Neuropsychologia 76 (2015) 17–26

Contents lists available at ScienceDirect

Neuropsychologia

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

Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes

Leonardo Fernandino [a,*], Colin J. Humphries [a], Mark S. Seidenberg [b], William L. Gross [c], Lisa L. Conant [a], Jeffrey R. Binder [a]

[a] Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA
[b] Department of Psychology, University of Wisconsin, Madison, WI, USA
[c] Department of Anesthesiology, Medical College of Wisconsin, Milwaukee, WI, USA

ARTICLE INFO

Article history:
Received 10 October 2014
Received in revised form
2 April 2015
Accepted 7 April 2015
Available online 8 April 2015

Keywords:
Lexical semantics
fMRI
Semantic memory
Embodiment
Concepts
Multimodal processing

ABSTRACT

While major advances have been made in uncovering the neural processes underlying perceptual representations, our grasp of how the brain gives rise to conceptual knowledge remains relatively poor. Recent work has provided strong evidence that concepts rely, at least in part, on the same sensory and motor neural systems through which they were acquired, but it is still unclear whether the neural code for concept representation uses information about sensory-motor features to discriminate between concepts. In the present study, we investigate this question by asking whether an encoding model based on five semantic attributes directly related to sensory-motor experience – sound, color, visual motion, shape, and manipulation – can successfully predict patterns of brain activation elicited by individual lexical concepts. We collected ratings on the relevance of these five attributes to the meaning of 820 words, and used these ratings as predictors in a multiple regression model of the fMRI signal associated with the words in a separate group of participants. The five resulting activation maps were then combined by linear summation to predict the distributed activation pattern elicited by a novel set of 80 test words. The encoding model predicted the activation patterns elicited by the test words significantly better than chance. As expected, prediction was successful for concrete but not for abstract concepts. Comparisons between encoding models based on different combinations of attributes indicate that all five attributes contribute to the representation of concrete concepts. Consistent with embodied theories of semantics, these results show, for the first time, that the distributed activation pattern associated with a concept combines information about different sensory-motor attributes according to their respective relevance. Future research should investigate how additional features of phenomenal experience contribute to the neural representation of conceptual knowledge.

1. Introduction

Brain-based theories of concept representation generally begin with the premise that concepts are learned through a process of generalization or abstraction from individual experiences. Experiences are comprised of identifiable sensory, motor, spatial, temporal, affective, and cognitive components, and the relative contribution of each of these experiential components to concept formation depends on the particular concept being acquired. For example, the concept of "thunder" is learned almost entirely through auditory experiences, whereas the concept of "octopus" has no connection with audition. Concepts for manipulable objects such as "pencil" and "spoon" are learned in part through motor actions performed with those objects, whereas concepts such as "moon" and "giraffe" have very little (if any) basis in motor experience. However, most experiences are inherently multimodal in the sense that they involve simultaneous, covarying information from multiple sensory-motor modalities. Using a pair of scissors, for example, typically involves concurrent sensory-motor experiences related to action planning, proprioceptive and tactile feedback, and the perception of characteristic shape, motion, and sound features. Thus, the concept of "scissors" is likely to include representations of all of these attributes.

Sensory information is processed according to perceptual attributes that can be strictly unimodal (e.g., color) or result from the combination of two or more sensory modalities (e.g., perception of object shape often combines visual and tactile information).

* Corresponding author.
E-mail address: lfernandino@mcw.edu (L. Fernandino).

http://dx.doi.org/10.1016/j.neuropsychologia.2015.04.009
0028-3932/

Embodied or "simulation" theories of concept representation propose that the information abstracted from experience during concept learning is represented to some degree in the same modality-specific and multimodal neural systems through which the learning occurred, and that concept retrieval involves some degree of activation of this sensory-motor information (Damasio, 1989; Barsalou, 2008; Glenberg and Gallese, 2012; Hoenig et al., 2011; Kiefer et al., 2007).

Although embodiment theories have garnered extensive empirical support (for reviews, see Fischer and Zwaan, 2008; Meteyard and Vigliocco, 2008; Binder and Desai, 2011; Kiefer and Pulvermüller, 2012; and Meteyard et al., 2012), most studies have aimed at demonstrating the existence of modality-specific representational systems (e.g., Warrington and Shallice, 1984; McCarthy and Warrington, 1988; Farah and McClelland, 1991), or at implicating a particular modality-specific cortical area in the representation of concrete concepts (e.g., Martin et al., 1996; Chao et al., 1999; Hauk et al., 2004; Aziz-Zadeh et al., 2006; Kiefer et al., 2008; Hsu et al., 2012). One question that remains unanswered is whether the distributed neural representation of a concept combines information originating from various sensory-motor attributes, as predicted by embodiment theories. In the present study, we explore this issue by adopting a novel approach to brain activation data relevant to this question. Instead of searching for localized activations associated with particular semantic attributes, we combine information about five different attributes to predict the activation patterns associated with individual concepts. In other words, we ask whether information about five sensory-motor attributes of lexical concepts is sufficient to predict patterns of brain activation elicited by isolated words. The five attributes selected for study – color, shape, visual motion, sound, and manipulation – are each associated with well-studied brain networks (for a review, see Fernandino et al., 2015). In a prior study, we obtained salience ratings on each of these attributes for a set of 900 English nouns, and showed that these attribute ratings parametrically modulate word-related brain activity in distinct cortical networks containing both unimodal and multimodal nodes (Fernandino et al., 2015). In the present study, attribute-specific activation maps, derived from a set of 820 words (the "modeling set"), were combined into an "encoding model" (Haxby et al., 2014; Naselaris et al., 2011), which was used to predict word-specific, whole-brain activation patterns for the 80 remaining words (the "test set"). The encoding model consisted of a linear combination of the five attribute maps from the training set, weighted by the test word's attribute rating values. The predicted activation patterns were then compared to the observed activation patterns for each test word. Successful prediction of the test word patterns would provide direct evidence that the overall, distributed neural representation of a concrete concept encodes information about the relative relevance of specific aspects of sensory-motor experience, as rated by an independent group of participants.

Another novel aspect of the present study is the focus on group-averaged activation patterns. Studies using multivoxel pattern analysis (MVPA) typically evaluate the performance of the encoding model (or classifier, in the case of pattern classification studies) separately for each participant, to account for individual differences in brain morphology and function-structure mapping, and perform group-level statistical tests on the resulting accuracy scores (e.g., Mitchell et al., 2008; Huth et al., 2012; Haxby et al., 2001). Here, we assess whether similarities in the neural code for concrete concepts across individuals would allow a group-level encoding model to successfully predict the group-averaged activation maps corresponding to different concepts. If prediction is successful, the encoding model can be seen as a first approximation to a subject-independent neural code for concept representation, and the predicted activation maps could be considered as rough neural signatures of the respective concepts.

