

# Structural Similarities between Brain and Linguistic Data Provide Evidence of Semantic Relations in the Brain

Colleen E. Crangle<sup>1,2\*</sup>, Marcos Perreau-Guimaraes<sup>3</sup>, Patrick Suppes<sup>3</sup>

**1** School of Computing and Communications, Lancaster University, Lancaster, United Kingdom, **2** Converspeech LLC, Palo Alto, California, United States of America, **3** Suppes Brain Lab, Center for the Study of Language and Information, Stanford University, Stanford, California, United States of America

## Abstract

This paper presents a new method of analysis by which structural similarities between brain data and linguistic data can be assessed at the semantic level. It shows how to measure the strength of these structural similarities and so determine the relatively better fit of the brain data with one semantic model over another. The first model is derived from WordNet, a lexical database of English compiled by language experts. The second is given by the corpus-based statistical technique of latent semantic analysis (LSA), which detects relations between words that are latent or hidden in text. The brain data are drawn from experiments in which statements about the geography of Europe were presented auditorily to participants who were asked to determine their truth or falsity while electroencephalographic (EEG) recordings were made. The theoretical framework for the analysis of the brain and semantic data derives from axiomatizations of theories such as the theory of differences in utility preference. Using brain-data samples from individual trials time-locked to the presentation of each word, ordinal relations of similarity differences are computed for the brain data and for the linguistic data. In each case those relations that are invariant with respect to the brain and linguistic data, and are correlated with sufficient statistical strength, amount to structural similarities between the brain and linguistic data. Results show that many more statistically significant structural similarities can be found between the brain data and the WordNet-derived data than the LSA-derived data. The work reported here is placed within the context of other recent studies of semantics and the brain. The main contribution of this paper is the new method it presents for the study of semantics and the brain and the focus it permits on networks of relations detected in brain data and represented by a semantic model.

**Citation:** Crangle CE, Perreau-Guimaraes M, Suppes P (2013) Structural Similarities between Brain and Linguistic Data Provide Evidence of Semantic Relations in the Brain. PLoS ONE 8(6): e65366. doi:10.1371/journal.pone.0065366

**Editor:** Kevin Paterson, University of Leicester, United Kingdom

**Received:** July 18, 2012; **Accepted:** April 30, 2013; **Published:** June 14, 2013

**Copyright:** © 2013 Crangle et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** One of the authors (Crangle) is employed by a commercial company, Converspeech LLC. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: crangle@converspeech.com

## Introduction

This paper presents a new method of analysis that assesses the extent to which brain data are structurally similar to linguistic data. It allows us to evaluate the relatively better fit of the brain data with one semantic model over another. The first model is derived from WordNet, a lexical database of English compiled by experts in which words are grouped into sets of cognitive synonyms called synsets, each set expressing a distinct concept, with relations such as synonymy being defined between the synsets [1], [2]. The second model is given by a statistical technique for extracting from large document collections a measure of how similar two words are to each other in terms of how similar their contexts of use are within those documents [3], [4]. Such a statistical method is said to reveal semantic relations latent in the texts and is referred to by the acronym LSA (latent semantic analysis). The brain data are drawn from experiments in which statements about the geography of Europe were presented auditorily to participants who were asked to determine their truth or falsity while electroencephalographic (EEG) recordings were made [5].

The theoretical framework for the analysis of the EEG data and the semantic data derives from the axiomatization of theories such as the theory of differences in utility preference or the theory of

differences in psychological intensity. Using brain data samples from individual trials time-locked to the presentation of each word and extracted from the EEG recordings, ordinal relations of similarity differences are computed for the brain data and for linguistic data derived from the two models of semantic similarity. In each case those relations that are invariant with respect to the brain data and the linguistic data, and are correlated with sufficient statistical strength, amount to structural similarities between the brain and linguistic data. Our results show that many more statistically significant structural similarities can be found between the brain data and the WordNet-derived data than the LSA-derived data.

The work reported here builds on earlier work by Suppes and colleagues in which machine learning was employed to recognize brain representations of phonemes, isolated words, sentences, words within sentences, and simple images and their names. That is, the following question was answered affirmatively for EEG and, in one instance, magnetoencephalographic (MEG) data. Can we train a classifier of brain data so that we can predict which word or sentence or word within a sentence the participant is seeing or hearing [6], [7], [8], [9], [10], [11], [12]? Similarly, the following questions were answered. Can we train a classifier on some participants and use it to make predictions for other participants [5], [13]? And, can we train a classifier using words and use that

classifier to make predictions for pictures depicting what the words refer to (and vice versa) [14]? A fundamental result that follows from this work is that when brain wave recordings are time-locked to the presentation of words in a sentential context, transformations of segments of those brain waves can be identified as representations (or brain images) of those words. Furthermore, these representations have approximate invariance across test participants and across a variety of sentential contexts. These are clearly representations not of the individuals' idiosyncratic perceptions of the orthography or sound of the words but of something that could be called the meaning of the words or their semantics.

There have been other recent attempts to study semantics and the brain, also using machine-learning approaches. Many have used functional magnetic resonance imaging (fMRI) data collected while participants were presented with words or word-image pairs. The semantics of a word was represented by its distributional properties in the data set contributed by Google Inc., which consists of English word n-grams and their frequencies in an approximately 1-trillion-word set of web pages [15], [16]. Specifically, for nouns referring to physical objects, their semantics was given by their co-occurrence patterns with 25 manually-selected sensory-motor verbs in the corpus. Taking 60 such noun, fMRI data were collected while participants paid attention to consistent sets of properties of the objects. Using either 58 or 59 words in a training set, statistically significant predictions were made for the semantic category (tool or mammal, for example) of the held-back words relative to each other or to a larger set of words outside the 60. This approach established that semantic features of words could be learned from brain data, thereby allowing a word outside the training set to be correctly classified in terms of its semantic features. More recent work by Just and others sought to reduce the 25-feature representation of the 60 words to a small set of semantic features [17]. The reduction was through factor analysis of the fMRI data to identify locations in the brain that give the best classification of fMRI data for the 60 words.

Other corpora and other ways of selecting the co-occurrence features have been investigated to see if they might offer improved recognition of the Mitchell and others fMRI data [15]. Pereira and others, [18], used semantic features obtained from a large text corpus consisting of articles from Wikipedia pertinent to the specific nouns. Latent Dirichlet allocation (LDA, [19]) gave the co-occurrence features, with words being represented as probabilistic distributions over topics and topics as probabilistic distributions over words. LDA is one of several unsupervised methods that, like LSA, find relations hidden or latent in the text. Kelly and others used parsed data to identify three-part rather than two-part associations, such as *flute-produces-sound*, rather than the simple *flute-sound* [20]. They also included three-part associations involving widely accepted computational and linguistic relations such as *is-a* and *part-of*, as in *Paris is a capital* and *Paris is part of France*. Wikipedia and the British National Corpus provided the language data. In a companion paper, Devereux and others, [21], showed that their semantic representation produces comparable predictions for the fMRI activity of the Mitchell and others dataset [15].

In the work of Murphy and others we see semantics and the brain investigated primarily with EEG data [22], [23], [24]. Using the words in the Mitchell and others experiments as a starting point, EEG data were collected from the presentation of 60 images corresponding to the objects named by the 60 words [15]. These EEG samples, both single-trial and averaged, were classified into two categories (tools and mammals) using co-occurrence data collected for the words describing the images. Several corpora were tested, including a large corpus of newspaper texts that

included linguistic information such as part-of-speech tagging and lemmatization. Patterns of co-occurrence were gleaned using singular value decomposition applied to over 22,000 frequent nouns and 5,000 frequent verbs to find the best 25 co-occurrence candidates in place of the pre-selected 25 verbs [15]. Their work also advanced understanding of the time latencies and intervals, frequency bands and scalp locations that improve classification rates, and of the relative advantage of MEG over EEG data. Murphey and others systematically investigated several different distributional models of semantics that could be applied to arbitrary corpora [25]. Of particular interest was the linguistic preprocessing that gave the best results, with dependency parsing and relative position (which word came first in a co-occurrence) proving most effective.

The work of Baroni and others represents efforts to build corpus-based models of semantics that capture something of the nature of the similarity between two words, not just a quantification of that similarity [26]. Properties that could be typed – dogs have tails as *parts*, barking as *behavior*, for example – were of interest and surface grammatical patterns were relied on to pinpoint the potentially interesting co-occurrence patterns. The corpus was a large part-of-speech tagged set of web pages. Outside knowledge sources such as the information in a corpus of child-directed speech and the hypernymy structure of WordNet were used to further constrain co-occurrence candidates.

In Jelodor and others, [27] we find WordNet used as a supplementary source of semantic information for the analysis of the dataset [15]. Several WordNet similarity measures, individually and in combination, were used to compute the similarity of each of the 60 nouns with each of the 25 sensory-motor verbs of Mitchell and others [15]. This approach was handicapped by the fact that in WordNet some similarity measures do not apply to verbs, and some verb-noun comparisons were approximated using noun-noun comparisons, which made sense only when the noun and the verb had strongly consonant meanings. There was also no attempt to limit the WordNet senses to those that were relevant to the specific nouns and verbs, with the highest scoring synset-to-synset similarity score simply being selected. In the end, the highest classification scores obtained in this work were with a combination of the WordNet similarity scores of the 60 nouns with the 25 sensory-motor words and the co-occurrence features used originally [15].

A move away from corpora is seen in the work of Chan and others [28]. EEG and MEG data were collected during a large number of trials (>1,000) involving a visual and auditory language task in which participants were required to assess the relative size of each of over 500 objects, which were equally divided between living and non-living. The challenge was to learn a binary semantic feature (living or non-living) from the brain data and then correctly classify words in terms of this feature. The amplitude at six time points, using all channels, provided the brain data. The high temporal resolution of EEG and MEG permitted to identify spatial and temporal features of the brain data that improved classification rates and additionally allowed the classification of individual words [28].

