Psychological Science
2021, Vol. 32(10) 1617–1635

DOI: 10.1177/09567976211003877
www.psychologicalscience.org/PS

Idiosyncratic Tower of Babel: Individual Differences in Word-Meaning Representation Increase as Word Abstractness Increases

Xiaosha Wang[1,2,3] and Yanchao Bi[1,2,3,4]
[1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University; [2]IDG/McGovern Institute for Brain Research, Beijing Normal University; [3]Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University; and [4]Chinese Institute for Brain Research, Beijing, China

Abstract
Humans primarily rely on language to communicate, on the basis of a shared understanding of the basic building blocks of communication: words. Do we mean the same things when we use the same words? Although cognitive neural research on semantics has revealed the common principles of word-meaning representation, the factors underlying the potential individual variations in word meanings are unknown. Here, we empirically characterized the intersubject consistency of 90 words across 20 adult subjects (10 female) using both behavioral measures (rating-based semantic-relationship patterns) and neuroimaging measures (word-evoked brain activity patterns). Across both the behavioral and neuroimaging experiments, we showed that the magnitude of individual disagreements on word meanings could be modeled on the basis of how much language or sensory experience is associated with a word and that this variation increases with word abstractness. Uncovering the cognitive and neural origins of word-meaning disagreements across individuals has implications for potential mechanisms to modulate such disagreements.

Keywords
intersubject consistency, word meaning, functional MRI, language, sensory experience, open data, open materials

Received 10/1/20; Revision accepted 2/27/21

Human beings transfer thoughts across individuals, time, and space using language. We often assume that differences in thoughts are reflected by different choices of words and that speakers of the same language have common conceptual understandings about the basic word elements. Such commonality is the basis of effective learning and communication, and word-meaning misalignment is usually discussed only within the context of cross-language speakers (Jackson et al., 2019; Thompson et al., 2020). However, the individual variations in how people understand a word within a language have intrigued classical philosophers (Locke, 1690; Russell, 1948). Indeed, it has recently been empirically shown that there are intersubject variations in understanding politically or emotionally related words, which are associated with related domains of nonlinguistic processing such as political position (Li et al., 2017) or emotional perception (Brooks & Freeman, 2018). It is unknown whether this is specific to these "subjective" domains or is a general mechanism of word-meaning representation. Here, using both behavioral and neural signatures, we empirically quantified the consistency and variations of word-meaning representations across speakers of the same language and from a relatively homogeneous culture and education group, and we investigated the underlying mechanisms leading to individual variation.

The nature of and variables affecting individual variation in word meaning are intrinsically related to the general principles of how word meanings are represented in the human brain. Meaningful variance in a system stems from the dimensions that make up the

Corresponding Author:
Yanchao Bi, Beijing Normal University, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research
Email: ybi@bnu.edu.cn

corresponding representation. For decades, research has focused on the common cognitive and neural basis of semantic (or conceptual) representations, converging on the consensus that these representations are compositional, entailing salient sensory, motor, and emotion-related attributes, and distributed over multiple systems of the cortex, despite the controversies about the sufficiency and necessity of the specific constituents (Binder et al., 2016; Lambon Ralph et al., 2017; Martin, 2016). Words referring to concrete objects comprise more specific sensorimotor attributes (e.g., the shape of a cup, the action associated with a cup), among other attributes, and tend to more strongly activate regions in the corresponding sensorimotor and association cortices (Fernandino et al., 2016; Martin, 2016; J. Wang et al., 2010). Abstract words (e.g., *i* , *i*), by comparison, tend to be associated with socioemotional attributes and depend more on linguistic context, and they more strongly activate language-related regions such as anterior temporal and inferior frontal cortices (Binder et al., 2016; Hoffman et al., 2013; Kousta et al., 2011; Schwanenflugel & Shoben, 1983; J. Wang et al., 2010). However, recent evidence suggests that words referring to external referents may also entail language-derived representations (Striem-Amit et al., 2018; X. Wang et al., 2020).

These current semantic theories do not postulate explicit hypotheses about individual variability, and it is not obvious what predictions can be generated without additional assumptions about the relationship between the underlying dimension compositions and the individual variation patterns. Is having richer properties of a particular attribute associated with greater or smaller variations? Consider the contrasts between words that have external referents (i.e., concrete words) and words that do not (i.e., abstract words). Although having external referents may boost consistency (through a common constraint), it is also possible that the knowledge about such referents is (at least partly) represented through sensorimotor experiences, which vary across individuals and actually introduce additional sources of variation. Furthermore, do various types of attributes themselves differ in their degree of intersubject variation, thus having different effects on a word's individual variations? With these theoretical and empirical possibilities, the approach here was to glean the potential organizational dimensions of word meanings from the current semantic theories and test the patterns in which these factors might account for individual consistency, including which dimensions produce significant effects and in what direction. Positive results would provide convergent evidence that the postulated dimension indeed effectively underlies meaning representation and that theories that do not incorporate those dimensions are to be challenged. Further, positive results would reveal the patterns of relationships of these dimensions and intersubject variations in word meaning.

Measuring people's internal representation of word meaning is notoriously challenging. Explicit-definition approaches are highly controversial (Marggolis & Laurence, 1999). The feature-based view makes the feature-listing approach appealing; this approach has been applied to test representations of object word meaning (Binder et al., 2016; Cree & McRae, 2003; Tyler & Moss, 2001), but it is very difficult to apply it to non-object words (Barsalou & Wiemer-Hastings, 2005). One widely adopted approach is to represent a word (at least partly) by its relationships with other words, which has been productive in natural-language processing (e.g., Landauer & Dumais, 1997; Mitchell et al., 2008; Thompson et al., 2020). This approach can be accomplished by subjective distance ratings in individual human subjects (Brooks & Freeman, 2018; Li et al.,

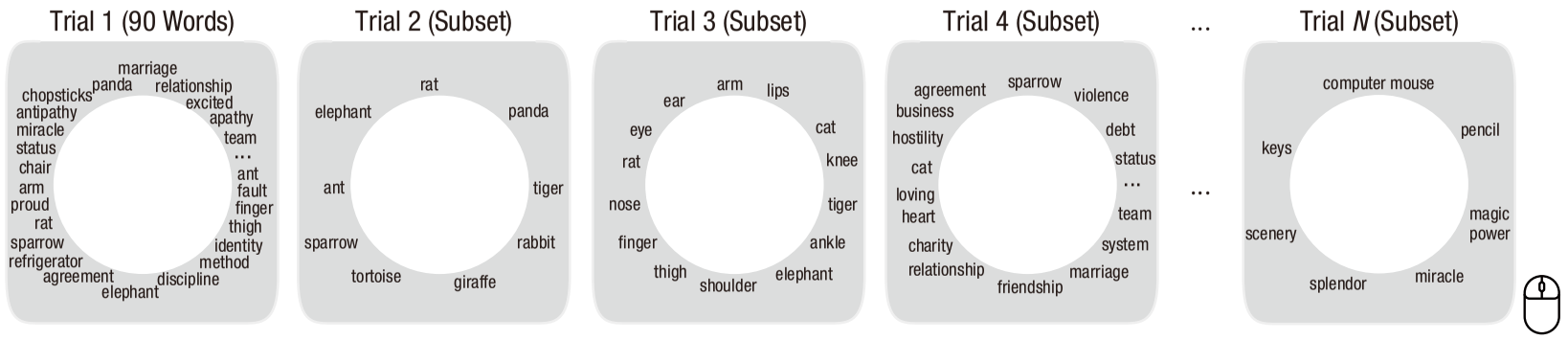
domains in which cognitive and brain mechanisms have been extensively studied; these stimuli were drawn from domains that have external sensory referents (varying in sensory and motor-related attributes: animals, face/body parts, and artifacts) and those that do not have specific external referents (nonobject abstract words with and without emotional associations, e.g., *i* vs. , respectively). We quantified their representations in all subjects (Chinese college students in Beijing) on the basis of both behavioral judgments (Experiment 1) and brain activation patterns measured by functional neuroimaging (Experiment 2), and we computed the intersubject consistency (ISC) for each word from behavioral data (ISC-behavior data) and brain data (ISC-brain data). We then asked independent groups of subjects to rate the extent to which each word was associated with each key representational dimension and examined how the ISC values across 90 words could be predicted by their rating means or variations (indexed by the standard deviation of ratings) of each dimension.