2. Methods

2.1. Attribute ratings

We focused on five semantic attributes related to sensory-motor experience: sound, color, manipulation, visual motion, and shape. Ratings for these attributes were available for a set of 900 words (see Fernandino et al., 2015, for details). The ratings reflect the salience of each attribute to the meaning of the word on a 7-point Likert scale ranging from "not at all important" to "very important". The data set included approximately 30 ratings of each attribute for each word. Fig. 1 shows mean ratings for six example words, and Table 1 lists the correlations between the five attribute ratings across all words.

2.2. fMRI data

All analyses were conducted on the data from Fernandino et al. (2015). Data collection procedures from that study are summarized below.

2.2.1. Participants

Forty-four healthy, native speakers of English (16 females; mean age 28.2, range 19–49) with no history of neurological or psychiatric disorders, participated in the study. All were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were compensated for their participation and gave informed consent in conformity with the protocol approved by the Medical College of Wisconsin Institutional Review Board.

2.2.2. Stimuli

The stimuli consisted of the 900 nouns for which attribute ratings were collected (see Section 2.1. above) and 300 pseudowords. All words were relatively familiar (mean CELEX frequency = 37.4 per million, SD = 118.5), and between 3 and 9 letters in length, with a flat distribution across this length range (i.e., 126–129 words of each length). Six hundred of the words were relatively concrete and 300 were relatively abstract, as determined by either published imageability ratings, which at the time were available for 748 words (Wilson, 1988; Bird et al., 2001; Clark and Paivio, 2004), or consensus judgment of the authors. The pseudowords were generated by a computer program (Medler and Binder, 2005) using constrained trigram statistics (i.e., third-order approximation to English), followed by exclusion of pseudohomophones. Pseudowords were matched to the words on length, orthographic neighborhood density, and bigram and trigram metrics (Table 2).

2.2.3. Task procedure

The stimuli were back-projected in white Courier font on a black background screen that was viewed by the participant through a mirror attached to the head coil. Stimuli subtended an average horizontal visual angle of approximately 2.5°. In addition to the 1200 task trials (900 words, 300 pseudowords), 600 passive fixation events ("+") were included to act as a baseline and provide jittering for the deconvolution analysis, resulting in a total of 1960 stimulus events distributed across 10 runs. On task trials, the stimulus string was presented for 1000 ms followed by a fixation cross for 1000 ms; on fixation events, the fixation cross appeared for 2000 ms. Thus, each stimulus was followed by a varied fixation interval of 1 s, 3 s, 5 s, etc. The probability distribution of these fixation intervals was an exponentially decaying function with a

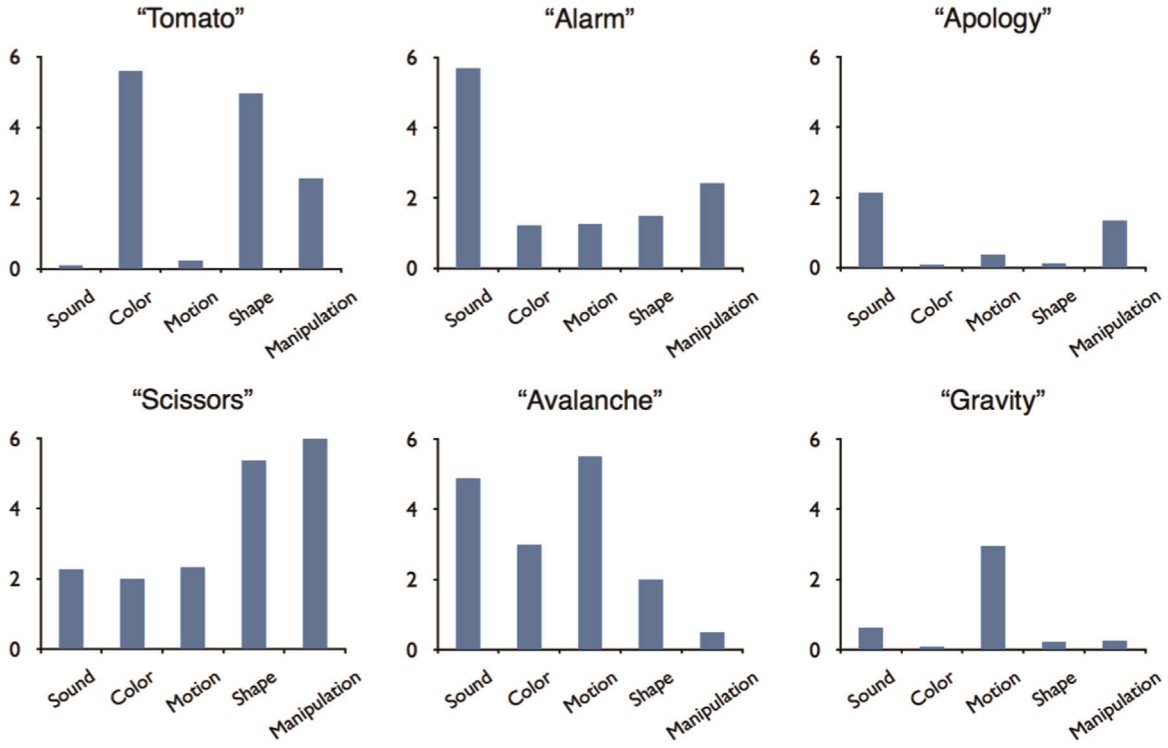

Fig. 1. Examples of words used in the study. Every word had a rating for each of the five semantic features.

Table 1
Correlations between semantic attribute ratings.

	Sound	Color	Manipulation	Visual motion
Color	0.12	–	–	–
Manipulation	0.33	0.31	–	–
Visual motion	0.59	0.37	0.25	–
Shape	0.28	0.76	0.58	0.53

median of 1 s and maximum of 13 s.

Participants were required to decide whether the stimulus (word or pseudoword) referred to something that can be experienced through the senses (i.e., a "concreteness decision"), and to respond as quickly as possible by pressing one of two response keys. They were informed that some of the items would be pseudowords, and that the answer should be "no" in these cases. All participants responded with their right hand using the index and middle fingers. Eprime 1.0 (Psychology Software Tools) was used for stimulus presentation and response registration.

