conceptual representation marked by several interesting properties. First, the inclusion of motor and sensory properties in the representation of an object (Hoenig et al., 2008; Pulvermuller, 2005; Pulvermuller et al., 2009) provides evidence of "embodied cognition," a theoretical position holding that conceptual representations contain perceptual and motor components corresponding to human interactions with real entities in the physical environment (e.g., Aziz-Zadeh and Damasio (2008); Barsalou (1999); Glenberg (1997)). Second, the inclusion of parietal and prefrontal regions in the representation of an object suggests that a visual imagery network may constitute part of the representation (Mechelli et al., 2004). Alternatively, these areas may be supporting a more abstract representation of object form (Pietrini et al., 2004). It is quite striking that single regions contain, on their own, enough information to decode the presented words and objects. It must be the case that sufficient information for category identification is represented in several different regions, lending a somewhat different interpreting to the notion of a distributed representation. We make no claim that the information in the different regions is equivalent. Further studies may help illuminate the representational content in regions that support category identification across stimulus formats, such as studies using item-repetition priming (Grill-Spector et al., 1999; James et al., 2002; Vuilleumier et al., 2002) or Dynamically Adaptive Imaging (Cusack et al., 2010).

Higher category identification accuracies for pictures than for words

Accuracy identifying the category of a picture a participant was viewing was higher than for identifying the category of a word that a participant was reading, both within and across participants. Higher identification accuracies for picture data may reflect the ability of pictorial stimuli to generate brain activation that is more stable over six presentations than the brain activation generated by word stimuli. The specific visual properties of a picture may more directly or reliably trigger the activation of the same properties of the neural representation of an object than a written word does. From a pattern classification perspective, classifiers are more accurate when tested on cleaner, more consistent data; in this case, identification of the category of an object based on the activation generated by its pictorial representation was more accurate than identification based on the less reliable activation generated by the written representation of the same object.

Within-participant vs. cross-participant category identification

The ability to identify object categories across participants, both within and across the two stimulus formats, reveals a common neural basis for representation of this type of semantic knowledge across people. Despite the individual differences in functional organization and the methodological difficulty of normalizing the morphological differences found among human brains, neural similarities arose in terms of the locations and activation amplitudes of voxels utilized by the classifier to identify the object category of a stimulus. Classification of mental states across individuals has been previously shown for visually depicted objects (Shinkareva et al., 2008), concrete nouns referring to physical objects (Just et al., 2010), lie detection (Davatzikos et al., 2005), attentional tasks (Mourao-Miranda et al., 2005), cognitive tasks (Poldrack et al., 2009), and voxel-by-voxel correspondence across individuals has been demonstrated during movie-watching (Hasson et al., 2004). The current results demonstrate the ability to identify the category of an object viewed as a picture or as a word based on neural activation data from other participants in the other stimulus format.

The category identification accuracies when training classifiers on data from across participants were, on average, lower than when training within participants, indicating that some small portion of the neural representation of the categories is idiosyncratic to individual participants. There is apparently systematic variation within an individual's neural activation (permitting better identification of that individual's cognitive state) that leaves room for individuality in object representation.

For a few participants with low within-participant identification accuracies, the identification accuracy was actually higher in the cross-participant identification (when the classifier was trained on data from all the other participants). For these individuals, signal-averaging over other participants provided a greater benefit to identification than the individual's own idiosyncratic activation pattern. It is intriguing that the neural activity underlying some individuals' thoughts of an object can be more similar to a group norm than to their own previous neural activity. This finding speaks both to the individuality of brain activation patterns as well as to the commonality.

The presented results provide an empirical demonstration of object category identification across stimulus formats and across participants. The experimental settings were restricted to 10 objects from 2 categories. Previously, we have shown cross-participant identification of visually depicted exemplars from the same category of objects (Shinkareva et al., 2008), and reliable object identification with 60 objects presented in a word form (Just et al., 2010) or combined word and picture form (Mitchell et al., 2008). Therefore, we expect the across stimulus format cross participant classification results to generalize to larger sets of exemplars.

The current study used words in a written form, however existing work suggests that cross format identification would work with auditorily presented stimuli as well. Despite differences in

[1] Anatomical regions that support both word and picture category identification, but not identification across formats, probably due to the coarseness of AAL regions or lack of statistical power, are not discussed.

representational format semantic information from words presented in auditory form and pictures has been shown to similarly integrate into a sentence context (Willems et al., 2008). Moreover, successful decoding has been shown for distinct activation patterns elicited by different speech sounds (Formisano et al., 2008) and syllables (Raizada et al., 2010). Thus the multivoxel pattern classification is revealing the many ways that physically different stimuli can evoke common neural substrates that correspond to the shared underlying conceptual basis.

Acknowledgments

This research was supported by the W. M. Keck Foundation and the National Science Foundation - Collaborative Research in Computational Neuroscience Grant IIS-0423070. We would like to thank Vladimir Cherkassky for helpful comments, and Stacey Becker and Rachel Krishnaswami for help in the preparation of the manuscript.