Method

Subjects

Twenty-one young, healthy college students (11 female; age: $M = 21.1$ years, range = 18–26 years) were recruited from several universities in Beijing for the study. All

a


Trial 1 (90 Words)
Trial 2 (Subset)
Trial 3 (Subset)
Trial 4 (Subset)
...
Trial N (Subset)
marriage
panda
relationship
chopsticks
excited
antipathy
apathy
miracle
team
status
...
chair
ant
arm
fault
proud
finger
rat
thigh
sparrow
identity
refrigerator
method
agreement
discipline
elephant
rat
elephant
panda
ant
tiger
sparrow
rabbit
tortoise
giraffe
arm
lips
ear
cat
eye
knee
rat
nose
tiger
finger
ankle
thigh
shoulder
elephant
agreement
sparrow
violence
business
debt
hostility
status
cat
...
loving
heart
team
charity
system
relationship
marriage
friendship
computer mouse
pencil
keys
magic
power
scenery
splendor
miracle

b

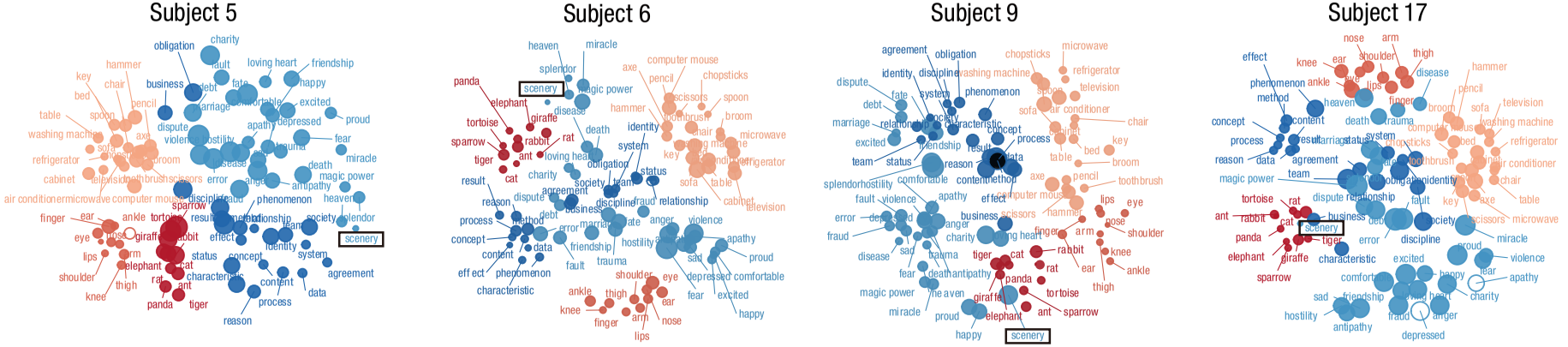
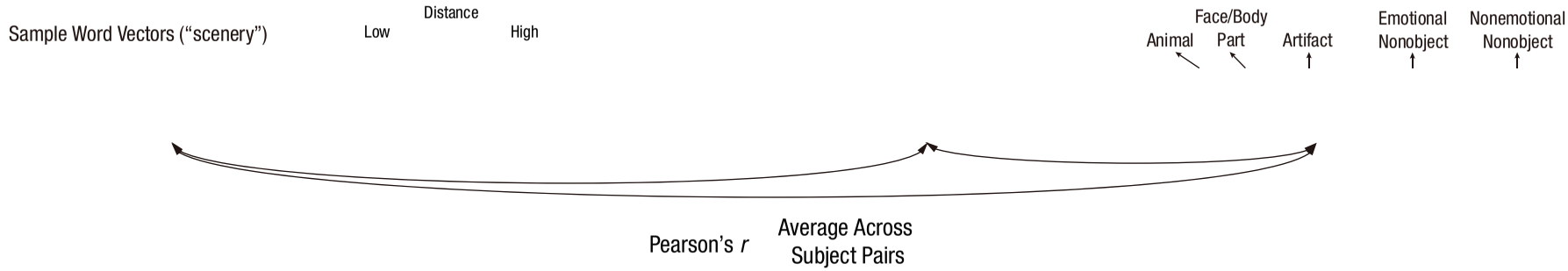

Subject 5
Subject 6
Subject 9
Subject 17
scenery

Fig. 1. (*i*)

C


Distance
Low
High
Sample Word Vectors (“scenery”)
Face/Body
Animal
Part
Artifact
Emotional
Nonobject
Nonemotional
Nonobject
Pearson’s r
Average Across
Subject Pairs

the words. In subsequent trials, subjects were shown adaptively selected word subsets that had been clustered together in previous trials, producing partial distance matrices. The task lasted for 1 hr, during which subjects completed various numbers of trials (M = 85, D = 71, range = 24–284). The final distance measure for each subject was calculated as the weighted average of distance measures of their multiple arrangements. Multidimensional scaling was carried out to visualize individual semantic distance matrices (number of dimensions [ndim] = 2, type = interval) using the package (de Leeuw & Mair, 2009) in the R programming environment (Version 4.0.0; R Core Team, 2020).

Word-level ISC-behavior computation. To compute the word-level ISC in behavior for each subject, we represented each word as an 89-dimensional vector of its semantic distance with the remaining words. Pearson's correlations of the word vector among each pair of subjects were then computed, Fisher transformed, and averaged across 190 subject pairs (20 subjects in total) to obtain ISC-behavior data for each word. The standard error of the ISC for behavior was assessed in two approaches: (a) bootstrapping the subject set with replacement 10,000 times, which evaluated ISC robustness across subjects, and (b) bootstrapping the word set with replacement 10,000 times, which evaluated ISC robustness across words included for judgment.

Validation of words' ISC-behavior computation. One issue that needed to be considered was whether a particular word's ISC-behavior pattern was affected by our choices of base words in its semantic-vector construction. In the main analyses using the 90-word set, for each word, the base words were the other 89 words (N – 1); the base words covered a wide range of words with varying types of relations with the word in consideration (both taxonomic and nontaxonomic neighbors). In this way, for each word under consideration, its 89 base words varied slightly (in a leave-one-out fashion). This validation analysis was then further conducted to check whether the ISC results obtained in this way were robust across different kinds of base-word list selections, especially when the common set of base words was used. We performed split-half analyses so that for each of the 45 words in the first half, the 45 words in the second half became the base words for its semantic vector (i.e., no leave-one-out method needed). ISC values were then computed from these data. This procedure was repeated 10,000 times.