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, $3.75 \times 3.75 \times 6.5$ mm^3 voxels), and the other 21 were scanned on a GE 3T Excite MRI scanner (TR=2000 ms, TE=25 ms, 40 axial slices, $3 \times 3 \times 3$ mm^3 voxels). T1-weighted anatomical images were obtained using a 3D spoiled gradient-echo sequence with voxel dimensions of $1 \times 1 \times 1$ mm^3.

Preprocessing and statistical analysis of fMRI data were done with the AFNI software package (Cox, 1996). EPI volumes were corrected for slice acquisition time and for head motion. Functional volumes were then aligned to the T1-weighted anatomical volume and transformed into Talairach standard space (Talairach and Tournoux, 1988), resampled at 3 mm isotropic voxels, and smoothed with a 6 mm FWHM Gaussian kernel. Each voxel time series was then rescaled to percent of mean signal level, so that subsequent regression parameter estimates reflect percent signal change.

2.3. Prediction analyses

For the prediction analyses, the 900 word stimuli were split into two groups: a modeling set, consisting of 820 items, and a test set, consisting of 80 items (40 concrete and 40 abstract). The 80 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 phonological neighborhood densities, and bigram frequency (Table 3).

The prediction analysis was conducted in four steps. First, for each participant, we generated activation maps for each of the five

Table 2
Lexical data [mean (standard deviation)] for words and pseudowords used in the study. Length: number of letters. Orth: number of orthographic neighbors. Orth_F: averaged frequency (per million) of the orthographic neighbors. N2_F: averaged frequency (per million) of the constrained bigrams for the wordform. N2_C: number of wordforms that share the same constrained bigrams. N3_F: averaged frequency (per million) of the constrained trigrams for the wordform. N3_C: number of wordforms that share the same constrained trigrams.

	Length	Orth	Orth_F	N2_F	N2_C	N3_F	N3_C
Words	5.99 (1.99)	4.12 (5.75)	54.4 (271)	964 (1087)	68.2 (62.6)	151 (240)	12.4 (19.5)
Pseudowords	5.99 (1.99)	4.26 (5.75)	43.4 (93)	976 (759)	68.7 (49.1)	141 (176)	11.4 (10.4)
T test (*p*)	0.98	0.71	0.49	0.86	0.91	0.51	0.40

Table 3
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 (Balota et al., 2007; http://elexicon.wustl.edu).

	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

attributes of word meaning (attribute maps, or AMs) based on the words in the modeling set (Fig. 2A). This was done by converting their attribute ratings into z-scores and including them as simultaneous predictor variables in a Generalized Least Squares (GLS) regression model. Word length, number of phonemes, number of syllables, word frequency, bigram frequency, orthographic and phonological neighborhood density, and the participant's z-transformed RT for each trial were included as nuisance regressors. Two binary regressors were also added to account for activity associated with early visual, orthographic, and phonological processing of the stimulus, as well as the subsequent motor response: one regressor coded for "word" events and the other for "pseudoword" events. Noise was modeled with linear, second-order, and third-order trends, as well as with the estimates of the motion parameters. For each attribute, a group-level AM was obtained by averaging the individual AMs (beta values) across participants.

Second, we computed predicted activation maps (predicted maps, PMs), for each of the 80 words in the test set, as linear combinations of the AMs, whereby each AM was weighted by the z-score of the word's corresponding attribute rating (Fig. 2B). The PM for a given test word corresponds 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).

Third, we generated activation maps for each of the 80 words in the test set, relative to a pseudoword baseline (observed maps, OMs; Fig. 2C). For each participant, a separate GLS regression was done for each of the 80 test words, with the following explanatory variables: (1) a binary regressor coding for the presentation of the selected test word; (2) a binary regressor coding for presentation of all the non-selected words (i.e., the other 899 words in the stimulus set); (3) a binary regressor coding for presentation of the pseudowords; (4) five continuous regressors coding for each of the five semantic attribute ratings for all non-selected words; and (5) a continuous regressor coding the response time for each trial. The resulting OM for a given test word, thus, revealed the unique activation pattern elicited by that word. For each test word, a group-level OM was obtained by averaging the individual OMs (beta values) across all participants.

Finally, we assessed the accuracy of the PMs by calculating the voxel-by-voxel pairwise correlation between each PM and each of the 80 OMs (Fig. 2D). For each PM, we then rank-ordered all the OMs according to correlation strength and noted the percentile rank of the OM for the corresponding target word (Fig. 2E). This percentile rank, scaled to a 0–1 range, was assigned to the PM as its accuracy score. Thus, each PM was assigned 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 instance, if the OM for the word "coffee" were the most highly correlated (among the 80 OMs) 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. We then conducted a one-tailed Wilcoxon signed rank test to verify whether the median prediction accuracy across all 80 words was significantly higher than chance (0.5).

In principle, any type of information about the stimuli that correlates with the attribute ratings could contribute to the prediction success of our encoding model. Thus, it is important to rule out other factors (e.g., word length) as possible drivers of prediction accuracy. For this purpose, 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 (i.e., the extent to which they referred to sensory-motor experiences). As Table 3 shows, inter-stimulus variance for the attribute ratings was much smaller among abstract than among concrete words. In other words, abstract words contained much less sensory-motor information. If prediction accuracy were indeed driven by the sensory-motor aspects of word meaning, performance should be higher for concrete than for abstract words. Finally, we also investigated the relative contribution of each of the five attributes to prediction accuracy (for concrete words) by evaluating encoding models based on all possible combinations of these attributes.

Voxel selection is an important step in MVPA, since the inclusion of non-informative voxels can severely degrade the model's performance (Cox and Savoy, 2003; Mitchell et al., 2004). We excluded non-brain voxels and voxels with low signal-to-noise ratio (SNR) by confining the analysis to a mask where the temporal SNR was higher than an arbitrary threshold (tSNR > 50). The resulting mask encompassed most of the brain, except for the regions that are typically affected by signal dropout (portions of the inferior temporal and orbitofrontal cortex).