Sudre and others also took a corpus-free approach to learning semantic features from brain data [29]. In their study of concrete nouns, scaled values (0 to 5) for semantic features associated with the objects and their names were obtained from a large number of informants. Simultaneous EEG and MEG data were separately collected from participants who were asked Yes/No questions about 20 of these features for 60 objects and their names. Features related to size, manipulability and animacy of the objects proved

most effective in predicting MEG data, and these were associated with different spatial and temporal features of the brain data.

In Huth and others we find the WordNet graph structure used to map a large set of categories (1,705), reduced by principal component analysis, onto fMRI data collected while subjects watched videos featuring objects and actions in those categories [30]. The surprising result in this work was how the brain activity related to category responses was so widespread, with activity for some categories localized and for others distributed.

The studies of language and the brain discussed above can each be characterized by whether or not they relied on corpora, specifically the distributional properties of words in corpora, to provide semantic features of the words. For the corpus-based studies, differences exist in the choice of corpus and the strategies used to glean distributional properties, including whether or not linguistic information such as part of speech or WordNet measures were exploited. Those that did not rely on a corpus solicited features either directly or indirectly from language users, relying on a large set of categories and binary or graded membership of categories to specify the semantic model.

Our work draws a comparison between the use of corpus-based distributional properties of words, specifically those provided by LSA, and the semantic information encoded in WordNet. Our interest is not in which gives better predictions of the brain data. It is in the ability of each semantic model – WordNet or the corpus-based distributional semantics of LSA – to represent the network of relations between a given group of words that a participant is considering – {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}, for example. LSA and other corpus-based distributional semantic models detect relations between words that are latent or hidden in text. WordNet makes these relations explicit based on the judgment of language experts. Ultimately, our interest is in evaluating the extent to which the network of relations we detect in the neural data corresponds to the network of relations represented by LSA compared to WordNet.

The approach taken in this paper is briefly as follows. We use machine learning to predict which word a given brain-data sample is associated with. That is, we classify the brain-data samples into classes that correspond to the words. We then use those classifications to compile confusion matrices, to capture not only which brain-data samples were classified correctly but also, for those that were not correctly classified, the words they were incorrectly associated with. This approach derives from the well-known experiments in Miller and Nicely in which confusion matrices were used to deduce important information about the perception of phonemes [31]. In our analysis, the pattern of mistaken predictions for the brain data is used to tell us something important about the perception of words in the brain. We can then compare what we learn about these perceptions with the two semantic models, WordNet and LSA.

The main contribution of this paper is in the new method it introduces by which structural similarities between brain data and linguistic data can be compared at the semantic level. Included in this method is a way to measure the strength of the structural similarities found between the brain data and the linguistic data, which allows us to assess the relatively better fit of the brain data with one semantic model over another.

## Materials and Methods

### 2.1 Brain Data

Brain data comes from the experiments described in Suppes and others [5], [14]. In the experiment of interest, 48 spoken sentences about the geography of Europe were presented to nine partic-

ipants in 10 randomized blocks, with all 48 sentences occurring once in each block. The 48 sentences were statements about commonly known geographic facts of Europe, with half of the sentences true, half false, half positive, and half negative (e.g., *The capital of Italy is Paris* and *Paris is not east of Berlin*). The possible forms of these trigger sentences are:  $X$  is [not]  $W$  of  $Y$ ,  $W$  of  $Y$  is [not]  $X$ ,  $X$  is [not]  $Z$  of  $X$ ,  $Y$  is [not]  $Z$  of  $Y$ , where  $X \in \{Berlin, London, Moscow, Paris, Rome, Warsaw, Madrid, Vienna, Athens\}$ ,  $Y \in \{France, Germany, Italy, Poland, Russia, Austria, Greece, Spain\}$ ,  $W \in \{the\ capital, the\ largest\ city\}$ ,  $Z \in \{north, south, east, west\}$ , and [not] indicates the optional presence of *not*. The maximum sentence length was 2.59 s and the minimum 1.35 s, with the interval from sentence onset to sentence onset 4.05 s. In a related experiment using the same sentences, the words were visually displayed one by one on a computer screen. The same timing was used in the auditory presentation of the sentences, providing a start and end time for each word. The sampling rate was 1,000 Hz and the data were down-sampled 16 times, giving an effective sample rate of 62.5 Hz.

With our focus on individual words in their sentential contexts, we extracted segments of the brain-wave data for each word of interest from every trial in which that word appeared. For classifying the wave-form segments, we needed to produce data samples of equal length. We let each word's data sample begin approximately 160 ms after the auditory onset of the word, that is, 10 data points beyond the word's initial start point. To determine the end point, we extended each word's data sample by 50 data points (that is, approximately 800 ms) and calculated the average length of the resulting data samples. This average length then determined the end point of each word's data sample. The data samples from all fifteen channels were concatenated, producing a single brain-data sample for each occurrence of a word. The choice of start and end points was informed by work showing the N400 component of the event related potential (ERP) to be associated with semantic processing and integration [32], [33], [34]. Furthermore, recent work by Brennan and Pykkänen confirmed the long-held assumption that since sentences appear to be understood incrementally, word-by-word, brain activity time-locked to the presentation of words in sentences can be analyzed to tell us something about language processing [35]. They additionally found that the activity 250 ms after word onset can be associated with the processing of words in sentences as opposed to words in lists.

The 21 country names, city names, and the four directions or relative locations are the words of interest for our analysis. We selected three city names, three country names and the four relative location words for classification as a group. The linguistic data for the ten words {*London, Moscow, Paris, north of, south of, east of, west of, Germany, Poland, Russia*} and the brain data for one subject (S18) are used to illustrate the methods throughout the paper. We chose S18 because that participant gave good classification rates. We use the data from this subject and the one set of ten words consistently throughout the paper to eliminate any possibility of cherry-picking participant-word samples in each instance to support our points. Results for all nine participants and all 21 words are reported in the paper, however, and form the basis for our conclusions. The words occurred within the trials with the frequencies here shown in parentheses: *Berlin* (70), *London* (80), *Moscow* (90), *Paris* (80), *Rome* (80), *Warsaw* (80), *Madrid* (70), *Vienna* (40), *Athens* (50), *France* (40), *Germany* (40), *Italy* (40), *Poland* (40), *Russia* (50), *Austria* (40), *Greece* (40), *Spain* (50), *north* (60), *south* (70), *east* (60), and *west* (70). Material S1 gives the trigger sentences and the timing of the presentation of each word for each sentence.

## 2.2 Models and Methods of Analysis

**2.2.2 Structural models.** The first step in constructing structural models for the brain data is to analyze the data using a statistical model that predicts to which class a brain-data sample belongs, where each class corresponds to one of the 21 word types. In order to classify the segments of data obtained from the individual trials, we use a linear discriminant model in a 5-fold cross-validation loop. The dimensionality of the data was reduced using a nested principal component analysis. Ocular artifacts were removed from the brain-data samples using blind source separation. The procedure is described in detail by Perreau-Guimaraes and others [11]. This classifier resulted from 10 years experimentation with different kinds of classifiers. It gave consistently higher recognition rates and it was used to derive the results that led to the current work [12].

For our 10 sample words, each of 640 brain-data samples is classified as a representation of one of the 10 words. More generally, the  $T$  brain-data samples  $s_1, s_2, \dots, s_T$  are classified into the  $N$  classes  $\omega_1, \omega_2, \dots, \omega_N$ . In this process, we do not directly compare the samples to each other to judge their similarity; rather we use a set of  $N$  prototypes, and these are what are designated by  $\omega_1, \omega_2, \dots, \omega_N$ . If test sample  $s_i$  is classified as  $\omega_i$  then  $s_i$  and  $\omega_i$  have a minimal similarity difference compared to the other possible classifications.

Let  $M = (m_{ij})$  be the confusion matrix for a given classification task, where  $m_{ij}$  is the number of test samples from class  $\omega_i$  classified as belonging to class  $\omega_j$ . By computing the relative frequencies  $m_{ij}/\sum_j m_{ij}$  we obtain  $N$ -by- $N$  estimates for the conditional probability densities that a randomly chosen test sample from class  $\omega_i$  will be classified as belonging to class  $\omega_j$ . Figure 1 gives the conditional probability estimates computed from the confusion matrix resulting from a single-trial classification of the brain data for  $\{London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia\}$  from participant S18. Following the conventions for heat maps, the higher the value of each element in the matrix the darker its shading.

Let these conditional probability estimates be given by the matrix  $P=(p_{ij})$ . For each class  $\omega_i$  we then define a quaternary relation  $R$  such that  $\omega_i \omega_j R \omega_k \omega_l$  if and only if  $p_{ij} < p_{kl}$ , that is, if and only if the probability that a randomly chosen sample from class  $\omega_j$  will be classified as belonging to class  $\omega_i$  is smaller than the probability that a randomly chosen test sample from class  $\omega_k$  will

be classified as belonging to class  $\omega_l$ .  $R$  is an ordinal relation of similarity differences, a partial order that is irreflexive, asymmetric, and transitive. It can of course be simplified to a ternary relation  $R^*(\omega_i \omega_j \omega_k)$ . For simplicity we continue to use  $R$  for this ternary relation.