References

Aziz-Zadeh, L., Damasio, A., 2008. Embodied semantics for actions: findings from functional brain imaging. J. Physiol. Paris 102, 35–39.
Barsalou, L.W., 1999. Perceptual symbols systems. Behav. Brain Sci. 22, 577–660.
Boronat, C.B., Buxbaum, L.J., Coslett, H.B., Tang, K., Saffran, E.M., Kimberg, D.Y., Detre, G.J., 2005. Distinctions between manipulation and function knowledge of objects: evidence from functional magnetic resonance imaging. Cogn. Brain Res. 23, 361–373.
Bright, P., Moss, H., Tyler, L.K., 2004. Unitary vs multiple semantics: PET studies of word and picture processing. Brain Lang. 89 (3), 417–432.
Carlson, T.A., Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15 (5), 704–717.
Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2 (10), 913–919.
Chee, M.W., Weekes, B., Lee, K.M., Soon, C.S., Schreiber, A., Hoon, J.J., Chee, M., 2000. Overlap and dissociation of semantic processing of Chinese characters, English words, and pictures: evidence from fMRI. Neuroimage 12 (4), 392–403.
Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270.
Cusack, R., Veldsman, M., Naci, L., Mitchell, D., 2010. Using Dynamically Adaptive Imaging with fMRI to rapidly characterize neural representations. In Proceedings of the ISMRM, 18th Scientific Meeting (Stockholm) p. 2346.
Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage 28 (3), 663–668.
Devlin, J.T., Rushworth, M.F.S., Matthews, P.M., 2005. Category-related activation for written words in the posterior fusiform is task specific. Neuropsychologia 43, 69–74.
Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., Kleinschmidt, A., 2009. Deciphering cortical number coding from human brain activity patterns. Curr. Biol. 19, 1608–1615.
Etzel, J.A., Gazzola, V., Keysers, C., 2008. Testing simulation theory with cross-modal multivariate classification of fMRI data. PLoS ONE 3 (11), e3690.
Formisano, E., Martino, F.D., Bonte, M., Goebel, R., 2008. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322, 970–973.
Gallagher, H.L., Happe, F.G., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: an fMRI study of "theory of mind" in verbal and nonverbal tasks. Neuropsychologia 38, 11–21.
Gates, L., Yoon, M.G., 2005. Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: an fMRI study. Neuroimage 24 (2), 473–486.
Glenberg, A.M., 1997. What memory is for. Behav. Brain Sci. 20, 1–55.
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24, 187–203.
Hanson, S.J., Halchenko, Y.O., 2007. Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. Neural Comput. 20 (2), 486–503.
Hanson, S.J., Matsuka, T., Haxby, J.V., 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage 23, 156–166.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430.
Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7 (7), 523–534.
Hodges, J.R., Patterson, K., 1997. Semantic memory disorders. Trends Cogn. Sci. 1, 68–72.
Hoenig, K., Sim, E., Bochev, V., Herrnberger, B., Kiefer, M., 2008. Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. J. Cogn. Neurosci. 20 (10), 1799–1814.
James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., Goodale, M.A., 2002. Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. Neuron 35, 793–801.
Just, M.A., Cherkassky, V.L., Aryal, S., Mitchell, T.M., 2010. A neurosemantic theory of concrete noun representation based on the underlying brain codes. PLoS ONE 5 (1), e8622.
Knops, A., Thirion, B., Hubbard, E.M., Michel, V., Dehaene, S., 2009. Recruitment of an area involved in eye movements during mental arithmetic. Science 324, 1583–1585.
Mechelli, A., Price, C., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cereb. Cortex 14, 1256–1265.
Mitchell, T.M., 1997. In: Liu, C.L. (Ed.), Machine learning. McGraw-Hill, Boston.
Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195.
Mourao-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. Neuroimage 28 (4), 980–995.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430.
O'Toole, A., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17 (4), 580–590.
O'Toole, A.J., Jiang, F., Abdi, H., Penard, N., Dunlop, J.P., Parent, M.A., 2007. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. J. Cogn. Neurosci. 19 (11), 1735–1752.
Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S.F., Fazio, F., 1999. Word and picture matching: a PET study of semantic category effects. Neuropsychologia 37 (3), 293–306.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209.
Pietrini, P., Furey, M.L., Ricciardi, E., Gobbini, M.I., Wu, W.H., Cohen, L., Guazzelli, M., Haxby, J.V., 2004. Beyond sensory images: object-based representation in the human ventral pathway. Proc. Natl Acad. Sci. USA 101 (15), 5658–5663.
Poldrack, R.A., Halchenko, Y.O., Hanson, S.J., 2009. Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychol. Sci. 20 (11), 1364–1372.
Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. Science 310, 1963–1966.
Pulvermuller, F., 2005. Brain mechanisms linking language and action. Nat. Rev. Neurosci. 6, 1–7.
Pulvermuller, F., Kherif, F., Hauk, O., Mohr, B., Nimmo-Smith, I., 2009. Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. Hum. Brain Mapp. 12, 3837–3850.
Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S., 1992. Human cerebral cortex: localization, parcellation and morphometry with magnetic resonance imaging. J. Cogn. Neurosci. 4, 352–374.
Raizada, R.D.S., Tsao, F., Liu, H., Kuhl, P., 2010. Quantifying the adequacy of neural representations for cross-language phonetic discrimination task: prediction of individual differences. Cereb. Cortex 20, 1–12.
Shinkareva, S.V., Mason, R.A., Malave, V.L., Wang, W., Mitchell, T.M., Just, M.A., 2008. Using FMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS ONE 3 (1) e1394-e1394.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.
Vandenberghe, R., Price, C., Wise, R., Josephs, O., Frackowiak, R.S., 1996. Functional anatomy of a common semantic system for words and pictures. Nature 383 (6597), 254–256.
Vuilleumier, P., Henson, R.N., Driver, J., Dolan, R.J., 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat. Neurosci. 5, 491–499.
Willems, R.M., Ozyurek, A., Hagoort, P., 2008. Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. J. Cogn. Neurosci. 20 (7), 1235–1249.