3. Results

3.1. Analyses including all five attributes

When all 80 words were analyzed together, activation patterns were predicted significantly better than chance (median = 0.63, 95% CI [0.58, 0.77], $V = 2413.5$, $p = 0.00007$), showing that information about these five aspects of sensory-motor experience is encoded in the activation pattern elicited by lexical concepts (Fig. 3). Since concrete concepts contain more sensory-motor information than abstract concepts (as reflected in the larger variance of attribute ratings for the former; Table 3), we expected that prediction accuracy would be driven mainly by the concrete words in the test set. As shown in Fig. 3, when analyzed separately, concrete words were predicted significantly better than chance (median = 0.67, 95% CI [0.55, 0.81], $V = 610$, $p = 0.0036$), while abstract words were not (median = 0.50, 95% CI [0.29, 0.63], $V = 360.5$, $p = 0.75$). This was further confirmed by a Wilcoxon rank sum test, which showed that prediction was significantly higher for concrete than for abstract words ($W = 1048$, $p = 0.0086$). Since concrete and abstract words were closely matched on non-semantic lexical properties of the stimuli (Table 3), this result confirms that prediction accuracy was driven by semantic information alone.

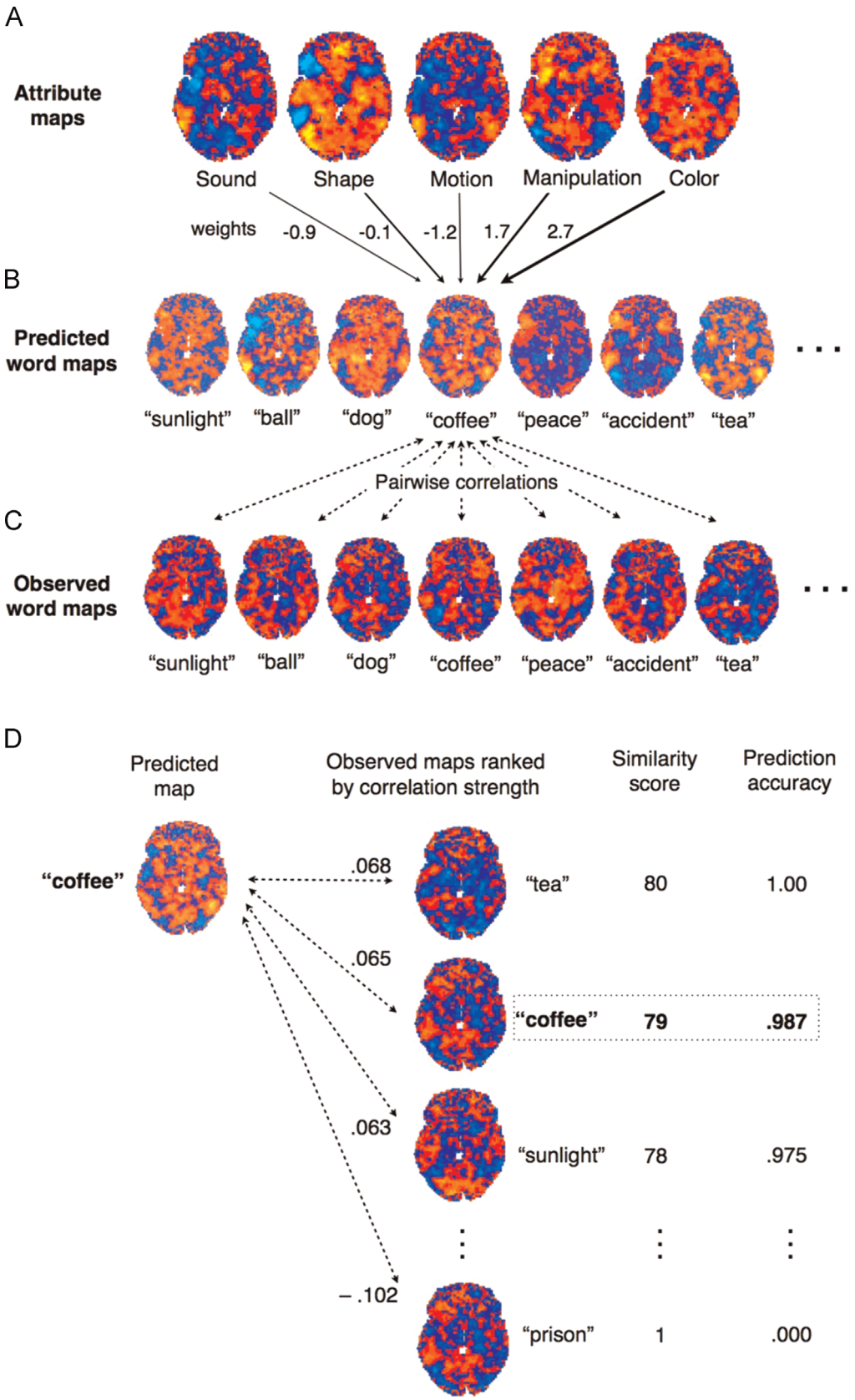

Fig. 2. Prediction analysis. (A) Group-level attribute maps were generated from the 820-word modeling set through least squares multiple regression. (B) In this example, the predicted map for "coffee" is generated by multiplying each attribute map by the z score of the corresponding attribute rating for that word and adding them together. (C) The voxel-by-voxel correlations between the predicted map for "coffee" and each of the 80 group-level observed maps are computed. (D) The observed maps are ranked by correlation strength, and prediction accuracy is determined from the percentile rank of the observed map for "coffee". Steps B–D are repeated for each word in the test set.

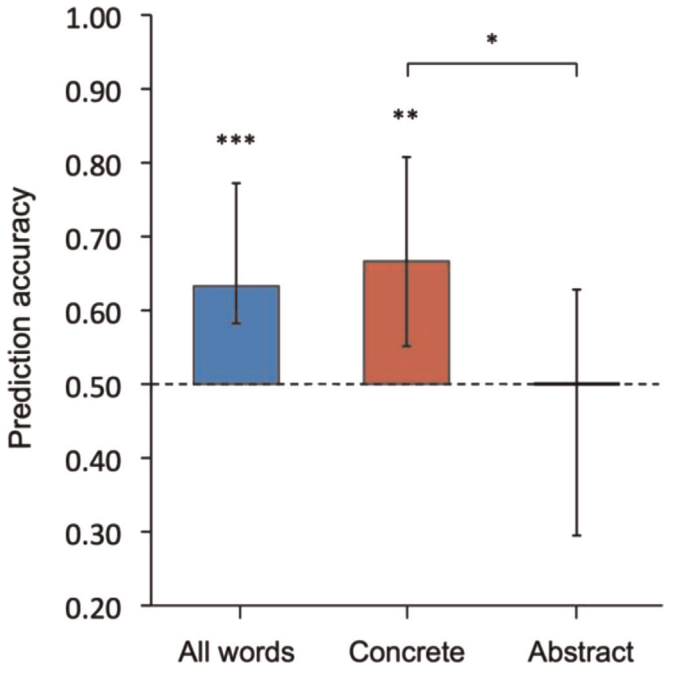

Fig. 3. Median prediction accuracy when all 5 attributes were included in the analysis. Error bars represent the 95% confidence interval. $^{*}p < .05$; $^{**}p < .005$; $^{***}p < .0005$.

