Slightly revised version to appear in J. Neuroscience.

Running title: Decoding multimodal concept representations

Heteromodal cortical areas encode sensory-motor features of word meaning

Leonardo Fernandino[1], Colin J. Humphries[1], Lisa L. Conant[1],

Mark S. Seidenberg[2], Jeffrey R. Binder[1]

[1]Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA

[2]Department of Psychology, University of Wisconsin, Madison, WI, USA

Correspondence: Leonardo Fernandino, Department of Neurology, Medical College of Wisconsin, 8701 Watertown Plank Road, MEB 4671, Milwaukee, WI, 53226, USA.

email: lfernandino@mcw.edu

3 figures, 1 table

Abstract: 249 words. Introduction: 500 words. Discussion: 707 words. Total: 4,432 words

No conflicts of interest to report.

Keywords: lexical semantics; fMRI; semantic memory; embodiment; concepts; multimodal processing; predictive machine learning.

Abstract

The capacity to process information in conceptual form is a fundamental aspect of human cognition, yet little is known about how this type of information is encoded in the brain. Although the role of sensory and motor cortical areas has been a focus of recent debate, neuroimaging studies of concept representation consistently implicate a network of heteromodal areas that seem to support concept retrieval in general rather than knowledge related to any particular sensory-motor content. We used predictive machine learning on fMRI data to investigate the hypothesis that the nodes in this "general semantic network" (GSN) encode multimodal information about the simultaneous activation of different sensory-motor features, functioning as convergence-divergence zones for distributed concept representation. A computational model based on five conceptual attributes directly related to sensory-motor processes (sound, color, shape, manipulability, and visual motion) was used to predict brain activation patterns associated with individual lexical concepts in a semantic decision task. When the analysis was restricted to voxels in the GSN, the model was able to identify the activation patterns corresponding to particular concepts significantly above chance. In contrast, a model based on five perceptual attributes of the word form performed at chance level. This pattern was reversed when the analysis was restricted to areas involved in the perceptual analysis of written word forms. These results indicate that heteromodal areas involved in semantic processing encode information about the relative importance of different sensory-motor attributes. These high-level association areas likely encode crossmodal conjunctions representing particular combinations of sensory-motor content.

Significance Statement

The present study employs a simple computational model of word semantics to decode conceptual information from neural activity in heteromodal cortical areas. The model is based on five sensory-motor attributes of word meaning (color, shape, sound, visual motion, and manipulability), and encodes the relative importance of each attribute to the meaning of a word. This is the first demonstration that heteromodal areas involved in semantic processing can discriminate between different concepts based on sensory-motor information alone. This finding indicates that the brain represents concepts as multimodal combinations of sensory and motor representations.

1. Introduction

The capacity to encode and retrieve conceptual information is an essential aspect of human cognition, but little is known about how these processes are implemented in the brain. Neuroimaging studies of conceptual processing have implicated areas at various levels of the cortical hierarchy, including sensory and motor areas (e.g., Hauk et al., 2004; Hsu et al., 2012) as well as multimodal (Fernandino et al., 2015) and heteromodal regions (Binder et al., 2009). Binder et al. referred to the latter as a “general semantic network” (GSN) because it responds more to meaningful input (words and sentences) than to meaningless input (nonwords and scrambled sentences), regardless of the particular sensory-motor content of the items. The GSN consists of portions of the inferior parietal lobule (IPL), lateral temporal cortex (LTC), ventrolateral prefrontal cortex (VLPFC), precuneus/posterior cingulate gyrus (Pc/PCG), parahippocampal gyrus (PHG), and medial prefrontal cortex (MPFC), all bilaterally activated, with stronger activations in the left hemisphere. According to embodied models of semantics, lower-level sensory and motor areas contribute to concept representation by encoding the sensory-motor features of phenomenal experience that characterize each concept, presumably derived from the experiences that led to the formation of the concept. However, the role of the GSN remains obscure. We propose that this network encodes high-level representations of the co-activation patterns exhibited by lower-level, sensory-motor cortical areas during concept retrieval, in line with the idea of convergence-divergence zones originally proposed by Damasio (1989) and further developed by Simmons and Barsalou (2003). Alternatively, it is possible that the GSN encodes conceptual representations in a qualitatively distinct format that does not rely on sensory-motor information. The existence of such a disembodied code for concept representation has been endorsed by some authors (see, for instance, Mahon and Caramazza, 2008).