Along the lines of the approach for the axiomatization of the theory of differences in utility preference or the theory of differences in psychological intensity, we put forward the relational structure  $(A, R)$ , which is constructed from  $R$  and the finite set  $A$  of classes  $\omega_i$  together with the real-valued functions given by the inequalities  $p_{ij} < p_{ik}$  [36]. (Specifically, the function defined by these inequalities is the function  $f$  defined on  $A$  such that  $xy R uv$  iff  $f(x) - f(y) < f(u) - f(v)$ , with  $f(x) - f(y)$  represented by  $p_{xy}$ .) Note that an ordinal relation of similarity differences, a partial order that is irreflexive, asymmetric, and transitive, is defined for each class  $\omega_i$ . We therefore have  $N$  partial orders and the notation  $(A, R)$  will from here on be understood to refer to a structure with  $N$  partial orders defined on the set  $A$ .

At this point we have a formal characterization of the brain data from the experiment. What of the linguistic data? The focus was on the recognition of phonemes and words [12]. Linguistic data were provided by the well-known Miller and Nicely experiment in which participants heard 16 English consonants under various conditions of noise or other distortion and then reported their perceptual judgments by classifying the consonants [31]. The confusion matrix of phonemes arising out of these perceptual judgments gives rise to a relational structure  $(A', R')$  constructed from  $R'$  (more precisely from  $N$  such relations) and the finite set  $A'$  of classes  $\omega'_i$  together with the real-valued function (more precisely, the  $N$  such functions) given by the inequalities  $p'_{ij} < p'_{ik}$ . The two relational structures  $(A, R)$  and  $(A', R')$  provide the means by which comparison can be made between the brain data and the linguistic data. For the Miller and Nicely experiments, the linguistic data were provided by human judgments of similarity. For the geography sentence experiments, the linguistic data are provided by similarity measures derived from two models of semantic similarity, as we describe in Section 2.3.

**2.2.3 Comparing structural models of brain and semantic data.** Isomorphism between  $(A, R)$  and  $(A', R')$  would constitute the strongest measure of similarity, but isomorphism is almost certainly too strong a requirement for brain data obtained under current experimental conditions. The confusion matrices of the

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	0.275	0.108	0.133	0.042	0	0.008	0	0.075	0.025	0.033
Moscow	0.133	0.408	0.167	0.008	0.025	0.025	0.008	0.025	0.025	0.033
Paris	0.05	0.158	0.208	0.05	0.017	0.05	0.033	0.05	0.008	0.033
north	0	0.033	0.067	0.617	0.017	0	0.017	0.05	0.017	0.033
south	0.05	0.033	0.017	0	0.367	0.05	0.083	0.067	0	0.067
east	0.033	0	0	0.1	0	0.517	0.283	0.017	0	0
west	0	0.033	0.05	0.017	0.017	0.25	0.433	0.017	0.017	0.017
Germany	0.167	0.15	0.1	0.05	0.05	0.033	0	0.217	0	0.1
Poland	0.017	0.017	0	0.033	0.017	0.017	0	0.017	0.6	0.017
Russia	0.017	0.083	0.083	0.033	0.017	0	0.017	0	0	0.6

**Figure 1. Conditional probability density estimates (shown as a heat map) computed from the confusion matrix resulting from the classification of 640 brain wave samples from S18 for the set of words  $\{London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia\}$ .** This table was taken from [12]. doi:10.1371/journal.pone.0065366.g001

brain and linguistic data could be compared directly using a statistical test such as the chi square, for example. And cluster or similarity trees can be computed from these matrices, allowing visual assessment of the congruence between the brain and linguistic data. However, direct comparison of the confusion matrices, whether visually or using a single measure of quantitative fit, is both too strong under current experimental conditions and too weak. It is too strong because data derived from electrophysiological measurements on the scalp almost certainly present an incomplete picture of brain activity. Direct visual or single-measure comparison is too weak because it tells little of the precise nature and strength of any congruence there might be between the brain and linguistic data. Instead we have devised a procedure that compares the partial orders of the two relational structures described in the previous section. This approach, we submit, has a stronger theoretical basis and offers a more detailed view of the similarities that do hold between brain data and perceptual or linguistic data.

We derive from the  $N \times N$  matrix of conditional probability densities for the brain data  $N$  strict partial orders  $R^B_i$  ( $i = 1, \dots, N$ ) of the similarity differences of the ordered pairs, one ordering for each of the  $N$  classes into which the samples are classified. Similarly for the perceptual data related to phonemes or, the focus of this paper, the semantic data related to the words. We derive from the perceptual or semantics matrix  $N$  strict partial orderings  $R^P_i$  ( $i = 1, \dots, N$ ) of the similarity differences of the ordered pairs, one ordering for each of the  $N$  classes into which the samples are classified. For simplicity, we omit the index  $i$  in the description that follows.  $R^B$  and  $R^P$  are understood to refer to the orderings for any one of the  $N$  classes.

The question now is how to compare the two partial orders. For each class  $\omega_i$ , we first compare  $R^B$  and  $R^P$  to see if they are order isomorphic. Order isomorphism is defined in terms of the partially ordered sets (or posets) corresponding to the partial ordering relations. Two partially ordered sets  $X$  and  $Y$  are order isomorphic iff there is a one-to-one and onto mapping  $f$  from  $X$  to  $Y$  such that for all  $x_1, x_2 \in X$ , the ordered pair  $(x_1, x_2) \in X$  iff the ordered pair  $(f(x_1), f(x_2)) \in Y$ . That is, the mapping  $f$  between the two posets is bijective and both  $f$  and  $f^{-1}$  are order preserving. If  $f$  also preserves the mapping between the classes in the relational structure for the brain data and those in the relational structure for the linguistic data, order isomorphism establishes a strong notion of similarity [37].

Order isomorphism at the level of the individual classes is more likely under current experimental conditions than isomorphism between the two full relational structures. Nonetheless, what is required to compare the brain and linguistic data is a generalization of the notion of order isomorphism to the notion of *invariant partial order*. An invariant partial order for the brain and linguistic data is the intersection of the partial order for the brain data,  $R^B$ , and the partial order for the perceptual data,  $R^P$ . Each intersection  $R^{INV} = R^B \cap R^P$ , one for each of the classes into which the samples are classified, is also a partial order that is irreflexive, asymmetric, and transitive. We devise a way to judge the strength of the structural similarity represented by these invariant partial orders.

We compute the Spearman rank correlation coefficient of the two partial orders  $R^B$  and  $R^P$  that give rise to each invariant partial order. Each invariant partial order for which the contributing  $R^B$  and  $R^P$  partial orders have a statistically significant correlation between them is counted as a significant invariant partial order. It is these significant invariant partial orders that constitute the structural similarity between the brain and linguistic data, and the total number of such significant partial orders is a measure of the strength of that structural similarity.

We have restricted ourselves to strict partial orders so far but it can be the case that confusion matrices produced by the brain or perceptual data contain ties in a row, that is, for any class  $\omega_i$ . If these equal numerical scores are permitted in the partial order, the result is a strict weak ordering instead of a strict ordering and the property of irreflexivity does not hold. Instead of permitting ties in the data to give rise to weak strict orders, a possibly more realistic response to equal numerical scores in the experimental context lies with semiorders. A semiorder requires that two numerical scores be within some threshold of each other before they are compared. Introduced in mathematical psychology, semiorders are intransitive and used as a model of human preference that permits nontransitive indifference [38], [39]. If the threshold is allowed to vary from one score to another, the order is an interval order. Semiorders or interval orders can prove useful in comparing relational structures for brain and perceptual data, [12]; in this paper, however, Suppes and others restrict ourselves to strict partial orders.

In summary, the procedure by which structural similarity between brain and linguistic data is judged is as follows. The similarity differences from the linguistic data and the brain data respectively are presented in the form of two confusion matrices. For each matrix, conditional probabilities are constructed from which we define  $N$  strict partial orders for the brain data and  $N$  for the linguistic data. For each  $N$ , the intersection of the brain and language partial orders give rise to one or more invariant partial orders with respect to the brain and linguistic data. Those invariant partial orders for which the contributing brain and language partial orders are significantly correlated, as measured by the Spearman rank correlation coefficient, constitute the structural similarity between the brain and perceptual data. The total number of significant invariant partial orders gives a measure of the strength of that structural similarity.

## 2.3 Semantic Data

**2.3.1 WordNet-based model of similarity.** WordNet is a lexical database of English compiled by experts using lexicographic practices to identify word senses, define relations between those senses, and compose glosses (or definitions) of those senses. In WordNet, nouns, verbs, adjectives and adverbs are each grouped into sets of cognitive synonyms (called synsets), each expressing a distinct concept [1], [2], [40]. The synsets are related to each other, primarily through the hypernymy and hyponymy relations for nouns. Other relations in WordNet are part-whole (holonym), member of (meronym), has instance, and so on. Hyponymy, often referred to as the *is-a* relation in computational discussions, is defined as follows: a concept represented by a lexical item  $L_i$  is said to be a hyponym of the concept represented by a lexical item  $L_k$  if native speakers of English accept sentences of the form "An  $L_i$  is a kind of  $L_k$ ." Conversely  $L_k$  is the hypernym of  $L_i$ . A hypernym is therefore a more general concept and a hyponym a more specific concept. The WordNet database has the structure of a directed acyclic graph, and the node for one concept  $c_x$  is said to be an ancestor of the node for another concept  $c_y$  if  $c_x$  can be reached from  $c_y$  by traversing one or more relational links, known as paths or edges. In WordNet one root node subsumes all of the noun concepts, and similarly for verbs, adjectives and adverbs.

The well-known example of *canaries can fly*, and others like it, suggests that humans represent words and their semantic relations to each other something like WordNet does. We can quickly verify that *canaries can sing* because a canary is a songbird, with only one level of hyponymy between. But to verify that *canaries can fly*, two levels of hyponymy must be traversed and to verify that *canaries have skin* several levels. These two judgments take correspondingly

more time. Words as represented in WordNet have multiple senses, each sense being part of one or more synsets. The WordNet senses for the city names, the country names, and the four direction or location terms (*east, west, north and south*) are listed in Material S2. For each word, the senses relevant to the geography sentences are listed first, with the non-relevant senses shown indented beneath. There are two senses of *London* in WordNet, for example. The first, designated London#1, concerns the capital of Britain, the second, designated London#2, the writer. Only the first sense of the word *London* is relevant to the geography sentences of our experiment.