Experiment 2: word-level ISC based on brain activation patterns

Task fMRI procedure. A condition-rich fMRI design was adopted to obtain activity patterns for each word (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). During the fMRI task (Fig. 2a), subjects were instructed to view each of 90 target words, think about their meanings, and perform an oddball one-back semantic judgment task. In the latter, subjects were instructed to determine whether occasional words in red were semantically related to the previous word by pressing buttons with their right index finger or middle finger (catch trials). There were 10 runs (360 s per run). Each run consisted of ninety 2.5-s-long word trials (0.8-s word followed by 1.7-s fixation), fourteen 2.5-s-long catch trials, and thirty 2.5-s-long null trials; the mean interval between two words was 3.23 s. Each target word appeared once within each run; the order of 90 target words was randomized in each run for each subject. Each run began with a 12-s fixation period and ended with a 13-s rest period during which subjects saw a verbal cue that the current run was about to end.

Image acquisition. All functional and structural MRI data were collected using a Siemens Prisma 3T scanner with a 64-channel head-neck coil at the Center for MRI Research, Peking University. Functional data were acquired with a simultaneous multislice echoplanar-imaging sequence supplied by Siemens (62 axial slices, repetition time [TR] = 2,000 ms, echo time [TE] = 30 ms, multiband factor = 2, flip angle [FA] = 90°, field of view [FOV] = 224 mm × 224 mm, matrix size = 112 × 112, slice thickness = 2 mm, gap = 0.2 mm, and voxel size = 2 mm × 2 mm × 2.2 mm). A high-resolution 3D T1-weighted anatomical scan was acquired using the magnetization-prepared rapid-acquisition gradient-echo sequence (192 sagittal slices, TR = 2,530 ms, TE = 2.98 ms, inversion time = 1,100 ms, FA = 7°, FOV = 224 mm × 256 mm, matrix size = 224 × 256 interpolated to 448 × 512, slice thickness = 1 mm, and voxel size = 0.5 mm × 0.5 mm × 1 mm).

Data preprocessing. Functional images were preprocessed using Statistical Parametric Mapping (SPM) software (Version 12; Wellcome Trust Center for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm12/). For each individual subject, the first four volumes of each functional run were discarded to reach signal equilibrium. The remaining images were corrected for slice timing and head motion and spatially normalized to Montreal Neurological Institute (MNI) space via unified segmentation (resampling into 2 mm × 2 mm × 2 mm voxel size). No subject had head motion larger than 2 mm/2°. These images were directly submitted to general linear models (GLMs) for multivariate pattern analyses and were further spatially smoothed using a 6-mm full-width half-maximum Gaussian kernel for univariate contrast analyses.

Computation of whole-brain activation patterns for each word. Whole-brain activation patterns for each word were obtained using a GLM with spatially

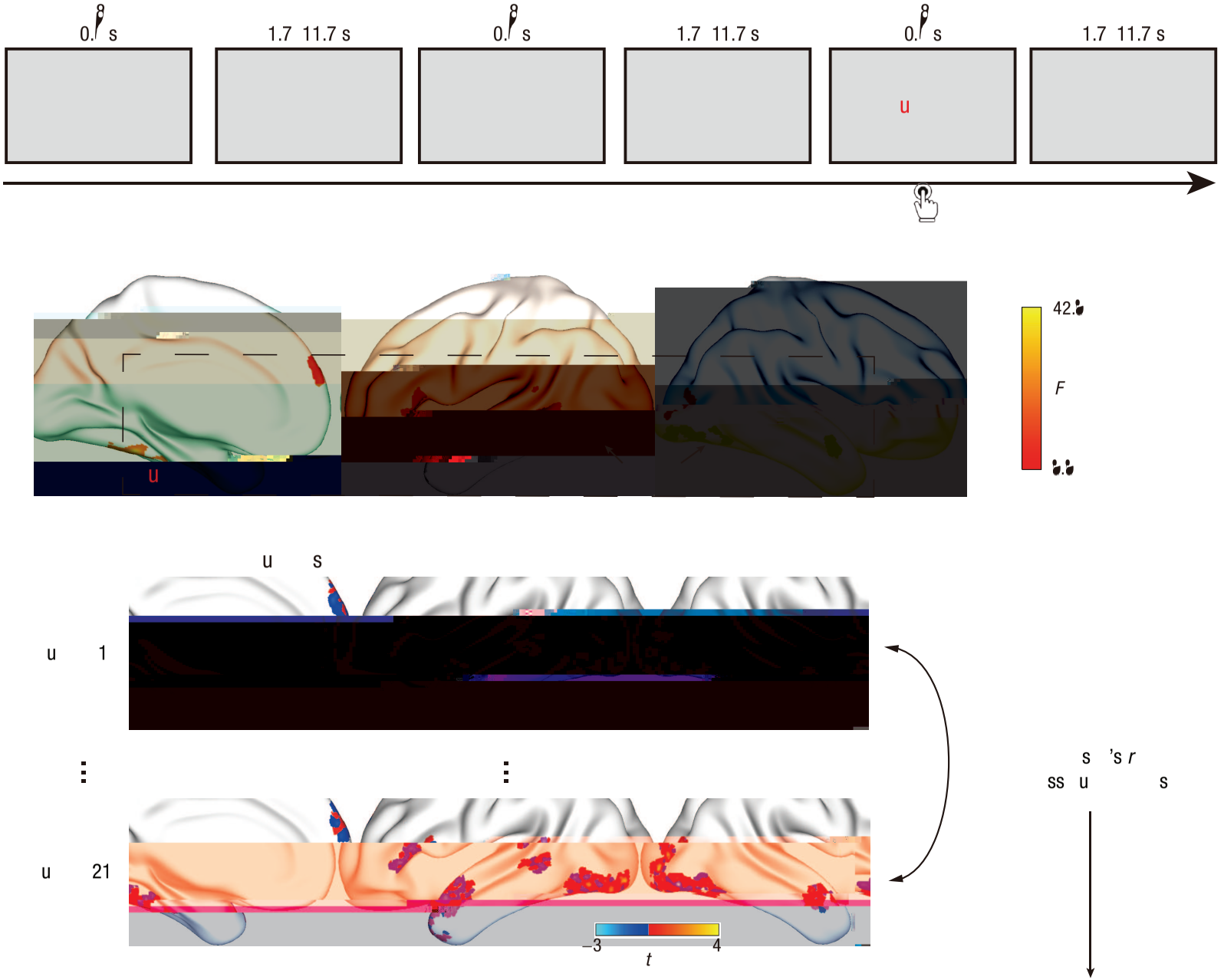
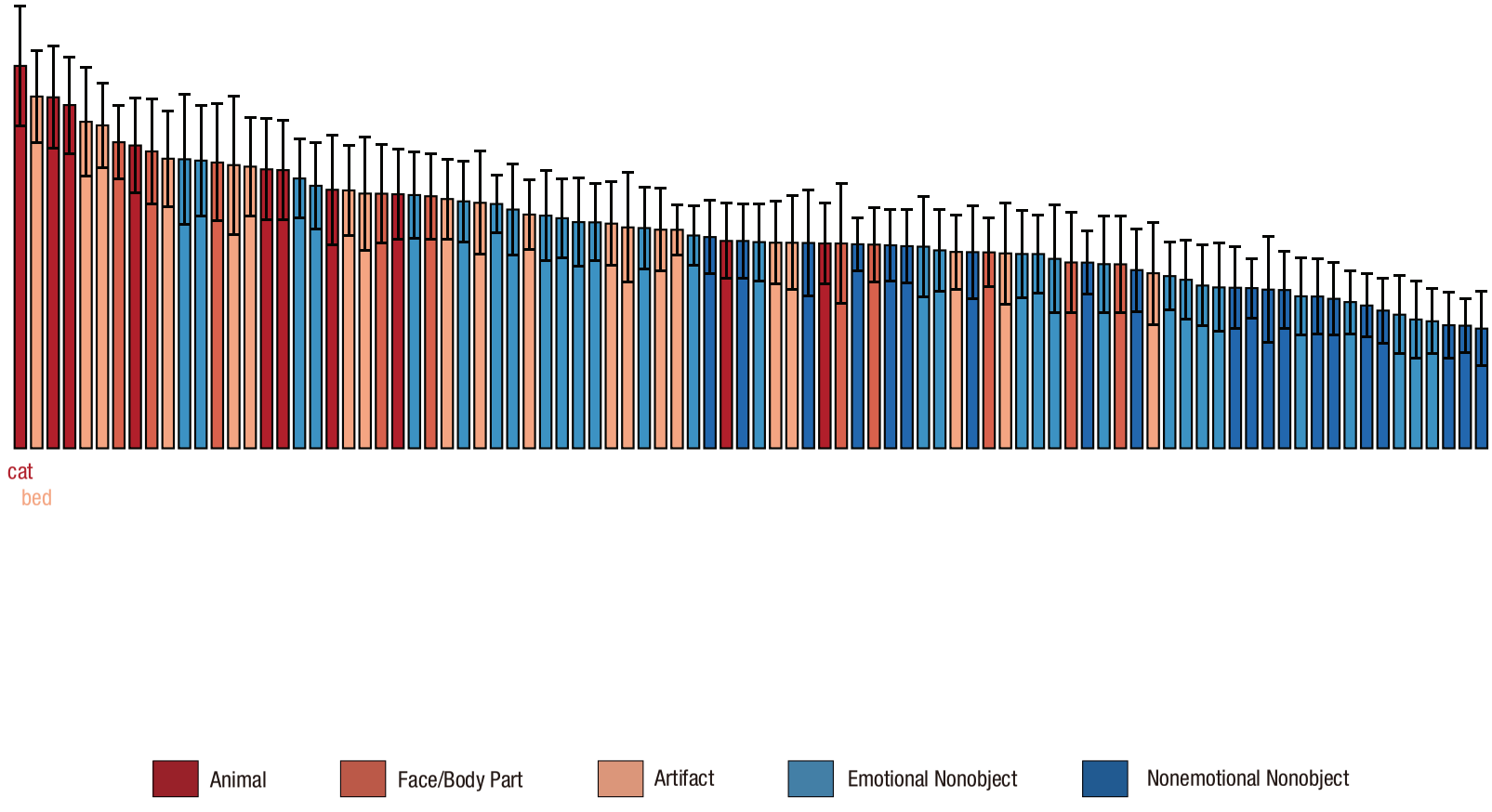