We set out to investigate whether the heteromodal cortical areas comprising the GSN encode sensory-motor information about concrete concepts during word-cued concept retrieval. We used a forward encoding model based on five sensory-motor attributes of word meaning (sound, color, visual motion, shape, and manipulability) to decode the distributed fMRI activation patterns associated with the meanings of 80 common nouns. We anticipated that this model of word meaning (from here on referred to as the "semantic model") would successfully identify individual concrete concepts from neural activity in the GSN. As a control, we predicted that an alternative model based on five orthographic and phonologic attributes of the word form (the "word-form model") would not decode activation patterns in the GSN above chance levels.

As an additional control, we also evaluated both encoding models in a different set of cortical regions, namely, those involved in the perceptual analysis of written word forms. This "word form network" (WFN) includes occipital and ventral temporal visual areas, as well as premotor and supplementary motor areas (e.g., Cohen et al., 2004). Thus, we expected the decoding accuracy of the two encoding models in these areas to show the opposite pattern relative to the GSN, that is, successful decoding for the word-form model but not for the semantic model.

2. Materials and Methods

2.1. Attribute ratings

The semantic model was based on five semantic attributes directly related to sensory-motor processes: sound, color, shape, manipulability, and visual motion. Ratings for these attributes were available for a set of 900 words (see Fernandino et al., 2015, for details). The ratings reflect the relevance of each attribute to the meaning of the word on a 7-point Likert scale ranging from “not at all important” to “very important”. Approximately 30 participants rated each attribute for each word.

2.2. Data source

We direct the reader to Fernandino et al. (2015) for details on the stimuli and data collection procedures, which are summarized below.

2.2.1. Participants

Participants were 44 healthy, right-handed, native speakers of English with no history of neurological or psychiatric disorders (16 females; mean age 28.2, range 19 to 49). They gave informed consent as approved by the Medical College of Wisconsin Institutional Review Board and were compensated for participation.

2.2.2. Stimuli

Stimuli consisted of the 900 nouns for which attribute ratings were available (see section 4.1. above) and 300 pseudowords. Six hundred nouns were relatively concrete and 300 were relatively abstract, as determined by either published imageability ratings or consensus judgment of the authors. Pseudowords were matched to the words on length, orthographic neighborhood density, and bigram and trigram metrics.

2.2.3. Task procedure

The stimuli were back-projected on a screen that was viewed by the participant through a mirror attached to the head coil. Participants performed 1200 trials (900 words, 300 pseudowords), distributed over 10 runs. Each stimulus was presented for 1000 ms and followed by a fixation cross for a jittered interval of 1-13 s.

Participants performed a speeded semantic decision task (“can it be directly experienced with the senses?”), and responded by pressing one of two response keys with their right hand. They were instructed to press the button for "no" in the case of pseudowords.

2.2.4. FMRI acquisition and preprocessing

Gradient-echo EPI images were collected in 10 runs of 196 volumes each. Twenty-three participants were scanned on a GE 1.5T Signa MRI scanner (TR = 2000 ms, TE = 40 ms, 21 axial slices, 3.75 x 3.75 x 6.5 mm voxels), and the other 21 were scanned on a GE 3T Excite MRI scanner (TR = 2000 ms, TE = 25 ms, 40 axial slices, 3 x 3 x 3 mm voxels). T1-weighted anatomical images were obtained using a 3D SPGR sequence with voxel dimensions of 1 mm isotropic.

EPI volumes were corrected for slice acquisition time and head motion. They were aligned to the T1-weighted volume and transformed into Talairach standard space (Talairach & Tournoux, 1988), resampled at 3 mm isotropic voxels, and smoothed with a 6 mm FWHM Gaussian kernel. Each voxel time series was rescaled to percent of mean signal level, so that subsequent regression parameter estimates reflected percent signal change.

2.3. Forward encoding models

The semantic model was designed to predict the activation pattern corresponding to a given word based on the ratings of the five semantic attributes for that word (see section *Attribute ratings* above). The word-form model was designed to predict activation patterns based on perceptual properties of the word form, regardless of meaning, thus serving as a control for the semantic model. It was based on five orthographic and phonologic attributes of the word form: number of letters, number of syllables, orthographic neighborhood density, phonologic neighborhood density, and bigram frequency.

2.4. Decoding algorithm

For the decoding procedure, we split the 900 word stimuli into a modeling set, consisting of 820 items, and a test set, consisting of 80 items (40 concrete and 40 abstract). Test words were selected randomly with the constraint that the concrete and abstract subsets were matched in word frequency, number of letters, number of phonemes, number of syllables, orthographic and phonologic neighborhood densities, and bigram frequency (Table 1).