Several different WordNet-based measures have been proposed to represent the degree to which two words are related. A detailed discussion of such measures [41] and online access is given to a program that computes these [42], [43]. (Pedersen, Patwardhan, Michelizzi, 2004). Three main approaches have dominated. The first uses path length between two synsets measured along WordNet’s taxonomic (*is-a*) paths, the second uses the notion of information content, a corpus-based frequency-related measure, and the third is a vector-space-based approach that uses the words in the definitions (or “glosses”) of the synsets. We combined five measures in our computations of similarity to take account of these different approaches, averaging the results from measures based on path-length, information-content, and vector-space, along with hybrids of these. Each similarity measure captures something important about the relations between words as represented in WordNet. Many studies have compared these and other measures of relatedness in WordNet, assessing their effectiveness for tasks as varied as word-sense disambiguation, information retrieval, indexing, and spelling correction, as done in Budanitsky and Hirst [41]. However, none emerges as clearly superior as a general measure of similarity-oriented relatedness. Recent work by Ballatore and others, furthermore, specifically compared the performance of several combinations of WordNet similarity scores against individual scores and found a clear pattern of success (that is, fidelity to human judgments) for combined scores [44]. Coincidentally, their work looked specifically at geography-related word. Further details of the five individual measures we used are given in Material S3. In Figure 2 we show the results for all 10 words averaged over all relevant senses. Figure 3 gives a hierarchical cluster tree computed for the data in Figure 2 using furthest (also known as complete) distance. It shows the expected

groupings of the city names, country names, and location or direction names.

**2.3.2 Statistical method for extracting latent semantic relations.** LSA is a statistical technique for extracting from large collections of documents a measure of how similar two words are to each other in terms of patterns of their co-occurrences within those documents [4], [3], [45]. The underlying idea is that if you take into account all the contexts in which a word does and does not appear within a set of documents, you capture something important about that word’s meaning in those documents. LSA has a plausible interpretation as a psychological theory of similarity and knowledge. The similarity judgments produced by latent semantic analysis have been shown to correspond to some extent to human judgments of similarity. For example, after training on about 2,000 pages of English text, it scored as well as average test-takers on the synonym portion of the Test of English as a Foreign Language (TOEFL). LSA’s judgments of similarity are derived from the specific document set used in the computations, a restriction that can be of practical advantage. For example, after training on a psychology textbook it achieved a passing score on a multiple-choice exam. LSA as generally implemented does not use syntactic information, which most likely limits what it can capture about semantic similarity, but there is no reason such information cannot be incorporated into an LSA-type model. We used the application at <http://lsa.colorado.edu/> (accessed January 15, 2013) to compute similarity matrices in term space for our set of words. See Figure 4. The computation was based on texts of general reading up to 1st year college level and a maximum of 300 factors was permitted in the analysis. Figure 5 gives the similarity tree computed from Figure 4. It shows the expected groupings of the city names, country names, and direction terms except for a displacement of *Moscow* into the country names.

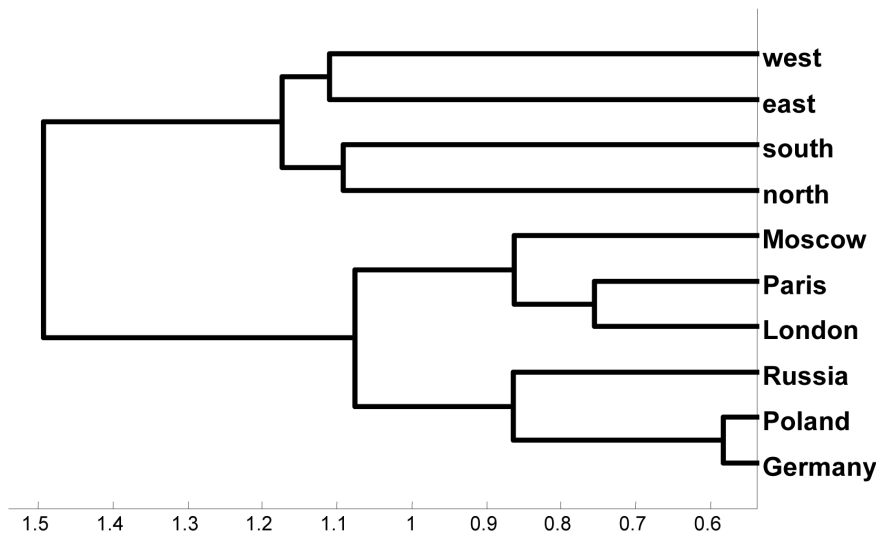
**2.4 Structural Models of Similarity between Brain Data and Semantic Data**

We now have two structural representations, one of the brain data and the other of the semantic data. We compare them.

**2.4.1 Invariant partial orders.** We start with the conditional probability density estimates in Figure 1 derived from the brain data and the similarity scores derived from WordNet (Figure 2) and LSA (Figure 4). Line graphs of the conditional probability densities and the semantic similarities for each word (that is, each row of the two tables) give a sense of how the

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	1	0.396	0.466	0.106	0.103	0.076	0.078	0.322	0.299	0.303
Moscow	0.396	1	0.393	0.095	0.094	0.062	0.07	0.286	0.281	0.288
Paris	0.466	0.393	1	0.106	0.104	0.074	0.077	0.327	0.308	0.307
north	0.106	0.095	0.106	1	0.228	0.179	0.21	0.123	0.132	0.111
south	0.103	0.094	0.104	0.228	1	0.172	0.212	0.115	0.107	0.109
east	0.076	0.062	0.074	0.179	0.172	1	0.216	0.093	0.08	0.077
west	0.078	0.07	0.077	0.21	0.212	0.216	1	0.087	0.082	0.083
Germany	0.322	0.286	0.327	0.123	0.115	0.093	0.087	1	0.589	0.409
Poland	0.299	0.281	0.308	0.132	0.107	0.08	0.082	0.589	1	0.403
Russia	0.303	0.288	0.307	0.111	0.109	0.077	0.083	0.409	0.403	1

**Figure 2. Semantic similarity matrix (shown as a heat map) derived from WordNet for the set of words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} using senses relevant to the geography of Europe.**  
doi:10.1371/journal.pone.0065366.g002



**Figure 3. Hierarchical cluster tree computed from the WordNet-based semantic similarity matrix for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} given in Figure 2.**  
doi:10.1371/journal.pone.0065366.g003

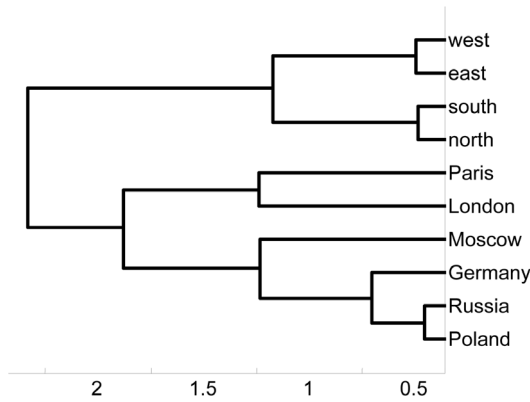
linguistic data and the brain data appear to track each other. See Figure 6 for the example of *London*. (Data for the other words can be found in Figure S1 a-j.) The Spearman rank correlation for the brain data and the WordNet-derived data in Figure 6 is 0.99, with one-sided significance of 1.84E-10. The Spearman rank correlation for the brain data and the LSA-derived data in Figure 6 is 0.8758 with one-sided significance of 8.9494e-004. The Spearman rank correlation, as an indication of the strength of the relationship between two partial orders, is a statistic of interest. But what is of primary interest based on the approach described in section 2.4 are the ordinal relations of similarity differences, partial orders that are irreflexive, asymmetric, and transitive, that are *invariant* with respect to the semantic and brain data. For *London*, the WordNet partial order is given by the 10-element sequence {*London Paris Moscow Germany Russia Poland north south west east*}. The brain data yields two nine-element sequences {*London Paris Moscow Germany north Russia Poland east south*} and {*London Paris Moscow Germany north Russia Poland east west*} arising from the tie between the scores for *south* and *west*. See Table 1. The partial orders that are invariant with respect to the brain data and WordNet-derived data are the

three seven-element sequences {*London Paris Moscow Germany Russia Poland south*}, {*London Paris Moscow Germany Russia Poland west*} and {*London Paris Moscow Germany Russia Poland east*}. These three invariant partial orders jointly constitute the similarity between the brain and semantic data for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}. A graphical representation is given in Figure 7.

Looking now at the partial orders derived from the LSA linguistic data and, again using *London* as an example, we find the partial orders that are invariant with respect to the brain data and these linguistic data. The LSA data yield eight seven-element sequences. The brain data yield the same two nine-element sequences as before. There are 16 partial orders that are invariant with respect to the brain and LSA-derived data. They jointly constitute the similarity between the brain and LSA semantic data for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}. A representation of the graphical join of these invariant partial orders is given in Figure 8.