3.2. Analyses on subsets of attributes

Since the predictive power of the model was specific to concrete words, we focused on this subset in the ensuing analyses. Prediction accuracies for models based on all possible combinations of the 5 attributes are shown in Fig. 4. When prediction maps were based on only one attribute, prediction accuracy across the 40 concrete words was indistinguishable from chance for all attributes, with no significant differences between them ($p=0.84$, one-way Kruskal–Wallis rank sum test), and with a mean (across attributes) of 0.53. When prediction was based on two attributes at a time, prediction accuracies were above chance level for most attribute dyads, except for Sound & Motion and for Manipulation & Motion, but with no significant differences between attribute pairs ($p=0.66$). Mean accuracy across pairs was 0.56. With three attributes included at a time, activation patterns were predicted better than chance for 5 out of the 10 triads, again with no significant differences between them ($p=0.40$), and with a mean accuracy of 0.60. Finally, when predictions were based on four attributes at a time, prediction was successful for all but one combination of attributes (the one in which Motion was left out). Mean prediction accuracy was 0.65, with no significant differences between the models ($p=0.49$). These results show that prediction accuracy increased as a function of the number of attributes included in the model (Fig. 5), but revealed no significant differences in the relative contribution of individual attributes.

4. Discussion

An encoding model based on the relative relevance of five sensory-motor attributes – sound, color, visual motion, manipulation and shape – to conceptual content was able to predict the distributed fMRI activation pattern elicited by concrete words. Prediction accuracy for a matched set of abstract words, on the other hand, was at chance level, confirming that prediction success for the concrete words was based solely on semantic information. These results show that the neural code for concept representation includes information about sensory-motor attributes. Finally, while prediction accuracy increased linearly with the number of sensory-motor attributes included in the model, we found no significant differences between the relative contributions of the attributes, suggesting that all of them contributed to prediction success.

These results make three main contributions to the existing literature on concept representation. First, while many previous studies focused on the important issue of whether activation in modal sensory or motor systems is modulated by sensory or motor conceptual content, they did not focus on the multimodal nature of the concept representation system. The present study offers the first evidence that the neural code for concept representation includes information about the relative relevance of specific aspects of sensory-motor experience, involving different modalities. Second, the present study establishes a novel method that can be used to assess the contribution of other attributes (sensory, motor, affective, temporal, spatial, social, etc.) to the neural representation of conceptual knowledge. Finally, our results show that the neural code for concepts is consistent enough across participants to allow successful prediction of activation patterns at the group level. This finding is important in the context of previous studies that have shown that the neural representation of a lexical concept is measurably affected by the individual's unique sensory-motor experiences during concept acquisition (Hoenig et al., 2011; Kiefer et al., 2007; Weisberg et al., 2007). These studies indicate that a lexical concept is not identically represented across different brains, but reflects each participant's unique life history. The relative extent of this inter-subject variability, however, remains an open issue. Is it large enough that inter-subject averaging would lead to complete loss of information? Our results indicate that, despite individual differences, the neural activation pattern associated with a lexical concept displays a certain degree of invariance across participants, making it possible to discriminate between concepts in group-averaged data. The generalizability of our results is further strengthened by the fact that the attribute ratings were derived from a separate group of participants.

The present approach to relating neural activation patterns to word meanings shares some similarities with previous studies on this topic but also differs in important ways. Mitchell et al. (2008) introduced the approach of predicting neural activation patterns for concepts using a linear combination of basis images weighted by semantic feature values. They showed that a set of 25 semantic features related to sensory-motor experience could predict neural activation associated with concrete concepts, in individual participants, with a mean accuracy of 0.77 (chance level=0.5). One major difference from the present study is that the semantic features of the test concepts were determined by text co-occurrence rather than explicit attribute judgments. For example, one semantic feature was the degree to which a test concept co-occurs with the word “eat” in a large text corpus. The features were thus defined by measuring co-occurrence with 25 words, primarily verbs: see, hear, listen, taste, smell, eat, touch, rub, lift, manipulate, run, push, fill, move, ride, say, fear, open, approach, near, enter, drive, wear, break, clean. Although some of these have strong relations to modality-specific neural processing systems (e.g., see, hear, listen, taste, smell, touch, manipulate, run, fear), many involve multiple sensory-motor attributes from various modalities (eat, rub, lift, fill, move, push, say, open, approach, enter) or are ambiguous as to experiential content (ride, drive, wear, break, clean). Thus the relationships between these features and particular neural processing systems are somewhat unclear. The co-occurrence method adds to this uncertainty in two ways. First, it is debatable whether co-occurrence with a set of verbs adequately captures the targeted content of a concept, since other verbs are available in most cases that could be used to convey the same