Table 1 here.

The decoding algorithm consisted of four steps:

1. *Generation of attribute maps*. Activation maps for each attribute in the encoding model (Attribute Maps, or AMs) were generated for each participant based exclusively on the words in the modeling set (Figure 1). This was done by including the z-transformed attribute values (sensory-motor ratings in the case of the semantic model, orthographic and phonologic measures in the case of the word-form model) as simultaneous predictor variables in a Generalized Least Squares (GLS) regression. For the semantic model, nuisance regressors included word length, number of phonemes, number of syllables, word frequency, bigram frequency, orthographic and phonological neighborhood density, and the participant's RT for each trial (all z-transformed). For the word-form model, nuisance regressors included the five sensory-motor ratings of word meaning and the participant's RT for each trial (all z-transformed). Two binary regressors — one coding for "word" events and the other for "pseudoword" events — were included to account for residual activity associated with early visual processing of the stimulus, as well as the subsequent motor response. Signal drift was modeled with linear, second-order, and third-order trends, and residual movement artifacts were modeled with the estimates of the motion parameters. A group-level AM for each attribute was created by averaging the individual AMs (beta values) across participants.

Figure 1 here.

2. *Computing predicted word maps*. For each of the 80 words in the test set, predicted activation maps (Predicted Maps, PMs) were computed as linear combinations of the AMs, whereby each AM was weighted by the word's corresponding attribute value (Figure 2A). In the case of the semantic model, the PM for a given test word corresponded to the hypothetical activation pattern that would be associated with the meaning of that word if the word's meaning were completely captured by the five attribute ratings (i.e., sound, color, manipulation, visual motion, and shape).

3. *Generation of observed word maps*. From the imaging data, activation maps were generated for each of the 80 words in the test set, relative to a pseudoword baseline (Observed Maps, OMs). For each participant, a separate GLS regression was conducted for each of the 80 test words, with the following explanatory variables: a binary regressor coding for the presentation of the selected test word; a binary regressor coding for presentation of all the non-selected words (i.e., the other 899 words in the stimulus set); a binary regressor coding for presentation of the pseudowords; five continuous regressors coding for each of the five attribute values for all non-selected words; and a continuous regressor coding the response time for each trial. Therefore, the resulting OM for a given test word corresponded to the unique activation pattern induced by that word. For each test word, a group-level OM was obtained by averaging the individual OMs (beta values) across participants.

Figure 2 here.

4. *Testing the PMs against the OMs*. The decoding accuracy of the model was evaluated separately for each word based on the similarity between the PM and its corresponding OM, relative to the similarity between the PM and all the other OMs (Figure 2B). Similarity was

defined as the voxel-by-voxel pairwise correlation between maps, and accuracy was defined as the percentile rank of the correlation strength between the PM and the corresponding OM. This percentile rank, scaled to a 0 to 1 range, was assigned to the PM as its accuracy score. Thus, each PM received an accuracy score corresponding to how similar it was to its respective OM relative to the other 79 OMs, with 0 corresponding to least similar and 1 corresponding to most similar. For example, if the OM for the word "tomato" were the most highly correlated to the PM for the same word, that word would receive an accuracy score of 79/79 = 1. If, instead, it were the second most highly correlated to its respective PM, its accuracy score would be 78/79 = 0.987. The Shapiro-Wilk normality test showed that the model's accuracy scores for the test words were not normally distributed, so we used non-parametric 95% confidence intervals and the Wilcoxon signed rank test to verify whether model performance (median decoding accuracy across the 80 test words) was significantly higher than chance (.5).

2.5. Voxel selection masks

Our hypothesis concerned the role of the GSN in representing sensory-motor information about concepts. Therefore, we created a mask based on the activation-likelihood estimation (ALE) meta-analysis by Binder et al. (2009), encompassing the cortical areas that were reliably associated with "general" semantic processing (Figure 3A). The map from Binder et al. was thresholded at $p < .05$ and converted into a binary mask, which was used to select the voxels included in the decoding analysis.

As a control, the models were also evaluated in a mask corresponding to the WFN, obtained from the contrast pseudowords > rest in the present data set, thresholded at $p < .05$ (corrected). This mask included visual, somatosensory, and motor/premotor areas, as well as the thalamus, and had minimal overlap with the GSN mask (Figure 3A). Since these regions are more strongly activated during bottom-up perceptual processing than during top-down processing (Goebel et al., 1998; O'Craven & Kanwisher, 2000), we expected their activation patterns to encode information primarily about word form, rather than semantic content.