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	1	0.17	0.37	0.13	0.12	0.12	0.14	0.17	0.16	0.16
Moscow	0.17	1	0.18	0.13	0.08	0.22	0.14	0.37	0.65	0.69
Paris	0.37	0.18	1	0.1	0.04	0.08	0.09	0.5	0.31	0.36
north	0.13	0.13	0.1	1	0.89	0.6	0.61	0.07	0.09	0.14
south	0.12	0.08	0.04	0.89	1	0.5	0.55	0.02	0.05	0.05
east	0.12	0.22	0.08	0.6	0.5	1	0.85	0.22	0.29	0.3
west	0.14	0.14	0.09	0.61	0.55	0.85	1	0.24	0.26	0.23
Germany	0.17	0.37	0.5	0.07	0.02	0.22	0.24	1	0.85	0.81
Poland	0.16	0.65	0.31	0.09	0.05	0.29	0.26	0.85	1	0.87
Russia	0.16	0.69	0.36	0.14	0.05	0.3	0.23	0.81	0.87	1

**Figure 4. Similarity scores computed using LSA for the words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} shown as a heat map.**  
doi:10.1371/journal.pone.0065366.g004



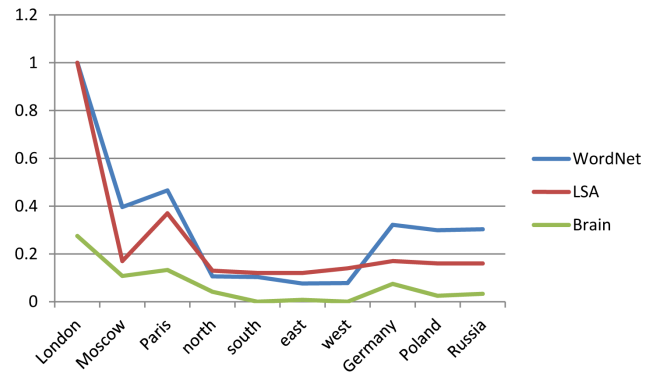
**Figure 5. Hierarchical cluster tree computed from the LSA scores of similarity for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} given in Figure 4.**  
doi:10.1371/journal.pone.0065366.g005

### 2.5 Computing Invariant Partial Orders

In this section we turn to the task of computing the invariant partial orders for the brain and linguistic data. To illustrate we use the semantic similarity matrix of Figure 2— that is, the Wordnet-derived semantic data for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} – and the conditional probability density estimates presented in Figure 9 below. The estimates in Figure 9 were computed using, as before, the linear discriminant model of Perreau-Guimaraes and others to classify the 640 single-trial data samples for the 10 geography words for participant S18 [11]. All computations described here were performed using the MATLAB programming environment. We use this new classification of the same ten words for the same participant to illustrate an important computational point that is also an important point about the brain data. We return to this point in section 2.6.

As stated earlier, for each word  $\omega$  we define a ternary relation  $R'$  such that  $R'(\omega, \omega_1, \omega_2)$  if and only if with respect to word  $\omega$  the conditional probability for word  $\omega_1$  is smaller than the conditional probability for word  $\omega_2$ , that is if and only if  $\omega_1$ 's similarity difference with  $\omega$  is smaller than  $\omega_2$ 's similarity difference with  $\omega$ , or in terms of semantic similarities  $\omega_1$  is more similar to  $\omega$  than is  $\omega_2$ .  $R'$  is an ordinal relation of similarity differences, a partial order that is irreflexive, asymmetric, and transitive.

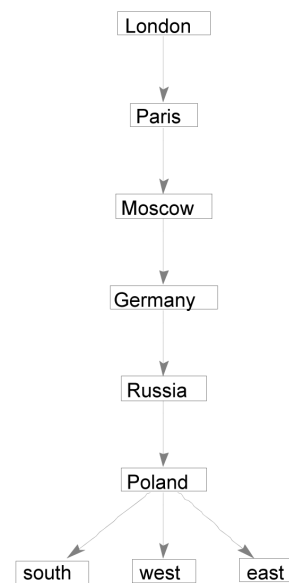
$R'$  can be represented computationally using a connection matrix  $C_\omega$  where  $C_\omega(\omega_1, \omega_2) = 1$  if and only if with respect to word  $\omega$  the conditional probability for word  $\omega_1$  is smaller than the conditional probability for word  $\omega_2$ , that is if  $\omega_1$ 's similarity difference with  $\omega$  is smaller than  $\omega_2$ 's similarity difference with  $\omega$ . Similarly for the semantic similarity matrix of Figure 2. These connection matrices can be computed automatically in a straightforward manner from Figure 2 and Figure 9. The connection matrix  $C_{London}^B$  for London for the brain data is given in Table 2. The connection matrix  $C_{London}^S$  for London for the semantic data is given in Table 3. Highlighted entries show where the partial orders differ. A directed acyclic graph can be derived from the connection matrix for a word  $\omega$ , with an arrow going from node  $\omega_1$  to node  $\omega_2$  if and only if  $\omega_1$ 's similarity difference with  $\omega$  is smaller than  $\omega_2$ 's similarity difference with  $\omega$ . Figure 10 shows the connection graph  $CG_{London}^B$  for the brain data for London, and Figure 11 the connection graph  $CG_{London}^S$  for the semantic data. These connection graphs may appear too complex to be potential models of semantic relations as they are represented in the brain. However, although these structures are visually complex in graph form, the connection matrices to which they are



**Figure 6. WordNet-based and LSA-based semantic similarities and EEG conditional probability estimates for London relative to London, Moscow, Paris, north, south, east, west, Germany, Poland, and Russia.** Data taken from Figure 1, Figure 2, and Figure 4.  
doi:10.1371/journal.pone.0065366.g006

logically equivalent are structurally simple and amenable to simple computations. More will be said in the Discussion about the role such networks may play in accounts of how the brain processes language.

To return to our task of computing the structural similarity between brain and linguistic data, connection graphs allow isomorphisms to be computed automatically using standard functions in the MATLAB programming environment. In MATLAB, biograph objects  $CG_{London}^B$  and  $CG_{London}^S$  are constructed from connection matrices  $C_{London}^B$  and  $C_{London}^S$  using the statements  $CG_{London}^B = biograph(C_{London}^B)$  and  $CG_{London}^S = biograph(C_{London}^S)$ . A graph isomorphism, a 1-to-1 mapping of the nodes in the graph  $CG_{London}^B$  and the nodes in the graph such that adjacencies are preserved, can then be computed, if it exists, using the *isomorphism* function in MATLAB. That is, the statement  $[Isomorphic, Map] = isomorphism(CG_{London}^B, CG_{London}^S)$  is used,



**Figure 7. Joint invariant partial order for London relative to the set of words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.** Brain data derived from FIGURE 1 and semantic data from the WordNet data in Figure 2.  
doi:10.1371/journal.pone.0065366.g007



**Table 1.** Partial orders for *London* derived from the WordNet semantic similarities of Figure 2 and the conditional probability estimates for the brain data of Figure 1.

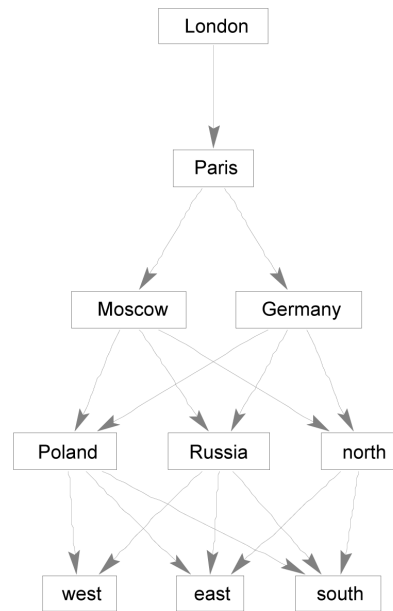
London			
Language data		Brain data	
London	1.000	London	0.275
Paris	0.466	Paris	0.133
Moscow	0.396	Moscow	0.108
Germany	0.322	Germany	0.075
Russia	0.303	north	0.042
Poland	0.299	Russia	0.033
north	0.106	Poland	0.025
south	0.103	east	0.008
west	0.078	south	0.00
east	0.076	west	0.000

doi:10.1371/journal.pone.0065366.t001

which returns logical 1 (true) if the two N-by-N adjacency matrices extracted from biograph objects  $CG_{London}^B$  and  $CG_{London}^S$  are isomorphic graphs and logical 0 (false) otherwise. When the Boolean value *Isomorphic* is true, *Map* is a row vector containing the node indices that map from  $CG_{London}^B$  to  $CG_{London}^S$ . When *Isomorphic* is false, the worst-case time complexity is  $O(N!)$ , where N is the number of nodes. When *Isomorphic* is true, the row vector *Map* must also be checked to see if the labeled nodes in the one structure are mapped to their correspondingly labeled nodes in the other structure. Without that, it is merely a structural isomorphism without mapping that has been detected, which is of limited value in comparing brain and perceptual data.

Isomorphism between the brain and perceptual data in fact lies simply in the two connection matrices being identical and this identity can be determined by a straightforward element-by-element comparison of the two connection matrices; for example, for *London*, an element-by-element comparison is made of  $CG_{London}^B$  for the brain data and  $CG_{London}^S$  for the semantic data. Failing isomorphism, which in current experimental situations is unlikely, it is useful to know what other structural similarities may hold for the brain and semantic data. These are best described in graph-theoretic terms. For two graphs G1 and G2 we would like to know what is the largest induced subgraph of G1 isomorphic to a subgraph of G2. However, this maximum common subgraph isomorphism problem is an NP-hard optimization problem. The associated decision problem, namely, given G1, G2 and an integer k, deciding if G1 contains a subgraph of at least k edges isomorphic to a subgraph of G2, is NP-complete, with no known polynomial-time solution guaranteed to work in general [46]. However, the importance of these problems to bioinformatics has given rise to other ways of approaching the problem – for example, by finding cliques [47] – and such algorithms may have application to the analysis of brain data.