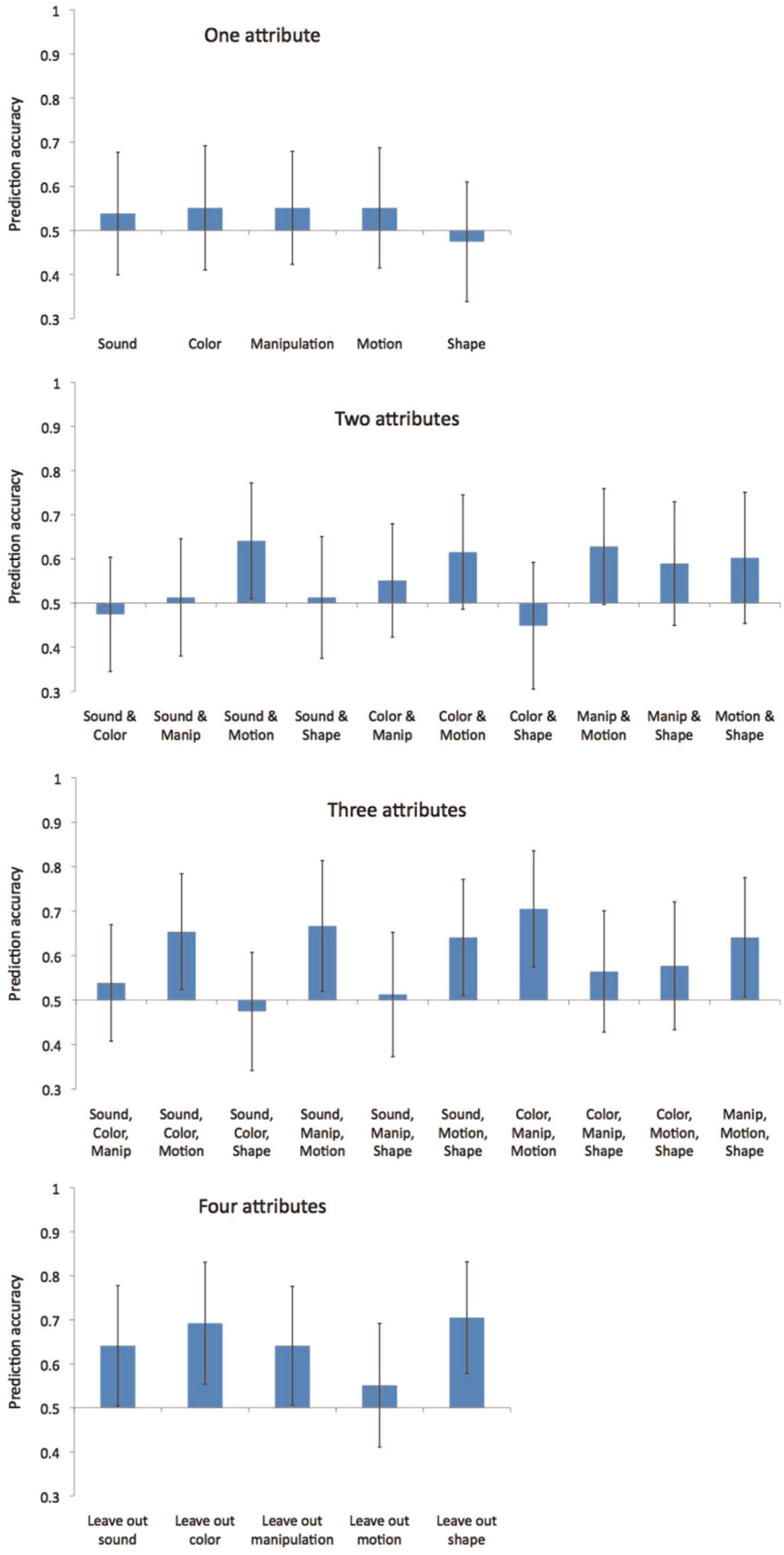

Fig. 4. Median prediction accuracy for concrete words, for different subsets of attributes. Error bars represent the 95% confidence interval.

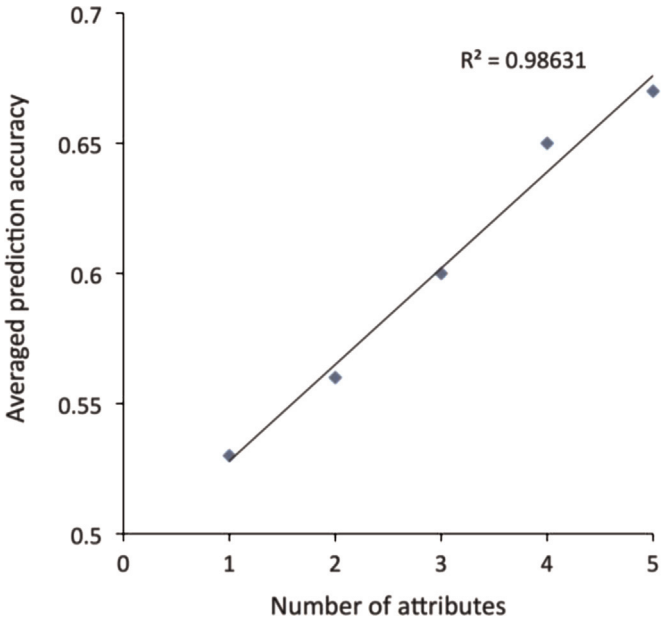

Fig. 5. Mean prediction accuracy for concrete words by number of attributes included in the model.

general meaning (e.g., "observe," "watch," "look," and "view" instead of "see"). Second, many of the words selected as co-occurrence targets have multiple meanings, some of which may be unrelated to the targeted neural modality. The target verb "run," for example, was likely selected to identify content related to lower limb motor actions, but usages such as "the car is running" ("car" was one of the test concepts) are unrelated to this targeted content.

The present approach attempts to remove such ambiguity by deriving the semantic representation of concepts directly from human judgments about attribute salience. Rather than inferring semantic content indirectly, the rating method asks human observers to make explicit links between concepts and experimenter-defined attributes. The limitation of this approach is that human judgments are inevitably subjective and open to error, although averaging responses from a sample of observers after removal of intra-class outliers mitigates this problem to some degree. Hoffman and Lambon Ralph (2013) provided evidence supporting the validity of such an approach. They obtained salience ratings on 8 attributes – visual form, observed motion, color, touch, sound, taste, smell, and performed actions – for 160 object concepts. These attribute representations predicted lexical processing time, and did so more accurately than traditional feature-based representations. It remains an open question, however, whether attribute-based representations derived from human judgments are more or less accurate than text-based co-occurrence measures at predicting neural activity patterns (for a review of the differences between approaches to word semantics based on explicit sensory-motor features and those based on statistical distributions of word co-occurrences – and how they can be integrated – see Andrews et al., 2014).

In a related study, Huth et al. (2012) investigated the cortical representation of semantic categories based on BOLD responses to natural movies. In that study, participants watched movie clips displaying a variety of objects, people, animals and actions, which were coded in the fMRI regression model using 1705 taxonomic category labels from the WordNet lexicon. The authors then identified four significant factors underlying the model fit using principal components analysis. The maps based on these four components showed a degree of correspondence between categorical structure and cortical representation (i.e., similar categories tended to be represented by similar sets of voxels) in individual participants. However, the underlying factors in such an analysis must be interpreted *post-hoc*, and the resulting maps offer little insight into how those factors are related to other known functions of the corresponding cortical areas. Our approach focused instead on five sensory-motor dimensions involved in concept acquisition (at least for concrete concepts), and investigated the extent to which these dimensions capture the neural representation of lexical concepts. This approach offers a representational space for the description of concepts in which the dimensions can be related directly to known functions of the brain.