2.6. Concrete versus abstract words

As mentioned in the *Decoding algorithm* section, the test set consisted of two matched subsets, one with 40 concrete words and the other with 40 abstract words. The two subsets were matched on all lexical attributes, except for concreteness. As shown in Table 1, the variance of the sensory-motor attribute ratings was much smaller among abstract than among concrete words, indicating that abstract word meanings contained much less information about the sensory-motor features included in the semantic model. Thus, if the accuracy of the semantic model were indeed driven by the sensory-motor aspects of word meaning, decoding performance should be high for concrete but low for abstract words.

Figure 3 here.

3. Results

3.1. Model performance in the GSN mask

Decoding accuracy for the two encoding models in the GSN mask is shown in Figure 3B. When all 80 test words were pooled together for the decoding procedure, decoding accuracy was significantly higher than chance for the semantic model (median = .68; 95%CI = [.58, .77]; V = 2319.5, p = .0004), indicating that the GSN encodes sensory-motor information about lexical concepts. When concrete and abstract words were decoded separately, decoding was successful for the concrete (median = .69; 95%CI = [.51, .77]; V = 608, p = .004) but not for the abstract words (median = .45; 95%CI = [.28, .64]; V = 438.5, p = .35). As expected, decoding accuracy for the word-form model was not significantly different from chance in the GSN (All words: median = .43; 95%CI = [.35, .58]; V = 1480.5, p = .75; Concrete: median = .45; 95%CI = [.31, .69]; V = 431.5, p = .39; Abstract: median = .43; 95%CI = [.28, .56]; V = 333, p = .85). Similar results were obtained when we excluded the GSN voxels that overlapped with the VWFN, with good decoding accuracy for the semantic (All words: median = .65; 95%CI = [.55, .75]; V = 2215, p = .002; Concrete: median = .67; 95%CI = [.53, .80]; V = 562, p = .02; Abstract: median = .40; 95%CI = [.25, .54]; V = 404.5, p = .53) but not for the word form model (all $p > .5$).

3.2. Model performance in the WFN mask

Decoding accuracy for the two encoding models in the WFN mask is shown in Figure 3C. Consistent with our hypothesis, the semantic model failed to decode activation patterns in this network (All words: median = .47; 95%CI = [.38, .67]; V = 1842.5, p = .14; Concrete: median = .32; 95%CI = [.21, .67]; V = 396, p = .58; Abstract: median = .50; 95%CI = [.44, .67]; V = 419.5, p = .45), whereas the word-form model was successful (All words: median = .66; 95%CI = [.59, .72]; V = 2292, p = .0006; Concrete: median = .67; 95%CI = [.59, .77]; V = 554.5, p = .03; Abstract: median = .64; 95%CI = [.54, .77]; V = 608.5, p = .004). Again, similar results were obtained when we excluded the WFN voxels that overlapped with the GSN from the analysis, for both the semantic (all $p > .2$) and the word form model (All words: median = .69; 95%CI = [.59, .79]; V = 2280.5, p = .0008; Concrete: median = .65; 95%CI = [.51, .80]; V = 542, p = .04; Abstract: median = .69; 95%CI = [.59, .77]; V = 637, p = .001).

4. Discussion

We evaluated two forward encoding models on their capacity to decode word-related information from fMRI activity patterns. The semantic model was based on five sensory-motor attributes of word meaning, while the word-form model was based on five orthographic and phonologic attributes of the word form. Each model was evaluated in two different sets of cortical areas: the GSN, a set of highly interconnected heteromodal areas that has been consistently implicated in semantic processing; and the WFN, involved in the perceptual analysis of word forms, comprising mainly visual and motor/somatosensory areas. We found that the semantic model successfully decoded fMRI activation patterns elicited by individual words in the GSN, but not in the WFN. As expected, decoding of GSN activity was successful for concrete but not for abstract words when the two sets were analyzed separately. The word-form model was successful in the WFN – for concrete and abstract words alike – but failed to decode activity in the GSN. This pattern of results strongly indicates that the GSN encodes information about sensory-motor attributes of concepts.