Fig. 2. (*i*)


cat
bed
Animal
Face/Body Part
Artifact
Emotional Nonobject
Nonemotional Nonobject

normalized, unsmoothed functional images. For each subject, the GLM for each run contained 90 regressors corresponding to the onset of each target word and one regressor indicating catch trials, convolved with a canonical hemodynamic response function, and six head-motion parameters. A high-pass-filter cutoff was set at 128 s. The resulting maps for each target word versus baseline were used to compute the ISC-brain data.

Word-level ISC-brain data computation. The procedure for ISC-brain computation consisted of the following steps: (a) Define word-associated voxels; (b) extract activation patterns of each word from these voxels in each subject; and (c) compute, for each word, the Pearson's correlations of activation patterns for each pair of subjects, which were Fisher transformed and averaged across subject pairs to obtain the ISC-brain data for this word. The key step here was the definition of word-associated regions, given that it is not necessarily obvious what voxels contain information about word (meaning) representations. We thus adopted multiple approaches that are described below to validate the robustness of the ISC-brain results. Functional activation maps were assessed at a voxel-wise threshold of $< .005$, family-wise error (FWE)-corrected cluster-extent $< .05$, unless explicitly stated otherwise.

In Approach 1, we defined word-related regions as those sensitive to major meaning differences between object words and nonobject words. For each subject, we built a GLM with spatially smoothed functional images and included two regressors corresponding to the onset of each word type (i.e., object or nonobject) and one regressor for catch trials, together with six head-motion parameters, for each run. The object-versus-nonobject contrast was computed, and the resulting β-weight images were submitted to an *F* test at the group level to identify the voxels whose activations were significantly different between object words and nonobject words.

In Approach 2, word-related regions were defined as gray-matter voxels showing the most stable responses across words in 10 repetitions (Mitchell et al., 2008). For each of the voxels with a probability higher than .4 in the SPM gray-matter mask, we computed a stability score to evaluate its response consistency regarding 90 words across 10 repetitions. For each subject, a gray-matter voxel was assigned a 90×10 matrix, where the entry at row *i*, column , was the β weight of this voxel during the th repetition (scanning run) of the *i*th word. The stability score for this voxel was then computed as the averaged pairwise correlations over all pairs of columns (scanning runs) in the matrix. This produced a stability gray-matter map for each subject. These stability maps were then submitted to a one-sample test at the group level, and the voxels with the top values (ranging from the top 100 to the top 5,000; for voxel distributions, see Fig. 3) were considered to show consistently high stability in response to word stimuli across subjects.

In Approach 3, word-associated voxels were identified in a meta-analysis of studies associated with word processing using Neurosynth (Yarkoni et al., 2011), an online platform for large-scale, automated meta-analyses based on the fMRI database of 14,371 studies (https://www.neurosynth.org/). Each study was automatically tagged with various terms (e.g., ;), and its activation coordinates were also automatically extracted. Using the term , Neurosynth divided the database into two sets: 944 studies were tagged with the term , and the other studies were not. The platform then produced an association test map showing scores from a two-way ANOVA to test for the association between each voxel and the term ; a higher score indicated that a voxel was more likely to be activated in studies tagged with the term than in those without. The association test map was assessed at a false-discovery-rate threshold of 0.01, and clusters with voxel sizes smaller than 10 were further removed. This method of functional region-of-interest localization has recently been widely used given the power offered by the large number of studies (e.g., Hung et al., 2020; Kragel & LaBar, 2016; Maimon-Mor & Makin, 2020).

In Approach 4, in case any regions sensitive to words' emotional meanings were not included, we redefined the word-associated mask as those clusters sensitive to any differences among object versus emotional nonobject versus nonemotional nonobject words. As in the object-versus-nonobject contrast, a GLM was built to include three regressors corresponding to the onset of each of the three word types for each run. The β maps for each word type versus baseline were submitted to a one-way ANOVA (within subjects) at the group level.

In Approach 5, instead of extracting activation patterns from a group-defined word-associated mask, we localized word-associated voxels in individual subjects using a group-constrained subject-specific approach (Fedorenko et al., 2010). Adopting a leave-one-subject-pair-out procedure, we first localized group-level word-associated parcels in 19 subjects on the basis of the object-versus-nonobject contrast. Within these parcels, we identified, for each of the remaining two subjects, the set of *N* voxels showing the largest differences between object and nonobject words. (Results of ISC-brain data were largely similar when the number of individual-defined voxels, *N*, increased from the top 50 to 400 voxels and to all the voxels in the group-defined mask; we reported ISC-brain results at $N = 300$ voxels.) We then united the two sets of voxels in the two subjects and calculated Pearson's correlations of activation patterns for this subject pair for each word. For a given