What can be computed when isomorphism fails is a topological sort of each connection graph, that is, an ordering of the graph's nodes such that if there is an edge from node x to node y, then y appears after x in the ordering. These topological sorts are exactly the partial orders of the relational structures  $(A, R^*)$  presented in Section 2.2. The intersection of these topological sorts reveals the structural similarity between the language and the brain data. In general, the node order produced by a topological sort is not



**Figure 8.** Joint invariant partial order for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}. Brain data derived from Figure 1 and the LSA data from Figure 4.

doi:10.1371/journal.pone.0065366.g008

unique. Specifically, in the case of the brain or semantic data, every tie in the similarity differences for a given word gives rise to two distinct orderings. There is no efficient way in general to find all topological sorts; it is an NP-complete problem. However, for directed acyclic graphs it can be solved in polynomial time and a simple brute-force generation of all topological sorts is easily done for the limited number of words at play at any time in sentence understanding. Once all topological sorts are available for the brain and semantic data, the common orderings of maximal length can be extracted, also in polynomial time. These computations are easily implemented in MATLAB using its *toporder* function. Returning to the brain and semantic data for *London*, we compute  $Ord_{London}^B = toporder(CG_{London}^B)$  and  $Ord_{London}^S = toporder(CG_{London}^S)$ . We then find the longest common subsequence of  $Ord_{London}^B$  and  $Ord_{London}^S$ . Suppose it has length L. All subsequences of this maximal length L that are common to both  $Ord_{London}^B$  and  $Ord_{London}^S$  jointly constitute the similarity between the brain and the semantic data.

### 2.6 Evaluating Structural Similarity

Comparing Figure 7 and Figure 8 and the Spearman rank correlation coefficients associated with these figures, it is hard to form a clear judgment of which model of semantic similarity has a better structural fit with the brain data. However, only one single-trial classification of the brain data produced the results given in these figures and the invariant partial orders were for *London* only. If the results for all the words are examined over a sufficiently large number of single-trial classifications, a different picture emerges.

We therefore performed the following computations. For each of the nine participants and for each of four sets of 10 words (three city names, three country names, four relative location words), we computed 60 single-trial classifications of the brain data using random resampling with replacement. For half of these classifications we found the partial orders that were significantly highly correlated ( $\rho = .6485, p < 0.05$ ) and invariant with respect to the WordNet-based semantic data. We did the same for the LSA data.

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	0.25	0.163	0.138	0.075	0.038	0.05	0.063	0.088	0.075	0.063
Moscow	0.144	0.333	0.133	0.033	0.011	0.056	0.056	0.089	0.067	0.078
Paris	0.175	0.188	0.125	0.038	0.038	0.038	0.05	0.1	0.138	0.113
north	0.067	0.017	0.017	0.283	0.167	0.15	0.1	0.1	0.067	0.033
south	0.014	0.014	0.071	0.114	0.271	0.171	0.157	0.029	0.086	0.071
east	0.05	0	0	0.133	0.15	0.383	0.15	0.05	0	0.083
west	0.057	0.043	0	0.171	0.114	0.143	0.229	0.086	0.1	0.057
Germany	0.075	0.175	0.1	0.05	0.125	0.025	0.05	0.275	0.025	0.1
Poland	0.15	0.1	0.125	0.05	0.05	0	0.025	0.15	0.25	0.1
Russia	0.16	0.04	0.08	0.04	0.1	0.02	0.12	0.1	0.08	0.26

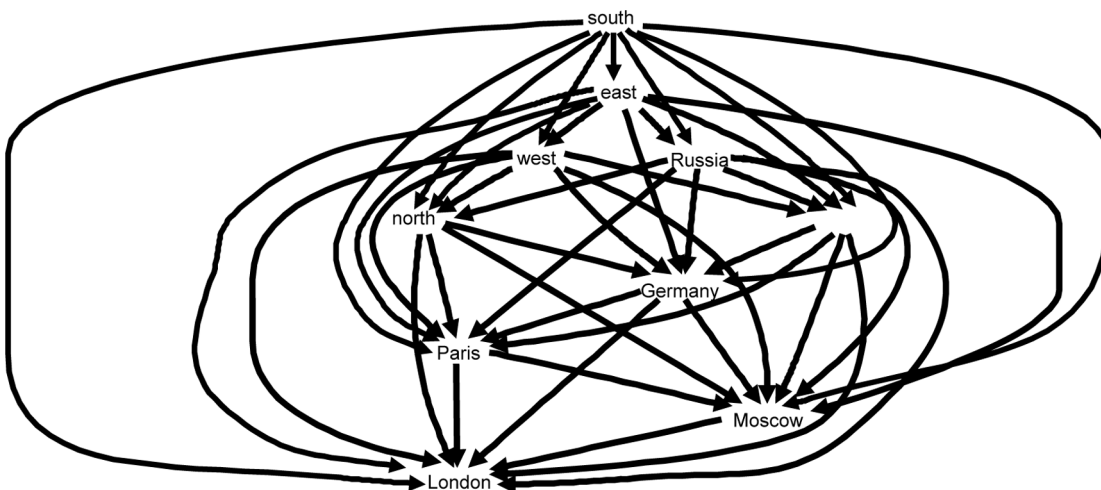
**Figure 9. Conditional probability density estimates (shown as a heat map) from a new single-trial classification of the brain data (S18) for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.**  
doi:10.1371/journal.pone.0065366.g009

By using the single set of observations (the brain data) repeatedly we were engaging in a form of bootstrapping. That is, we were estimating the property of interest – the conditional probability densities – using an approximating distribution. The assumption behind this approach is that the observed data for each participant are from a population of identically distributed and independent sets of such observations.

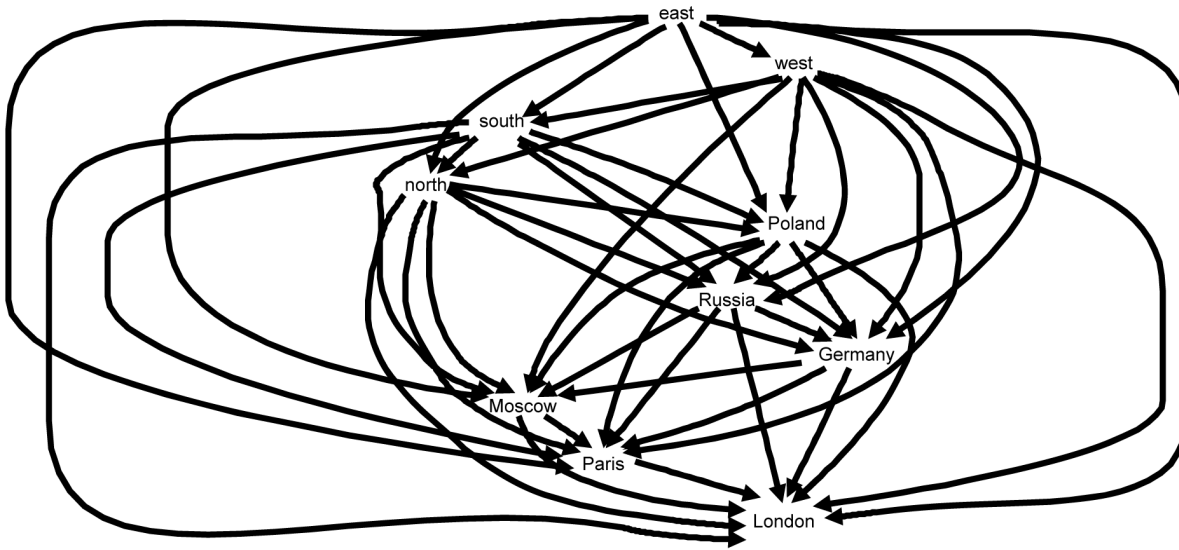
We used four sets of words to ensure that every word appeared at least once in our computations. The sets were {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}; {*Paris, Vienna, Athens, north, south, east, west, Italy, Spain, Austria*}; {*Berlin, Rome, Warsaw, north, south, east, west, France, Greece, Poland*}; and {*Madrid, Rome, Vienna, north, south, east, west, Spain, Italy, Austria*}.

The average classification percent rate overall was 20.81. The average classification rates for individual participants were 16.76, 17.27, 16.97, 18.68, 16.26, 25.41, 24.69, 19.83, and 31.5. The average classification rates obtained for each of the four sets of words over the nine participants were 21.13, 21.14, 20.48, and 20.54. All rates are well above chance (10%). Furthermore, these

results are stable, as illustrated in Figure 12, which presents the results of 120 single-trial classifications of the sample ten geography words for S18. These repeated computations produced consistent classification rates in the range 25% to 29%, with a mean classification rate of around 24.5%,  $p < 10E-10$ . We also examined specific conditional probabilities obtained over a set of 120 single-trial classifications for four sets of 30 single-trial classifications at a time. Figure 13 shows for each set of 30 single-trial classifications (using S18) the probability that a test sample is classified as belonging to the set *Moscow* when the sample actually belongs to the set *London* (dashed line), *Paris* (double line), or *Germany* (solid line). These conditional probability estimates show similar constancy over time. As another test of the stability of the successive single-trial classifications, we computed the cumulative averages of the conditional probability estimates over two sets of 60 single-trial classifications and examined the series of similarity trees generated from these conditional probability estimates. For the first set of 60 classifications, from the 7<sup>th</sup> to the 23<sup>rd</sup> classification the similarity tree switched between the two



**Figure 10. Ordinal relation of similarity differences represented as a directed acyclic graph  $CG^B_{London}$  for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}, derived from the probability density estimates for the brain data in Figure 1.**  
doi:10.1371/journal.pone.0065366.g010



**Figure 11. Ordinal relation of similarity differences represented as a directed acyclic graph  $CG^S_{London}$  for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}, derived from the WordNet data in Figure 2.**  
doi:10.1371/journal.pone.0065366.g011

configurations (left and right) in Figure 14. In both trees, *Paris* and *London* are closely coupled and *east, south, west, and north* are grouped identically. In the tree on the left, *Russia, Moscow, Germany* and *Poland* are intermingled. In the tree on the right, *Germany's* grouping with the cities is the only anomaly. After the 23<sup>rd</sup> classification the cumulative tree remained the same as the tree on the right. This similarity tree in fact first emerged on the 2<sup>nd</sup> single-trial classification. For the second set of 60 single-trial classifications, the stable configuration emerged after the 2<sup>nd</sup> classification, shown in Figure 15, and never changed.