The GSN was identified by Binder et al. (2009) in an ALE meta-analysis of 120 neuroimaging studies of semantic word processing. It overlaps considerably with the "default mode network", a set of cortical areas typically deactivated during attention-demanding tasks relative to rest (for a review, see Buckner et al., 2008). Resting state connectivity and MRI tractography studies have shown that the core nodes of the network (IPL, LTC, Pc/PCG, and MPFC) are strongly interconnected (Greicius et al., 2009; Horn et al., 2014), and graph theoretical analyses have identified these regions as central connector hubs for more specialized, modular cortical networks (Hagmann et al., 2008; Sepulcre et al., 2012). Based on these findings, we have argued that the GSN supports multimodal conceptual representations by encoding patterns of co-activation across lower-level, modality-specific areas (Fernandino et al., 2015). The present results show that the GSN can discriminate between individual concrete concepts based exclusively on sensory-motor information, thus providing substantial support for this proposal. Although this finding, by itself, does not preclude the existence of a disembodied representational code, as proposed by Mahon and Caramazza (2008), it raises questions about the necessity of such a code and about its possible neural substrates.

The WFN mask included sensory-motor areas that have previously been found to encode information about word semantics (e.g., Hauk et al., 2004; Hsu et al., 2012; Fernandino et al., 2015). Why, then, did the semantic model fail to decode neural activation in this mask? We believe the answer lies in the nature of the task. Since perceptual word processing and concept retrieval took place virtually simultaneously in the present study – the two processes were modeled as a single event in the GLM estimation of beta values, due to the low temporal resolution of the BOLD signal – activity in the WFN was driven much more strongly by the perceptual features of the stimuli (bottom-up activation) than by their semantic attributes (top-down activation), thus greatly reducing the signal-to-noise ratio of the semantic activation patterns in those areas. Future studies should investigate this issue by dissociating concept retrieval from complex sensory stimulation.

Finally, we should note that the failure of the semantic model to decode activation patterns for abstract words does not necessarily imply that concrete and abstract concepts are based on qualitatively different codes; rather, it could reflect the fact that the relationship between the meaning of an abstract word and specific features of sensory-motor experience is much more complex and context-dependent than that of concrete words (Badre & Wagner, 2002; Barsalou and Wiemer-Hastings, 2005; Hoffman, 2015). Low prediction accuracy was predicted for abstract words based on the relatively low variance of the sensory-motor ratings across these words (Table 1).

Our results provide the first demonstration that heteromodal areas involved in semantic processing can discriminate between individual concepts based on sensory-motor information alone. They provide strong support for the view that conceptual representations are grounded, at least in part, in elementary sensory-motor attributes of phenomenal experience. Furthermore, they indicate that the neural architecture of these representations is hierarchically organized, with higher-level heteromodal areas encoding information about the activation patterns exhibited by lower-level sensory-motor areas – patterns that are, presumably, established during concept formation and partially reinstated during retrieval.

5. Acknowledgements

This project was supported by National Institute of Neurological Diseases and Stroke grant R01 NS33576 and by NIH General Clinical Research Center grant M01 RR00058.

6. References

Badre, D., & Wagner, A. D. (2002). Semantic retrieval, mnemonic control, and prefrontal cortex. *Behavioral and Cognitive Neuroscience Reviews*, *1*(3), 206–218.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thought* (pp. 129–163). New York: Cambridge University Press.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.

Cohen, L., Jobert, A., Le Bihan, D., and Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*, 23, 1256-1270.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1-2), 25–62.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., et al. (2015). Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics. *Cerebral Cortex*, 1–17.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, *10*(5), 1563–1573.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, *19*(1), 72–78.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), 1479–1493. Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hoffman, P. (2015). The meaning of “life” and other abstract words: Insights from neuropsychology. *Journal of Neuropsychology*.

Horn, A., Ostwald, D., Reisert, M., & Blankenburg, F. (2014). The structural–functional connectome and the default mode network of the human brain. *NeuroImage*, *102*(P1), 142–151.

Hsu, N. S., Frankland, S. M., & Thompson-Schill, S. L. (2012). Chromaticity of color perception and object color knowledge. *Neuropsychologia*, *50*(2), 327–333.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris*, *102*(1-3), 59–70.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise Connectivity of the Modal Cortex Reveals the Multimodal Organization of the Human Brain. *The Journal of Neuroscience*, *32*(31), 10649–10661.

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, *20*(3), 451–486.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging. New York: Thieme.