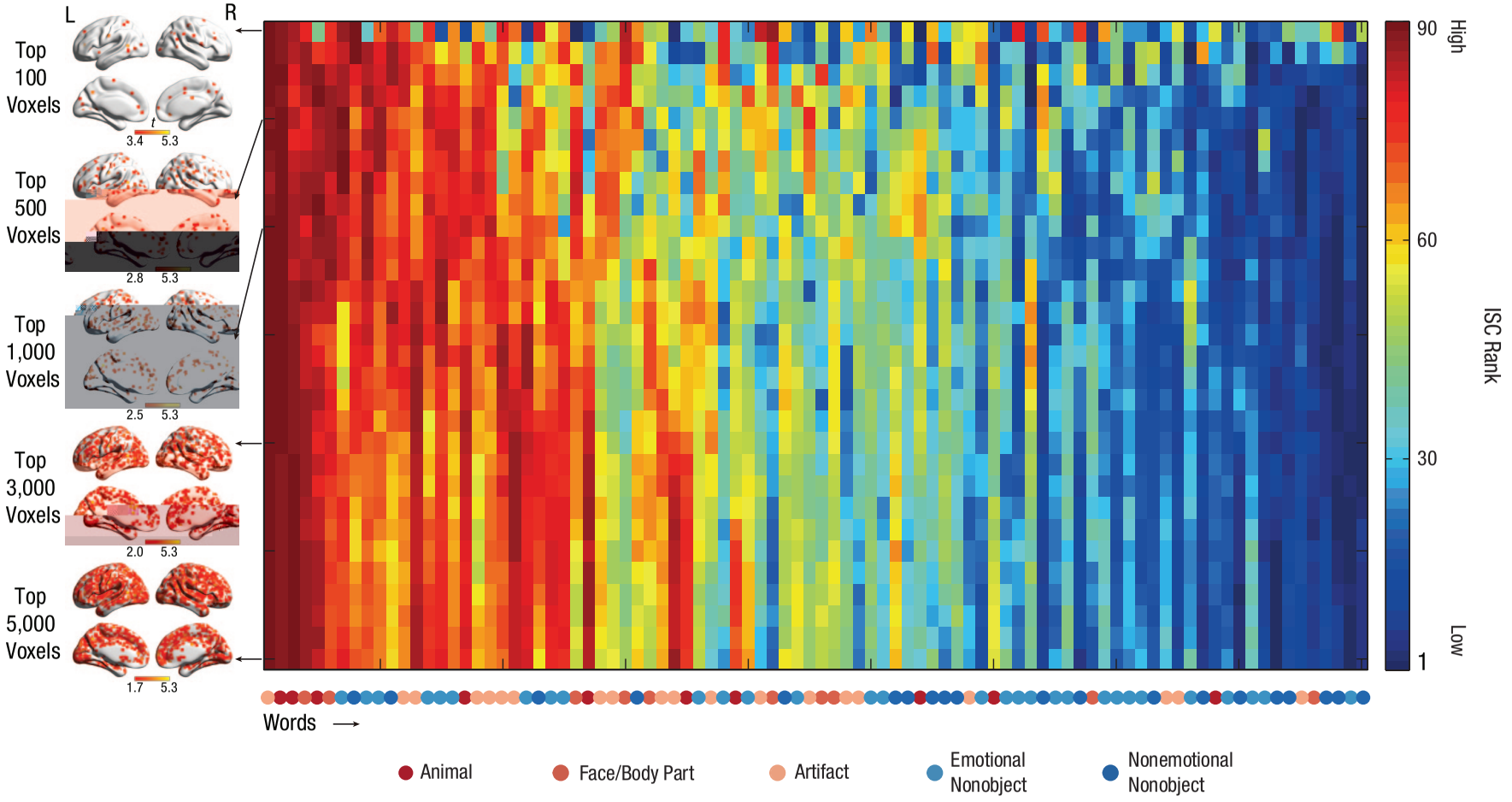


Fig. 3. Ranking of intersubject consistency (ISC) values from brain data, calculated from activation patterns in gray-matter voxels showing consistently high stability in response to words across subjects. The brain images (visualized using BrainNet with the "Maximum Voxel" algorithm; Xia et al., 2013) show the distribution of gray-matter voxels with the top-*N* highest stability scores (for details, see the Method section). The heat map shows the ranking of ISC-brain values across 90 words in the top-*N* voxels we sampled (from the top 100 to the top 1,000 voxels, in steps of 100 voxels each, and from the top 1,000 to the top 5,000 voxels, in steps of 200 voxels each). Words were sorted in descending order according to the averaged ISC-brain values from the top 100 to the top 5,000 voxels. L = left hemisphere; R = right hemisphere.

word, the correlations across all subject pairs were Fisher transformed and averaged to obtain the ISC-brain data.

Brain visualization. The brain maps and results were projected onto the MNI brain surface using BrainNet Viewer (Version 1.7; Xia et al., 2013; https://www.nitrc.org/projects/bnv/) with the default "interpolated" mapping algorithm, unless stated explicitly otherwise.

Ratings of candidate organizing principles of semantic representations in the brain

To explain the cognitive origins of word-meaning variation across individuals, we collected ratings on the following dimensions relevant to semantic representations. Each word was rated on a scale concerning emotional valence, ranging from 1 (i) to 4 ( ) to 7 (i i), and a scale for other ratings ranging from 1 ( ) to 7 (i). The rating instructions were as follows.

For sensory experience, subjects rated "to what extent the concept denoted by the word evokes a sensory experience (including vision, audition, taste, touch, and smell)." For navigation, they rated "to what extent the concept denoted by the word could offer spatial information to help you explore the environment." For manipulation, the instruction was to rate "to what extent the concept denoted by the word could be grasped easily and used with one hand." For stress-related actions, subjects rated "to what extent the concept denoted by the word would make you have a stress response, e.g., run away, attack, or freeze." For emotional valence, they rated "to what extent the concept denoted by the word evokes positive or negative feelings; very positive feelings mean that you are happy, satisfied, contented, hopeful; very negative feelings mean that you are unhappy, annoyed, unsatisfied, despaired, or bored." For arousal, they were asked to rate "to what extent the concept denoted by the word makes you feel aroused. Low arousal means that you feel completely relaxed, very calm, sluggish, dull, or sleepy; high arousal means that you are stimulated, excited, frenzied, jittery, or wide-awake." For language descriptiveness, the instruction was to rate "to what extent the concept denoted by the word could be described and explained using language."

We recruited independent groups of 26 to 30 college students from Beijing Normal University for each rating (N = 196) via an online survey (https://www.wjx.cn/). We computed a quality metric by correlating each subject's ratings with the averaged ratings from all subjects (except the subject being assessed) across all rated words. Subjects whose ratings were not significantly correlated with others' mean ratings (> .05) were excluded from the subsequent analyses, leaving 24 to 28 college students for each rating (N = 184).

Results

Cognitive representations of word meaning: individual consistency

(manipulation, navigation, and stress-related actions; Bi et al., 2016; Lambon Ralph et al., 2017; Martin, 2016); emotion-related (Kousta et al., 2011), including emotional valence and arousal; and language-related (X. Wang et al., 2020; i.e., language descriptiveness). We asked independent groups of subjects (from the same linguistic and cultural background as subjects in the main experiments) to rate the 90 words on each dimension on a 7-point scale (for details, see the Method section). We computed the mean and variation (indexed by standard deviation; Fig. 4a; see also Fig. S3 in the Supplemental Material) for each word across subjects' ratings as candidate sources for the ISC for behavior.