Perhaps the most important difference between our study and those of Mitchell et al. (2008) and Huth et al. (2012) is that concepts in our study were cued solely by word stimuli, with no pictorial representations of the concepts themselves. Mitchell et al., on the other hand, presented word-picture pairs as stimuli (e.g., the word "airplane" accompanied by a line drawing of an airplane), while Huth et al. presented movie clips. The use of pictorial stimuli in these studies leaves open the possibility that the elicited activation patterns in part reflected perceptual rather than purely semantic aspects of the objects (Haxby et al., 2001, 2011; Spiridon and Kanwisher, 2002; Cox and Savoy, 2003; Kriegeskorte et al., 2008). Since similar categories of objects tend to share similar visual properties, similarity in cortical representation could have been driven to a large extent by perceptual similarity. The present study avoided this confound by relying exclusively on word stimuli, thus ensuring that any sensory-related information contributing to prediction accuracy originated from the concept representation itself, rather than the physical stimulus used to cue it. This was further confirmed by the contrasting results obtained for concrete and abstract words.

A final difference between studies is that our analysis was done on group-averaged activation maps, and thus relied entirely on concept-related neural activation patterns that are common across individuals. This approach offers the potential of predicting a participant's activation pattern for a concept based solely on attribute maps generated from other participants. However, the averaging procedure certainly results in loss of information about inter-subject variability, and this may explain in part why the accuracy levels obtained in the present study are generally lower than those obtained by Mitchell et al. (2008). Another factor that probably contributed to the difference in accuracy was the number of semantic features used to generate the predictions (5 in the present study, 25 in Mitchell et al.). Presumably, increasing the number of attributes in our analysis would lead to higher prediction accuracies, provided the extra attributes capture significant aspects of the concept's neural representation.

One potential criticism of the present study is that the task may have induced participants to engage in mental imagery, and that the activation patterns may thus reflect perceptual, rather than conceptual, representations. This argument is based on the assumption that mental imagery and concept retrieval constitute categorically distinct kinds of phenomena, such that imagery engages perceptual representations while conceptual processing does not (e.g., Machery, 2007). In light of this criticism, previous studies have taken steps to minimize conscious, deliberate conceptual elaboration and imagery by using more implicit tasks such as lexical decision and masked word presentation, thus focusing on the automatic, unconscious aspects of conceptual processing (e.g., Kiefer et al., 2012; Trumpp et al., 2013; Willems et al., 2010).

While these studies have been important in demonstrating that early automatic processes in concept retrieval also involve sensory-motor representations, we see the categorical distinction between concepts and imagery as an artificial dichotomy. It seems

more likely that concept retrieval can involve a variable amount of sensory-motor reenactment, where the extent of the reenactment depends on the specific demands of the task. We know, for instance, that word imageability (a direct measure of how readily mental images come to mind) affects behavior even on "shallow" tasks such as lexical decision (e.g., Evans et al., 2012), suggesting that "images" are activated to some degree even during such tasks. When more extensive or detailed sensory-motor information is required (e.g., during narrative comprehension), retrieval of this information may be experienced as conscious imagery. In all cases, however, sensory-motor information about a concept is being retrieved. Thus, we see automatic, unconscious activation of conceptual features, on the one hand, and deliberate, vivid imagery, on the other, as two ends of a continuous range of conceptual knowledge retrieval. Studies that have focused on implicit concept retrieval have made a fundamental contribution to our understanding of conceptual processing; however, implicit activation of conceptual features is only one aspect of a larger, more complex phenomenon.

5. Conclusions and future directions

Our results show that the distributed cortical representation of concrete lexical concepts can be predicted, to a considerable extent, by information about the relative relevance of five attributes of sensory-motor experience – sound, color, visual motion, shape and manipulation – to the content of those concepts. Furthermore, these results show that the activation patterns elicited by concrete concepts are consistent enough across participants to be useful in predicting group-averaged data.

This approach advances our understanding of the neural substrates of conceptual knowledge by framing individual concepts as points in a multidimensional representational space, defined by elementary features of phenomenal experience that can be mapped onto known functions of the brain. By adding new attributes to the encoding model, we hope to determine which ones capture relevant aspects of this representational space based on how much they contribute to prediction accuracy. By restricting the analysis to different brain regions, we also plan to explore how their contributions to concept representation relate to their known functions in other domains.

Acknowledgments

We would like to thank two anonymous reviewers for their helpful comments and suggestions.

This project was supported by National Institute of Neurological Diseases and Stroke Grant R01 NS33576, by National Institutes of Health General Clinical Research Center Grant M01 RR00058, and by National Institute of General Medical Sciences Grant T32 GM89586.

References

Andrews, M., Frank, S., Vigliocco, G., 2014. Reconciling embodied and distributional accounts of meaning in language. Top. Cognit. Sci. 6 (3), 359–370. http://dx.doi.org/10.1111/tops.12096.

Aziz-Zadeh, L., Wilson, S.M., Rizzolatti, G., Iacoboni, M., 2006. Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Curr. Biol. 16 (18), 1818–1823. http://dx.doi.org/10.1016/j.cub.2006.07.060.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R., 2007. The English Lexicon Project. Behav. Res. Methods 39, 445–459.

Barsalou, L.W., 2008. Grounded cognition. Annu. Rev. Psychol. 59, 617–645. http://dx.doi.org/10.1146/annurev.psych.59.103006.093639.

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends Cognit. Sci. 15 (11), 527–536. http://dx.doi.org/10.1016/j.tics.2011.10.001.

Bird, H., Franklin, S., Howard, D., 2001. Age of acquisition and imageability ratings for a large set of words, including verbs and function words. Behav. Res. Methods Instrum. Comput. 33 (1), 73–79.

Brysbaert, M., Warriner, A.B., Kuperman, V., 2014. Concreteness ratings for 40 thousand generally known English word lemmas. Behav. Res. Methods 46 (3), 904–911. http://dx.doi.org/10.3758/s13428-013-0403-5.

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. http://dx.doi.org/10.1038/13217.

Clark, J.M., Paivio, A., 2004. Extensions of the Paivio, Yuille, and Madigan (1968) norms. Behav. Res. Methods, Instrum. Comput. 36 (3), 371–383.

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 (2 Pt 1), 261–270.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. Int. J. 29 (3), 162–173.

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.