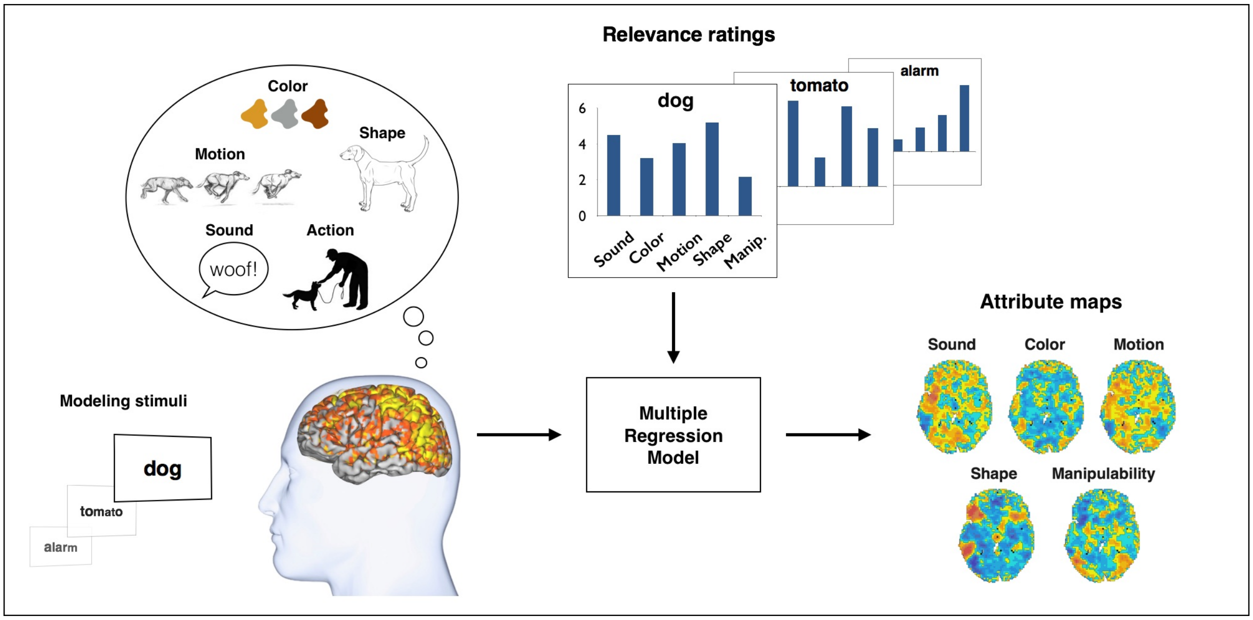
	Concrete	Abstract	T test (p)
Number of letters	5.75 (1.97)	5.75 (2.02)	0.70
Number of phonemes	4.47 (1.75)	4.87 (1.9)	0.33
Number of syllables	1.67 (0.8)	2 (1.01)	0.11
Log Frequency HAL	9.86 (1.31)	9.44 (1.5)	0.18
Orth. neighborhood	4.87 (5.51)	4.27 (5.81)	0.64
Phon. neighborhood	9.57 (10.22)	8.72 (11.52)	0.73
Bigram frequency	1722 (842)	1902 (933)	0.37
Concreteness	4.81 (0.19)	2.21 (0.65)	< .0001
Sound rating	2.39 (1.47)	0.96 (0.84)	< .0001
Color rating	3.32 (1.07)	0.60 (0.61)	< .0001
Manipulation rating	2.43 (1.42)	0.78 (0.53)	< .0001
Motion rating	2.42 (1.68)	0.81 (0.77)	< .0001
Shape rating	3.90 (1.27)	0.33 (0.26)	< .0001