Each word's ISC-behavior value was predicted using multiple linear regression models with these variables as predictors. The means of language descriptiveness and sensory experience were highly correlated across the 90 words (= .94) and were collapsed by taking the average values into a single mean language/sensory-experience variable (see Fig. S3). The significant mean predictors (mean language/sensory experience, mean arousal, and mean valence) and standard-deviation predictors (standard-deviation language, standard-deviation manipulation, and standard-deviation valence) were obtained separately first and then considered together (see Table S1 in the Supplemental Material). The mean language/sensory-experience, mean arousal, and mean valence predictors were significant in the final model, together explaining 76.2% of the variance in the ISC for behavior: ISC behavior = 0.74 × Mean Language/Sensory Experience − 0.33 × Mean Arousal − 0.17 × Mean Valence + 0.59—regression-model significance test: $F(3, 86) = 91.64$, $= 1.07 \times 10^{-26}$; coefficient (β) significance tests: mean language/sensory experience, $(86) = 13.55$, $= 4.72 \times 10^{-23}$; mean arousal, $(86) = -5.76$, $= 1.27 \times 10^{-7}$; mean valence, $(86) = -3.14$, $= .002$ (for the partial regression plot between mean language/sensory experience and the ISC for behavior, see Fig. 4b). These effects persisted when we included word frequency and familiarity as nuisance variables (see Table S1). As an alternative approach to dealing with the correlated variables, we employed principal component analysis (see Table S2 in the Supplemental Material), and the results converged on the findings that the principal component with high loadings of mean language-descriptiveness and sensory-experience ratings was a significant predictor for the ISC for behavior and revealed that the principal component composed of standard deviations of emotion-related variables was another significant predictor (see the Results and Table S3 in the Supplemental Material). Note that we took extra caution to consider potential Chinese-specific orthographic properties that may contribute to the ISC effect. The majority of Chinese words are compound words made up of two or more characters, and some of the characters contain a semantic radical (indicative of meaning of the whole character, e.g., an animal). We obtained the average frequency measures of all characters or the first character and the frequency measures of semantic radicals in all characters or the first character from the Chinese lexical

topt-h1( )1f(ia6(tl)1(r))5f


Language
Navigation
Stress Action
Valence
Sensory
Manipulation
Arousal

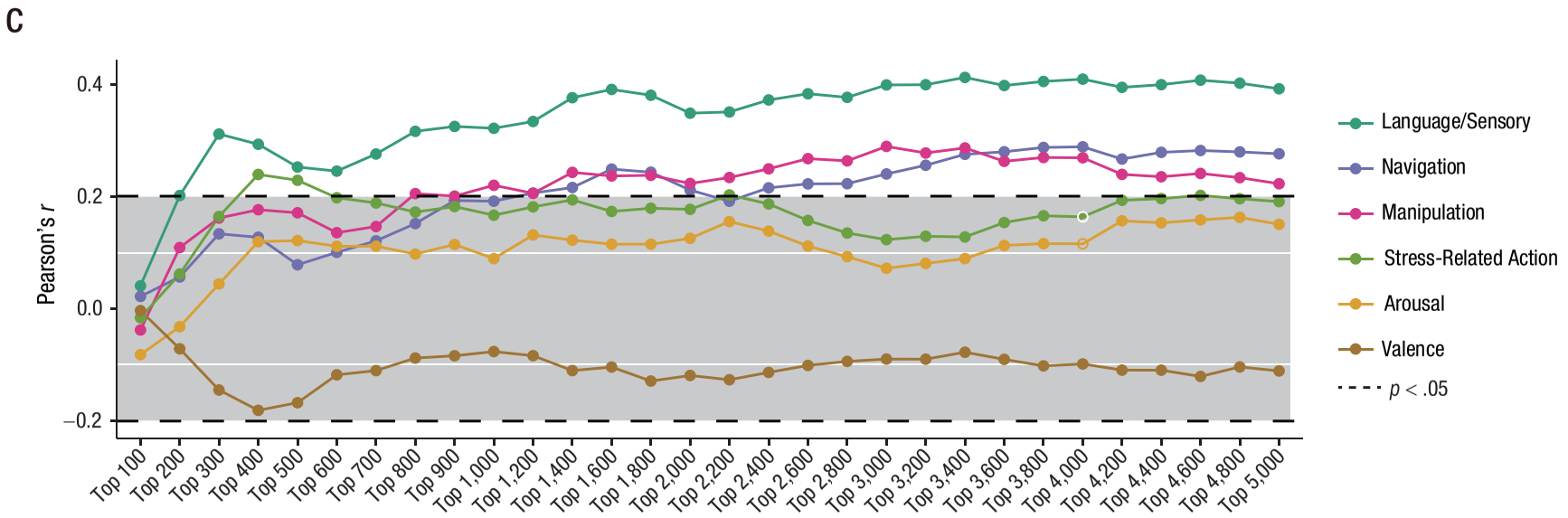


Fig. 4. Cognitive mechanistic origins of intersubject consistency (ISC) from behavioral data (ISC-behavior data) and brain data (ISC-brain data). Distributions of 90 words (a) are shown for each of seven semantic dimensions gleaned from cognitive and neuroscience studies. For each semantic dimension, ratings were collected from independent groups of 24 to 28 subjects (from the same linguistic and cultural background as subjects in the main experiments) on a scale ranging from 1 to 7. Both rating means and standard deviations across subjects were calculated as candidate sources for individual variations. Language- and sensory-rating means were transformed into scores and averaged to obtain the mean language/sensory-experience variable because of high correlation between them. Black outlines indicate the density of the data in each distribution. Partial regression scatterplots between ISC values and mean language/sensory experience are shown in (b). Results are shown separately for ISC-behavior values (covarying mean valence and arousal) and ISC-brain values in object versus nonobject regions. Diagonal lines show best-fitting regressions, and the error bands indicate 99% confidence intervals. Pearson's correlations between rating means of semantic dimensions and ISC-brain values in gray-matter voxels with high stability to words across subjects (sampled from the top 100 to 5,000 voxels) are shown in (c). The gray area marks the range in which correlations are nonsignificant (> .05).

and relatively nonobject-preferring regions (left posterior middle temporal gyrus, bilateral anterior temporal lobes, left inferior frontal gyrus, and dorsal medial prefrontal cortex; Fig. 2b), which were highly consistent with findings reported in the semantic literature (J. Wang et al., 2010; X. Wang et al., 2019). For each word, we obtained its activation pattern in this mask for each subject, calculated Pearson's correlations of the activation patterns across all subject pairs and then Fisher-transformed and averaged the values to form the ISC from brain data for each word.

As shown in the bar plots in Figure 2c, words referring to concrete referents (objects) such as and *i* were again highly significantly more consistent across individuals than words without external referents (mean Fisher- -transformed $M_{object} = .039$, $D = .008$ vs. $M_{nonobject} = .029$, $D = .007$), (88) = 6.23, = 1.59×10^{-8}, Cohen's = 1.33. The ISC-brain and ISC-behavior values were significantly correlated across words (= .43, = 2.20×10^{-5}). We examined which properties of word meanings account for the magnitude of ISC-brain values across words. The mean language/sensory-experience property was the only significant predictor in the final multiple regression model (Fig. 4b; see also Table S4 in the Supplemental Material), explaining 37.4% of the variance in the ISC from brain data: ISC brain = 0.61 × Mean Language/Sensory Experience + 0.033, $F(1, 88) = 52.57$, = 1.53×10^{-10}. The effect of mean language/sensory experience persisted when we included psycholinguistic confounds (see Tables S4 and S5) and when we used semantic principal components as predictors (see the Results and Table S3 in the Supplemental Material). That is, the more likely it was that a word could be described using language and/or was associated with sensory experiences (typically those with an external referent), the more similar brain activation patterns it induced across individuals.