**Results**

For each of the nine participants (identified as S10, S12, S13, S16, S18, S24, S25, 26 and S27) and each of four sets of words, we computed two sets of 30 single-trial classifications of the brain data

as described above. One set of results from the brain data for each participant was compared with the linguistic data derived from WordNet and the other LSA. We then assessed the degree of structural similarity between the brain data and the linguistic data. The maximum number of significant invariant partial orders that could be obtained in each set of single-trial classifications was 300 (30 classifications of 10 words). Figure 16, Figure 17, Figure 18 and Figure 19 give the results for WordNet and LSA for each participant for the four sets of words. A higher number of significant structural similarities were found between the brain data and the WordNet data than the LSA data. This pattern held for all but two of the 72 experimental runs. It is interesting to note that the difference between the number of structural similarities for WordNet and LSA was smallest for the set of words that had for each city the country of which it was the capital, namely the set {*Madrid, Rome, Vienna, north, south, east, west, Spain, Italy, Austria*}.

**Table 2.** Ordinal relation of similarity differences represented as a connection matrix  $C^B_{London}$  for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} shown by initial letters, derived from the probability density estimates for the brain data in Figure 9.

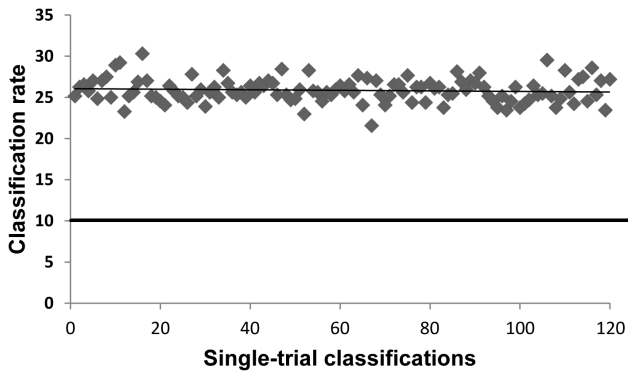
	L	M	P	n	s	e	w	Ge	Po	Ru
L	0	0	0	0	0	0	0	0	0	0
M	1	0	0	0	0	0	0	0	0	0
P	1	1	0	0	0	0	0	0	0	0
n	1	1	1	0	0	0	0	1	0	0
s	1	1	1	1	0	1	1	1	1	1
e	1	1	1	1	0	0	1	1	1	1
w	1	1	1	1	0	0	0	1	1	0
Ge	1	1	1	0	0	0	0	0	0	0
Po	1	1	1	0	0	0	0	1	0	0
Ru	1	1	1	1	0	0	0	1	1	0

doi:10.1371/journal.pone.0065366.t002

**Table 3.** Ordinal relation of similarity differences represented as a connection matrix  $C^S_{London}$  for *London* relative to the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} shown by initial letters, derived from the WordNet data in Figure 2.

	L	M	P	n	s	e	w	Ge	Po	Ru
L	0	0	0	0	0	0	0	0	0	0
M	1	0	1	0	0	0	0	0	0	0
P	1	0	0	0	0	0	0	0	0	0
n	1	1	1	0	0	0	0	1	1	1
s	1	1	1	1	0	0	0	1	1	1
e	1	1	1	1	1	0	1	1	1	1
w	1	1	1	1	1	0	0	1	1	1
Ge	1	1	1	0	0	0	0	0	0	0
Po	1	1	1	0	0	0	0	1	0	1
Ru	1	1	1	0	0	0	0	1	0	0

doi:10.1371/journal.pone.0065366.t003



**Figure 12. Single-trial classification rates over 120 trials for the set of words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.** 640 samples (S18) classified into 10 classes. Chance accuracy is shown by the thick solid line. doi:10.1371/journal.pone.0065366.g012

**Discussion**

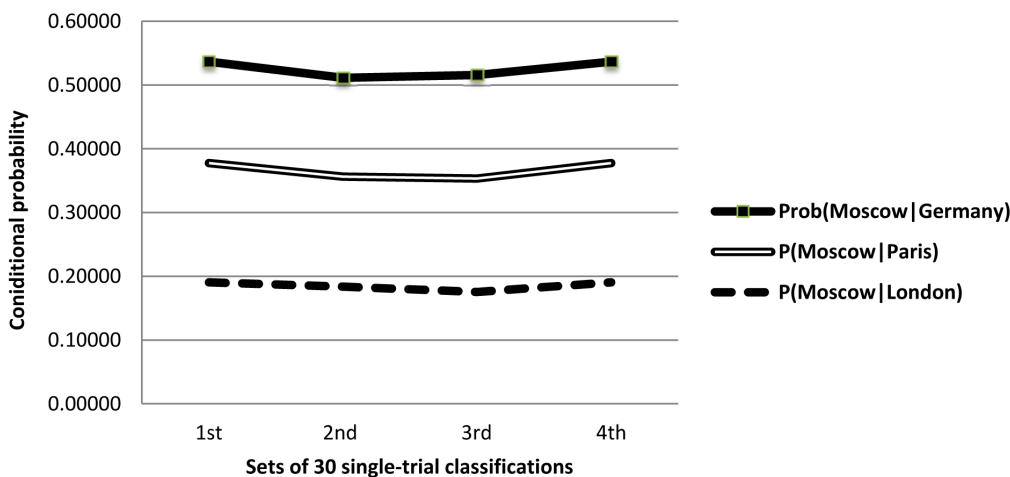
We start with some possible objections to the results presented in this paper. First is the consideration that some other property of the stimuli co-varies with the words of interest and is responsible for the recognition rates and confusion matrices. Two possible confounding variables are the length of the words and the number of syllables. The relative location words *north*, *south*, *east* and *west* are all one syllable and shorter than most of the city and country names, only a few of which are one syllable. With the modified sample lengths as described in section 2.1, however, the data samples for these words correspond roughly to the two-word and two-syllable utterances *north of*, *south of*, *east of*, and *west of*. The three-syllable names – *Vienna*, *Italy*, *Germany* and *Austria* – are misclassified relative to the other words at approximately the same rate as the two-syllable words and so their possible confounding effect is taken to be minor for this study. A third possible confounding variable is that of grammatical class, which is typically highly correlated with word meaning. In our geography sentences, the grammatical roles of the country and city names are different from those of the relative location terms (*east*, *west*, *north*, and *south*). It is not immediately clear to what extent, if any, our analysis has been

affected by grammatical class. It is significant, however, that the city names and country names are clearly differentiated despite being from the same grammatical class. Furthermore, any confounding effect would be present in the structural similarity evaluation for both semantic models and so would not invalidate the comparative conclusion we have reached.

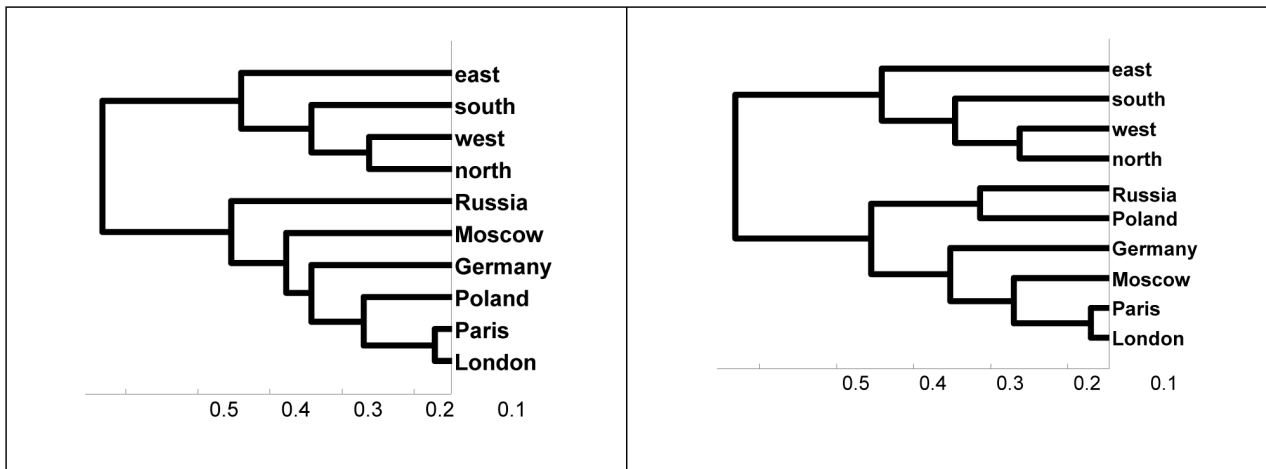
Another possible objection to the results presented in this paper is that the recognition rates and confusion matrices reflect the probabilities of each word’s occurrence in the sentences. That is, *north* and *south* and *east* and *west* are confused with each other because of their expected occurrences in their given positions in the sentences. Similarly for the country names and city names. In this case, a probabilistic semantic grammar could perhaps explain the confusion matrices, without any recourse to positing WordNet or LSA as an underlying semantic model. The notion of a word’s expected appearance at a particular point in a sentence is related to the notion of cloze probability, which is defined as the probability of a specific word completing a particular part of a sentence. Kutas and Hillyard reported that the N400 ERP amplitude for a word has a nearly inverse linear relationship with its cloze probability [32]. The higher the expectation that a word will appear in a given context, the lower its N400 is relative to words with lower expectation. Generally for nouns, verbs, adjectives, and adverbs, the N400 is also lower for words appearing later in a sentence compared to earlier. Further experimentation is needed to investigate the N400 components for categories of words and their occurrences in different parts of our geography sentences.