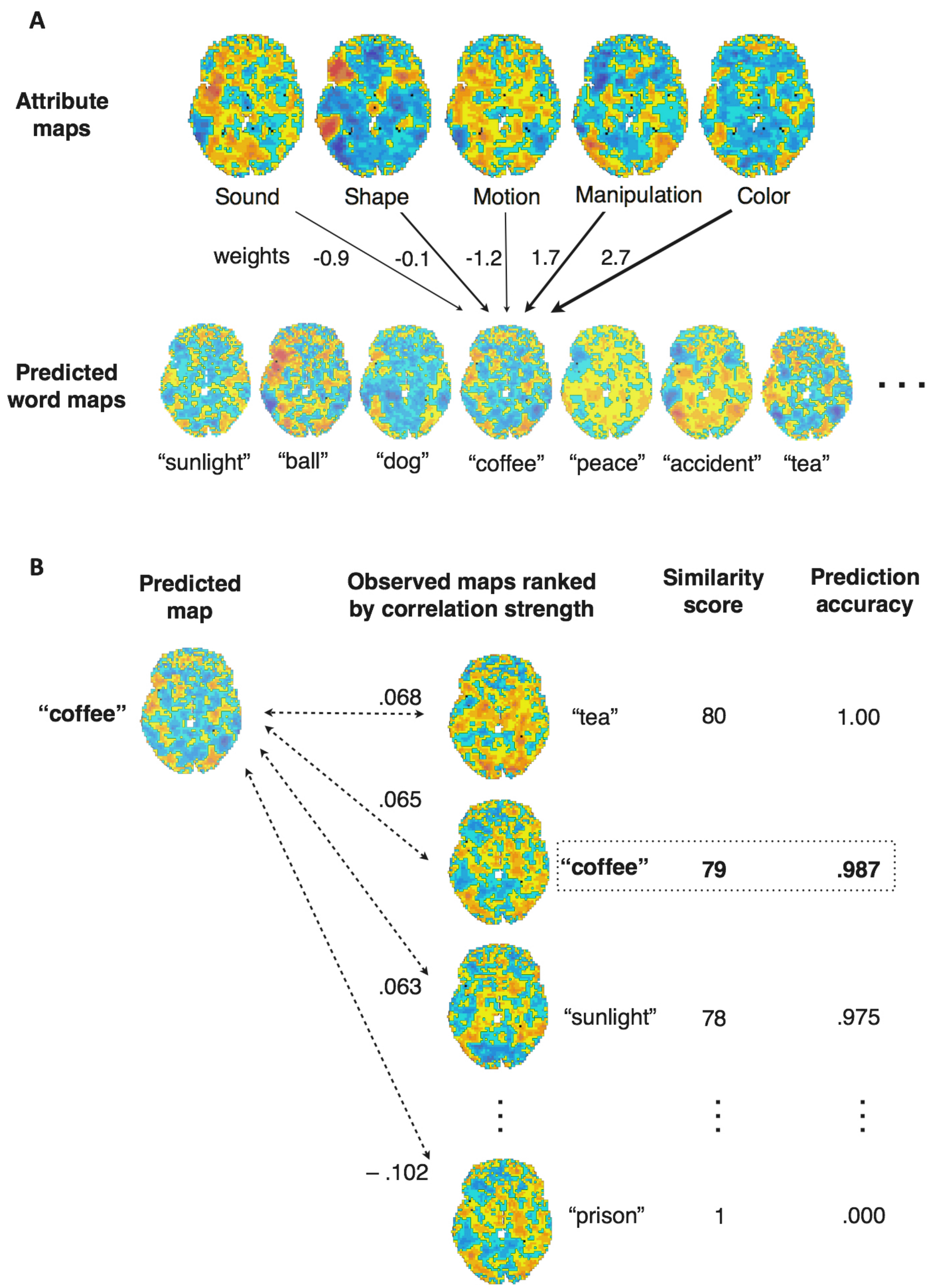
Table 1: Lexical and semantic attributes [mean (standard deviation)] for the two subsets of test words. Concreteness data is from Brysbaert et al. (2014). All other lexical attributes were obtained from the English Lexicon Project (http://elexicon.wustl.edu).