Validation analyses using four different word-related brain-mask definitions (for details, see the Method section) yielded largely similar results to the analyses above. For Validation 1, without focusing on voxels showing different activations to predefined word types, we considered whole-brain ISC, selecting gray-matter voxels showing consistently high stability in response to words across subjects (following Mitchell et al., 2008). Figure 3 shows that the ISC rankings for object words and nonobject words were largely consistent across the size of the brain mask (number of voxels being selected). The positive correlations between ISC-brain value and mean language/sensory experience were statistically confirmed by the analyses shown in Figure 4c. Note that the significant correlation results for mean navigation and manipulation ratings were driven by their intercorrelations with mean language/sensory experience, as revealed by partial correlation analyses: The effects of mean language/sensory experience still held when analyses controlled for navigation or manipulation ratings (s < .034, for the top 200 to 5,000 voxels), and the effects of navigation or manipulation disappeared when analyses controlled for mean language/sensory experience (s > .17). For Validation 2, we used the search term in Neurosynth to identify brain areas consistently shown to be involved in word processing across a large number of studies in the neuroimaging literature (see Fig. S4 in the Supplemental Material). For Validation 3, in case any regions sensitive to words' emotional meanings were not included in the main contrast above, we redefined the word-meaning-associated mask as those clusters sensitive to any differences among object versus emotional nonobject versus nonemotional nonobject words (see Fig. S5 in the Supplemental Material). For Validation 4, we calculated ISC-brain values using voxels showing the greatest sensitivity to object versus nonobject words in individual subjects (rather than the group) for each subject pair (within the group mask identified in the remaining 19 subjects; Fedorenko et al., 2010; see Fig. S6 in the Supplemental Material). ISC-brain values obtained in these ways were highly correlated with the main results and all were significantly predicted by mean language/sensory experiences (Figs. 3 and 4; see also Figs. S4–S6). Finally, to examine the possibility that ISC-brain values may be driven by activation strength so that words with higher activations may show higher ISC, we extracted the overall activation strength for each word in a given mask and found that overall activation strength indeed significantly positively correlated with the ISC from brain data across 90 words in various brain-mask definitions (except for ISC-brain values computed with fewer than 800 stable voxels in gray matter in Validation 2; range = .23–.68). After we controlled for overall activation strength using partial correlation, the ISC from brain data still significantly correlated with mean language/sensory-experience ratings (range = .24–.67), indicating that the observed effect of activation-pattern consistency across individuals was not fully attributed to overall activation-strength differences.

Discussion

We found that speakers of the same language from a relatively homogeneous cultural and educational background exhibit substantial differences in their understanding of what a word means, measured both by behavioral judgment about relations with other words and by the patterns of brain activation when reading the words. Both behavioral and brain measures showed that the magnitude of ISC for a given word can be significantly positively predicted by how much the word is associated with sensory experience and language

descriptiveness. Behavioral and neural response patterns for words that refer to concrete entities (e.g., , - ,), which are associated with richer sensory experiences and are more easily described by language, are more similar across different people, compared with words without external referents (e.g., *i* *i* , *i*). These results were robust when other psycholinguistic variables, including familiarity and word frequency (and visual complexity in the fMRI experiment), were included as covariates and when multiple methods were used to construct behavioral measures or define brain masks.

There are debates about how to measure the internal representation of word (conceptual) meaning. Explicit-definition approaches and feature-listing approaches are highly controversial (Marggolis & Laurence, 1999; Tyler & Moss, 2001). The behavioral measure of word meaning based on relational structure with other words, although requiring no explicit definition, may be argued to be affected by potential task biases, such as the 2D spatial constraints of the testing environment and the sampling of other words. It is thus worth highlighting that our fMRI experiment is more invulnerable to these potential task biases because the subjects were asked to simply think about the word meaning, with the brain activity pattern for that word taken as the internal word representation. It may still be argued that the activity pattern of some regions may not necessarily be related to meaning, although we controlled for the effects of surface visual properties and validated the results across multiple brain mask definitions. The convergence of findings that we obtained using these multiple approaches and control analyses is thus particularly reassuring.

Where do intersubject differences about word representations come from? Decades of research on the general cognitive neural basis of word-meaning (semantic) representation (i.e., common across individuals) have led to a consensus of a decompositional structure entailing dimensions including salient sensory, motor, and emotion-related attributes (Binder et al., 2016; Kousta et al., 2011; Martin, 2016) and nonsensory language-derived representations (Landauer & Dumais, 1997; Striem-Amit et al., 2018; X. Wang et al., 2020). One source of individual variation may thus come from differences in experiences along these dimensions—different people may have different types or amounts of sensory, emotional, or language experiences with or *i* . Indeed, we found that sensory and language properties of words (group-mean judgments) were significant positive predictors of how similar or different they were across individuals. These measures of language descriptiveness and richness of sensory experience were highly correlated and were higher for words referring to objects (concrete words) than for words without external referents (abstract words). Although their effects on intersubject variability could not be disentangled at present, each may contribute to different aspects. For sensory representations, the more sensory experiences associated with a word, the more likely different people are to have at least some similar experiences, that is, the word is likely to be more robust to differences. Taking the word as an example, although people may have different quantities or qualities of tactile experiences with cats, they still have more common visual experiences with the form of a cat. If there is little sensory experience associated with a word to begin with, the same amount of experiential variation may lead to greater (sensory-derived) representation differences. For language, the rating was designed to capture how much of the word meaning could be derived from language inputs, that is, "to what extent the concept denoted by the word could be described and explained using language." The result that words referring to concrete referents tend to have higher ratings on this dimension is consistent with the classic context-availability theory (Schwanenflugel & Shoben, 1983), which proposes that the quantity and availability of verbal contextual information is lower for abstract concepts than for concrete concepts (see also Hoffman et al., 2013). The results here that increasing language descriptiveness is associated with greater intersubject agreements corroborate the findings that language-derived, nonsensory representations are one way of representing knowledge space (Striem-Amit et al., 2018; X. Wang et al., 2020). Intriguingly, we did not observe positive effects of emotion-related properties (arousal or valence) or action-response properties (manipulation, navigation, or stress) in predicting words' individual variability; however, previous literature showed that these dimensions contribute to word representation (Kousta et al., 2011) and that people differ in terms of emotional perception and concepts (Brooks & Freeman, 2018). These null results here are difficult to interpret and may be related to word sampling in the current experiment.

The current observations are likely not exhaustive in revealing the origins of the intersubject variations in word understanding. The results by themselves do not speak to whether the meaning representation differences arise from people's individual experiences ("nurture") or from genetic differences in terms of how neural circuits of various meaning components are hardwired (e.g., Briscoe et al., 2012). Also, it is unclear how the intersubject variation patterns of brain functionality (Mueller et al., 2013) and of word-meaning representations observed here are related. Finally, although modern semantic theories do not directly inherit earlier philosophical discussions, it is nonetheless worth noting that the current results are more in line with Locke's (1690) speculation that words denoting "complex ideas" (e.g., abstract words) may have lower ISC and not with Russell's (1948), who asserted that words entailing more

"abstractness of logic" may have greater individual consistency. Russell's arguments that nonsensory concepts have greater agreements may be relevant to specific sets of terms in which the definition is more logically transparent (e.g., math terms). The predictive power of a word's specific intrinsic property (language specificity/sensory experience) regarding agreement across people highlights the need to further test factors that specifically modulate these properties, including culture and ideology (Jackson et al., 2019; Thompson et al., 2020). Particularly worth highlighting are the potential effects of contemporary artificial intelligence algorithms that are widely applied, that is, automated individually tailored language (and sensory) inputs, which may symmetrically increase differences in language experiences and in turn lead to more drastic differences across people in word understanding.

To conclude, we have identified the extent and characteristics of intersubject variations in word understanding, showing that the agreements and disagreements of word representations systematically differ across different types of words. The magnitude of variability can be modeled with the association strength of words with sensory experiences and language descriptiveness, greater variability being associated with words without rich sensory experience or specific language descriptiveness (abstract words). Such disagreements on single-word meaning may at least partly underlie potential human communication failures, especially in settings that rely largely on terms without external referents such as politics, sociology, or legal domains. Increasing language descriptiveness and sensory experiences may help reduce miscommunication originating from these basic elements and facilitate more productive information exchanges and discussions.