Evans, G.A.L., Lambon Ralph, M.A., Woollams, A.M., 2012. What's in a word? a parametric study of semantic influences on visual word recognition. Psychon. Bull. Rev. 19 (2), 325–331. http://dx.doi.org/10.3758/s13423-011-0213-7.

Farah, M.J., McClelland, J.L., 1991. A computational model of semantic memory impairment: modality specificity and emergent category specificity. J. Exp. Psychol.: Gen. 120 (4), 339–357. http://dx.doi.org/10.1037/0096-3445.120.4.339.

Fernandino, L., Binder, J.R., Desai, R.H., Pendl, S.L., Humphries, C.J., Gross, W.L., Connant, L.L., Seidenberg, M.S., 2015. Concept representation reflects multimodal abstraction: a framework for embodied semantics. Cereb. Cortex . http://dx.doi.org/10.1093/cercor/bhv020.

Fischer, M.H., Zwaan, R.A., 2008. Embodied language: a review of the role of the motor system in language comprehension. Q. J. Exp. Psychol. 61 (6), 825–850. http://dx.doi.org/10.1080/17470210701623605.

Glenberg, A.M., Gallese, V., 2012. Action-based language: a theory of language acquisition, comprehension, and production. Cortex 48 (7), 905–922. http://dx.doi.org/10.1016/j.cortex.2011.04.010.

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

Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37 (1), 435–456. http://dx.doi.org/10.1146/annurev-neuro-062012-170325.

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. http://dx.doi.org/10.1126/science.1063736.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., et al., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72 (2), 404–416. http://dx.doi.org/10.1016/j.neuron.2011.08.026.

Hoenig, K., Müller, C., Herrnberger, B., Sim, E.-J., Spitzer, M., Ehret, G., Kiefer, M., 2011. Neuroplasticity of semantic representations for musical instruments in professional musicians. NeuroImage 56 (3), 1714–1725. http://dx.doi.org/10.1016/j.neuroimage.2011.02.065.

Hoffman, P., Lambon Ralph, M.A., 2013. Shapes, scents and sounds: quantifying the full multi-sensory basis of conceptual knowledge. Neuropsychologia 51 (1), 14–25. http://dx.doi.org/10.1016/j.neuropsychologia.2012.11.009.

Hsu, N.S., Frankland, S.M., Thompson-Schill, S.L., 2012. Chromaticity of color perception and object color knowledge. Neuropsychologia 50 (2), 327–333. http://dx.doi.org/10.1016/j.neuropsychologia.2011.12.003.

Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224. http://dx.doi.org/10.1016/j.neuron.2012.10.014.

Kiefer, M., Pulvermüller, F., 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. Cortex 48 (7), 805–825. http://dx.doi.org/10.1016/j.cortex.2011.04.006.

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., Hoenig, K., 2008. The sound of concepts: four markers for a link between auditory and conceptual brain systems. J. Neurosci. 28 (47), 12224–12230. http://dx.doi.org/10.1523/JNEUROSCI.3579-08.2008.

Kiefer, M., Sim, E.-J., Liebich, S., Hauk, O., Tanaka, J., 2007. Experience-dependent plasticity of conceptual representations in human sensory-motor areas. J. Cognit. Neurosci. 19 (3), 525–542. http://dx.doi.org/10.1162/jocn.2007.19.3.525.

Kiefer, M., Trumpp, N., Herrnberger, B., Sim, E.-J., Hoenig, K., Pulvermüller, F., 2012. Dissociating the representation of action- and sound-related concepts in middle temporal cortex. Brain Lang. 122 (2), 120–125. http://dx.doi.org/10.1016/j.bandl.2012.05.007.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., et al., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60 (6), 1126–1141. http://dx.doi.org/10.1016/j.neuron.2008.10.043.

Machery, E., 2007. Concept empiricism: a methodological critique. Cognition 104 (1), 19–46. http://dx.doi.org/10.1016/j.cognition.2006.05.002.

Martin, A., Wiggs, C.L., Ungerleider, L.G., Haxby, J.V., 1996. Neural correlates of category-specific knowledge. Nature 379 (6566), 649–652. http://dx.doi.org/10.1038/379649a0.

McCarthy, R.A., Warrington, E.K., 1988. Evidence for modality-specific meaning systems in the brain. Nature 334 (6181), 428–430. http://dx.doi.org/10.1038/334428a0.

Medler, D.A., Binder, J.R., 2005. MCWord: An On-Line Orthographic Database of the English Language. http://www.neuro.mcw.edu/mcword/.

Meteyard, L., Vigliocco, G., 2008. The role of sensory and motor information in semantic representation In: Calvo, P., Gomila, T. (Eds.), Handbook of Cognitive Science: An Embodied Approach. Elsevier, San Diego, Oxford, Amsterdam.

Meteyard, L., Cuadrado, S.R., Bahrami, B., Vigliocco, G., 2012. Coming of age: a review of embodiment and the neuroscience of semantics. Cortex 48 (7), 788–804. http://dx.doi.org/10.1016/j.cortex.2010.11.002.

Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., 2004. Learning to decode cognitive states from brain images. Machine Learning 57, 145–175.

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 (5880), 1191–1195. http://dx.doi.org/10.1126/science.1152876.

Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2011. Encoding and decoding in fMRI. NeuroImage 56 (2), 400–410. http://dx.doi.org/10.1016/j.neuroimage.2010.07.073.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.

Spiridon, M., Kanwisher, N., 2002. How distributed is visual category information in human occipito-temporal cortex? an fMRI study. Neuron 35 (6), 1157–1165.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Thieme, New York.

Trumpp, N.M., Traub, F., Kiefer, M., 2013. Masked priming of conceptual features reveals differential brain activation during unconscious access to conceptual action and sound information. PLoS One 8 (5), 1–10. http://dx.doi.org/10.1371/journal.pone.0065910.

Weisberg, J., van Turennout, M., Martin, A., 2007. A neural system for learning about object function. Cereb. Cortex 17 (3), 513–521. http://dx.doi.org/10.1093/cercor/bhj176.

Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107 (Pt 3), 829–854.

Willems, R.M., Toni, I., Hagoort, P., Casasanto, D., 2010. Neural dissociations between action verb understanding and motor imagery. J. Cognit. Neurosci. 22 (10), 2387–2400. http://dx.doi.org/10.1162/jocn.2009.21386.

Wilson, M., 1988. MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. Behav. Res. Methods Instrum. Comput. 20 (1), 6–10. http://dx.doi.org/10.3758/BF03202594.