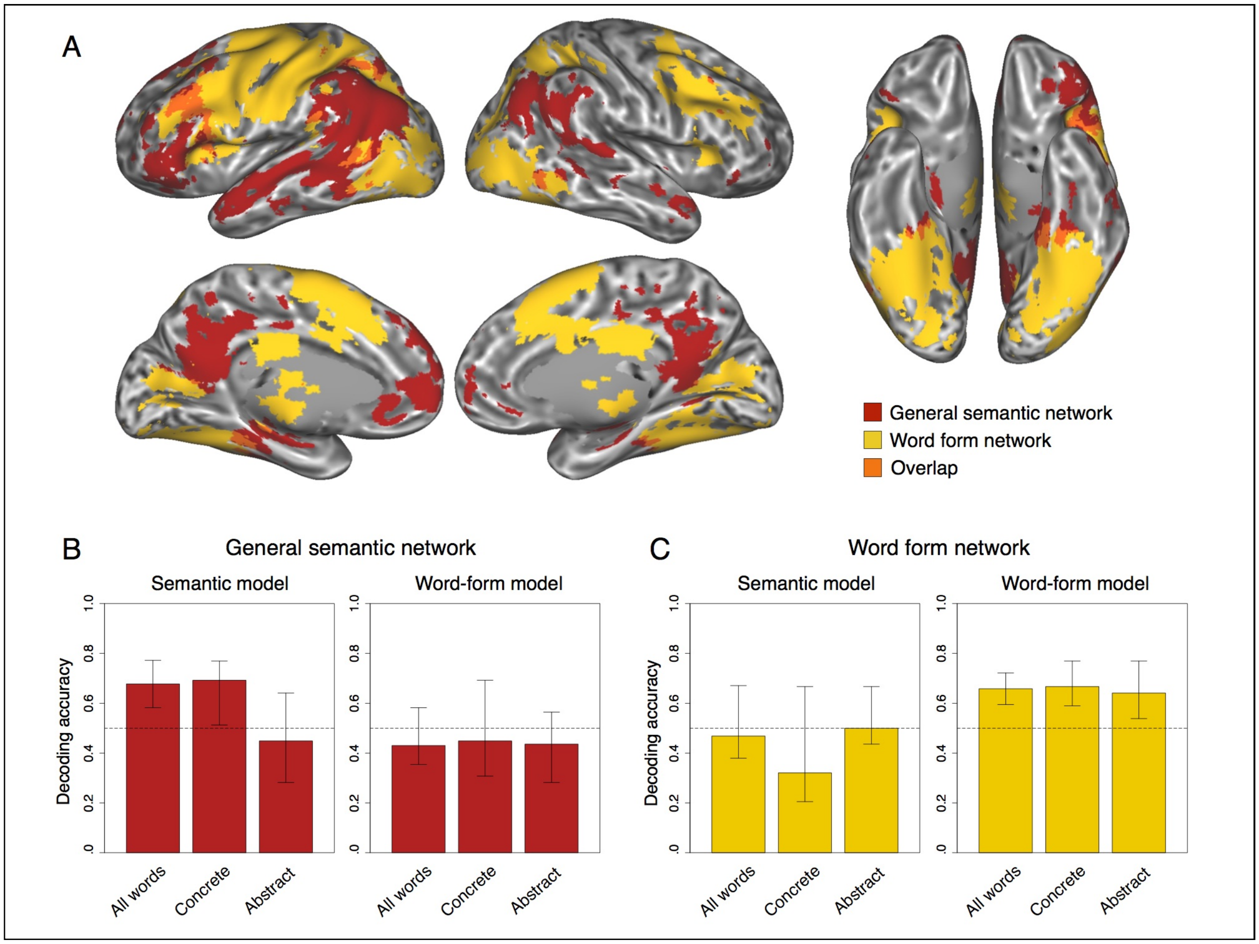
Figure 1. Activation images for all the 820 words in the modeling set were combined with their corresponding attribute ratings in a least squares multiple regression model, resulting in 5 attribute maps.

Figure 2. A. For each word in the test set (e.g., "coffee"), a predicted map was generated by a weighted sum of the 5 attribute maps, where each map was weighted by its corresponding attribute rating for that word. B. The voxel-by-voxel correlations between the predicted map and each of the 80 observed maps were computed, and the observed maps ranked by correlation strength. Decoding accuracy was determined from the percentile rank of the observed map for the corresponding predicted map.

Figure 3. A. Masks used for voxel selection. B. Decoding accuracy for each model in each of the two masks.


Relevance ratings
Color
Shape
Motion
Sound
Action
woof!
dog
tomato
alarm
6
4
2
0
Sound
Color
Motion
Shape
Manip.
Modeling stimuli
dog
tomato
alarm
Multiple Regression Model
Attribute maps
Sound
Color
Motion
Shape
Manipulability


A
Attribute maps
Sound
Shape
Motion
Manipulation
Color
weights -0.9 -0.1 -1.2 1.7 2.7
Predicted word maps
"sunlight"
"ball"
"dog"
"coffee"
"peace"
"accident"
"tea"
. . .
B
Predicted map
Observed maps ranked by correlation strength
Similarity score
Prediction accuracy
"coffee"
.068
"tea" 80 1.00
.065
"coffee" 79 .987
.063
"sunlight" 78 .975
– .102
"prison" 1 .000


A
General semantic network
Word form network
Overlap
B
General semantic network
Semantic model
Word-form model
Decoding accuracy
0.0
0.2
0.4
0.6
0.8
1.0
All words
Concrete
Abstract
C
Word form network
Semantic model
Word-form model
Decoding accuracy
All words
Concrete
Abstract