Appendix

Table A1. Chinese Words (Along With English Translations) Used in The Present Study

Words with external referents ($N = 40$)			Words without external referents ($N = 50$)	
Animals (= 10)	Face/body parts (= 10)	Artifacts (= 20)	Emotional nonobject words (= 30)	Nonemotional nonobject words (= 20)
(ant)	(ankle)	(air conditioner)	(anger)	(agreement)
(cat)	(arm)	(ax)	(antipathy)	(business)
(elephant)	(ear)	(bed)	(apathy)	(characteristic)
(giraffe)	(eye)	(broom)	(charity)	(concept)
(panda)	(finger)	(cabinet)	(comfortable)	(content)
(rabbit)	(knee)	(chair)	(death)	(data)
(rat)	(lips)	(chopsticks)	(debt)	(discipline)
(sparrow)	(nose)	(computer mouse)	(depressed)	(effect)
(tiger)	(shoulder)	(hammer)	(disease)	(identity)
(tortoise)	(thigh)	(key)	(dispute)	(method)
		(microwave)	(error)	(obligation)
		(pencil)	(excited)	(phenomenon)
		(refrigerator)	(fate)	(process)
		(scissors)	(fault)	(reason)
		(sofa)	(fear)	(relationship)
		(spoon)	(fraud)	(result)
		(table)	(friendship)	(society)
		(television)	(happy)	(status)
		(toothbrush)	(heaven)	(system)
		(washing machine)	(hostility)	(team)
			(loving heart)	
			(magic power)	
			(marriage)	
			(miracle)	
			(proud)	
			(sad)	
			(scenery)	
			(splendor)	
			(trauma)	
			(violence)	

Transparency

A i E i Sachiko Kinoshita
E i Patricia J. Bauer
A C i
Y. Bi conceived the study. Both authors designed the study. X. Wang conducted the research and analyzed the data. Both authors wrote the manuscript and approved the final version for submission.
D i C i I
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
F i
This work was supported by the National Natural Science Foundation of China (Grant Nos. 31925020 and 31671128 to Y. Bi, Grant No. 31700943 to X. Wang), the Changjiang Scholar Professorship Award (Grant No. T2016031 to Y. Bi), the 111 Project (Grant No. BP0719032 to Y. Bi), the Fundamental Research Funds for the Central Universities (Grant No. 2017EYT35 to Y. Bi), and the China Postdoctoral Science Foundation (Grant No. 2017M610791 to X. Wang).
O P i
All data and materials have been made publicly available via OSF and can be accessed at https://osf.io/cyusp. The design and analysis plans for the studies were not preregistered. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Xiaosha Wang https://orcid.org/0000-0002-2133-8161

Acknowledgments

We thank Bijun Wang for assistance with data collection and Shuang Tian for assistance with figure preparation. We thank Chi Zhang for insightful discussions.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211003877

References

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. A. Zwaan (Eds.), *G i i : i* , (pp. 129–163). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.007

Bi, Y., Wang, X., & Caramazza, A. (2016). Object domain and modality in the ventral visual pathway. *i C i i* , *20*(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *C i i N* , *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Briscoe, J., Chilvers, R., Baldeweg, T., & Skuse, D. (2012). A specific cognitive deficit within semantic cognition across a multi-generational family. *P i B: Bi i* , *279*(1743), 3652–3661. https://doi.org/10.1098/rspb.2012.0894

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *N H B i* , *2*(8), 581–591. https://doi.org/10.1038/s41562-018-0376-6

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *N N i* , *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *i* - , , *i* , , and (and many other such concrete nouns). *J E P : G* , *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

de Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *J i* , *31*(3). http://www.jstatsoft.org/v31/i03/

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *J N i* , *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *C C* , *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *B i M* , *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hung, J., Wang, X., Wang, X., & Bi, Y. (2020). Functional subdivisions in the anterior temporal lobes: A large scale meta-analytic investigation. *N i Bi i i* , *115*, 134–145. https://doi.org/10.1016/j.neubiorev.2020.05.008

Jackson, J. C., Watts, J., Henry, T. R., List, J., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *i* , *366*(6472), 1517–1522.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *J E P : G* , *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *i C i i* , *20*(6), 444–455. https://doi.org/10.1016/j.tics.2016.03.011

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. F i P , 3, Article 245. https://doi.org/10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. F i N i , 2, Article 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. N , 60(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. N i N i , 1 (1), 42–55. https://doi.org/10.1038/nrn.2016.150

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. P i i , 104(2), 211–240.

Li, P., Schloss, B., & Follmer, D. J. (2017). Speaking two "languages" in America: A semantic space analysis of how presidential candidates and their supporters represent abstract political concepts differently. B i M , 4 (5), 1668–1685. https://doi.org/10.3758/s13428-017-0931-5

Locke, J. (1690). A i . Oxford University Press.

Maimon-Mor, R. O., & Makin, T. R. (2020). Is an artificial limb embodied as a hand? Brain decoding in prosthetic limb users. PLO Bi , 1 (6), Article e3000729. https://doi.org/10.1371/journal.pbio.3000729

Marggolis, E., & Laurence, S. (Eds.). (1999). C : C i . MIT Press.

Martin, A. (2016). GRAPES—grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. P i B i , 23(4), 979–990. https://doi.org/10.3758/s13423-015-0842-3

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. , 320(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., & Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. N , (3), 586–595. https://doi.org/10.1016/j.neuron.2012.12.028

R Core Team. (2020). : A i i i (Version 4.0.0) [Computer software]. http://www.R-project.org

Russell, B. (1948). H : I i . George Allen & Unwin.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. J E P : L i , M , C i , 9(1), 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Striem-Amit, E., Wang, X., Bi, Y., & Caramazza, A. (2018). Neural representation of visual concepts in people born blind. N C i , 9(1), Article 5250. https://doi.org/10.1038/s41467-018-07574-3

Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese Lexical Database (CLD): A large-scale lexical database for simplified Mandarin Chinese. B i M , 50(6), 2606–2629. https://doi.org/10.3758/s13428-018-1038-3

Sun, H. L., Huang, J. P., Sun, D. J., Li, D. J., & Xing, H. (1997). Introduction to language corpus system of modern Chinese study. In M. Y. Hu (Ed.), P i F C i i (pp. 459–466). Peking University Publishers.

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. N H B i , 4(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. i C i i , 5(6), 244–252. https://doi.org/10.1016/S1364-6613(00)01651-X

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. H B i M i , 31(10), 1459–1468. https://doi.org/10.1002/hbm.20950

Wang, X., Men, W., Gao, J., Caramazza, A., & Bi, Y. (2020). Two forms of knowledge representations in the human brain. N , 10 , 383–393. https://doi.org/10.1016/j.neuron.2020.04.010

Wang, X., Wang, B., & Bi, Y. (2019). Close yet independent: Dissociation of social from valence and abstract semantic dimensions in the left anterior temporal lobe. H B i M i , 40(16), 4759–4776. https://doi.org/10.1002/hbm.24735

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. PLO ONE, (7), Article e68910. https://doi.org/10.1371/journal.pone.0068910

Xiao, X., Zhou, Y., Liu, J., Ye, Z., Yao, L., Zhang, J., Chen, C., & Xue, G. (2020). Individual-specific and shared representations during episodic memory encoding and retrieval. N I , 21 , Article 116909. https://doi.org/10.1016/j.neuroimage.2020.116909

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. N M , (8), 665–670. https://doi.org/10.1038/nmeth.1635