The new methods introduced in this paper have been illustrated using two different semantic models. The stronger structural similarity between the brain data and the WordNet-derived data seems to support the claim that during sentence comprehension the representation of words in the brain has a WordNet-like quality. To evaluate what that may mean, we have to ask what the crucial differences are between the two semantic models and the measures of similarity afforded by each.

Wordnet was developed using human judgments about the meanings of the words and relations between them. WordNet encodes fine-grained sense distinctions, grouping two words into the same synonym set if and only if in some context they can be substituted for each other without changing the truth value of the proposition. WordNet has been criticized for encoding sense



**Figure 13. Probability that a test sample is classified as Moscow given that it is London (dash), Paris (double) or Germany (solid) over 4 sets of 30 single-trial classifications (S18) of the words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.** doi:10.1371/journal.pone.0065366.g013



**Figure 14. Hierarchical cluster trees computed from the accumulated average conditional probability density estimates obtained after 22 and 23 classifications of 640 brain wave samples (S18) for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.**  
doi:10.1371/journal.pone.0065366.g014

distinctions that some think are too fine-grained even for humans. For example, consider the following two senses of *north* used in our work:

(n) north#3, due north#1, northward#1, N#2 (the cardinal compass point that is at 0 or 360 degrees).

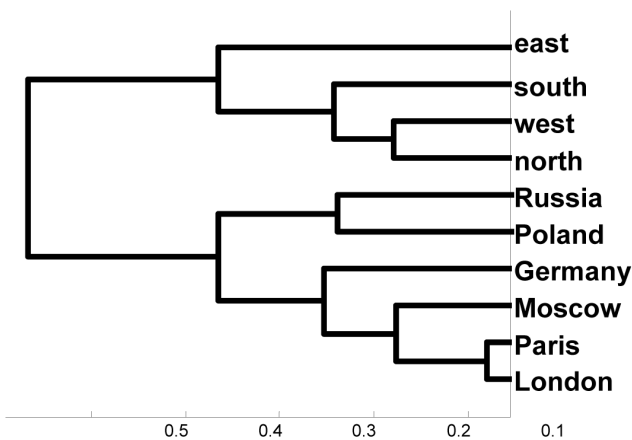
(n) north#5 (the direction corresponding to the northward cardinal compass point).

However, a point is fundamentally different from a direction; the two senses cannot be substituted for each other in all contexts. There is an irreducible distinction between the two senses, even if very few people could articulate it and it is most likely seldom at work in people’s uses of the word. WordNet has been particularly successfully applied computationally in word-sense disambiguation, that is, in determining which of the candidate senses of a word is most appropriate in a given context. WordNet senses have been aggregated in some applications, as indeed they were in our work here. In this way the fine-grained distinctions prevail in the

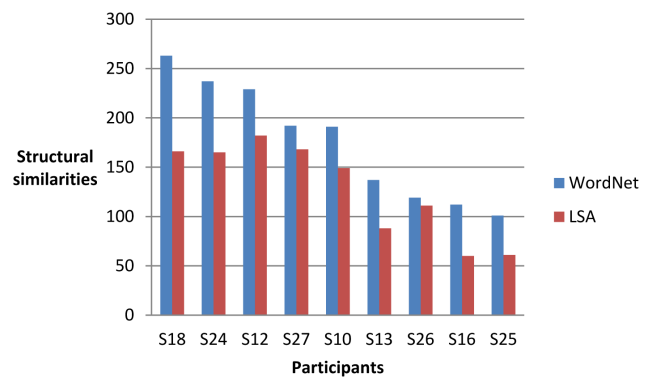
network of sense connections present in WordNet, but the distinctions do not dominate.

The strongest criticism of WordNet has been related to its inadequacies as an ontology [48]. For example, WordNet does not distinguish between one thing being a subtype as opposed to an instance of another thing, as with the Bronze Age, the Iron Age and so on being instances, not subtypes, of time periods. Its taxonomic organization is imperfect. Wordnet also has limited coverage of specialized vocabularies. Another criticism is that WordNet gives obscure senses the same prominence as more frequent senses, as with Paris the city and Paris the character in literature. The last criticism has no impact on our results as only relevant senses were used. As to the ontological criticism, it would be a surprise to discover that when it came to ordinary people, brain representations met strict logical conditions. Only further experimentation will tell.

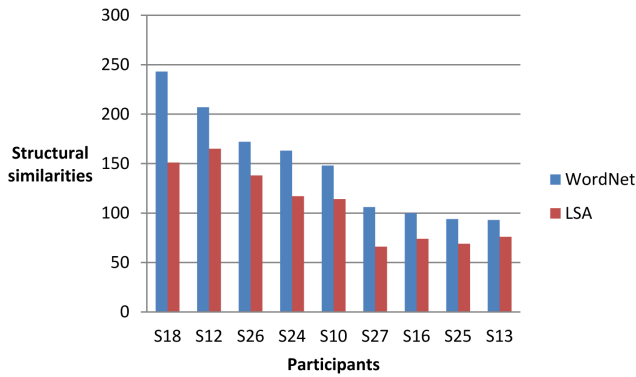
Since Lund and Burgess, distributional corpus-based models, ranging from simple co-occurrence vectors to probabilistic topic-based approaches such as LDA, have increasingly gained acceptance [49]. LSA and other distributional models do not in



**Figure 15. Hierarchical cluster tree computed from the accumulated conditional probability density estimates obtained after 2 classifications of 640 brain wave samples (S18) for {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}.** For all successive runs, the accumulated similarity tree did not change.  
doi:10.1371/journal.pone.0065366.g015



**Figure 16. Significant structural similarities (partial orders of similarity differences invariant between the brain data and the linguistic data) for the set of words {London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia} from 60 single-trial classifications for each participant.**  
doi:10.1371/journal.pone.0065366.g016

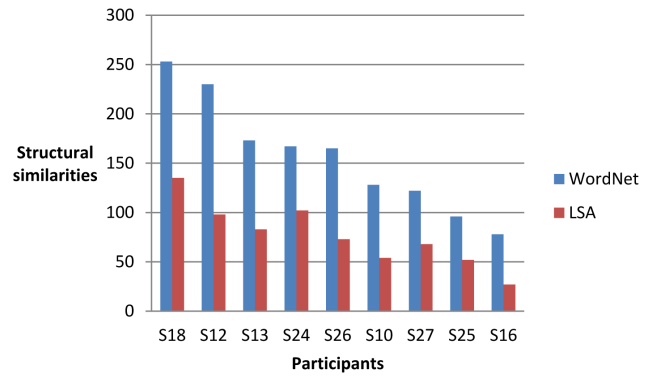


**Figure 17. Significant structural similarities (partial orders of similarity differences invariant between the brain data and the linguistic data) for the set of words {Paris, Vienna, Athens, north, south, east, west, Italy, Spain, Austria} from 60 single-trial classifications for each participant.**  
doi:10.1371/journal.pone.0065366.g017

any obvious way connect all words and all senses of all words via semantic relations. Nor do they systematically represent fine-grained distinctions between different words and different senses of a word. There are, however, many ways in which one distributional model differs from another, some of which were discussed in the Introduction. The size and content of the corpus vary, as do the normalization techniques, dimensionality reduction strategies, and measures of similarity. When we point out the greater structural similarity of the WordNet-derived data with the brain data over the LSA-derived data, our conclusions have to be circumscribed. LSA cannot function as a proxy for all distributional models. What we can say is that the WordNet model seems to have an advantage over corpus-based distributional models, particularly when it comes to looking at the network of relations between words in a sentence comprehension task.

One important point about distributional models is that they require validation. Typically, validation entails seeing how well a given model correlates with the semantic judgments of native speakers. When speakers' predictions diverge from a model, questions arise as to whether that divergence points to limitations in the model or inadequacies in how those judgments were solicited, whether they were fine-grained enough, for example. Judgments obtained directly from neural activity would potentially be more accurate, an approach considered in the work of Devereux and others discussed in the Introduction [21].

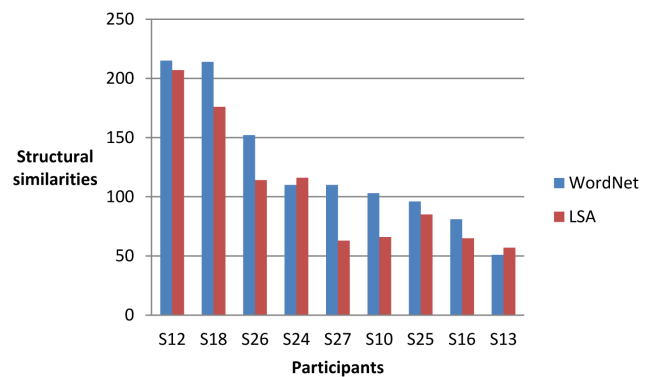
At this point it is useful to explore what differentiates our work from that of Kriegeskorte *et al.* [50]. In both cases, representational models of brain activity are constructed and compared with representational models of some aspect of the experimental situation, whether the stimuli that elicited the neural activity or the behavior exhibited in response to the stimuli, for example. There are, however, important differences between the Kriegeskorte approach and ours. Kriegeskorte and others, two relational structures are compiled, one a representation of brain activity and the other a representation of some aspect of the experimental setup [50]. Their relational structures are in fact representational dissimilarity matrices, one constructed for measures of neural activity and the other for the stimuli eliciting the neural activity or the behavior accompanying the neural activity. Dissimilarity matrices, like our conditional probability density estimates and semantic similarity matrices, offer a general way to characterize the information carried by a given representation, whether it is a representation of the brain activity or a representation of the



**Figure 18. Significant structural similarities (partial orders of similarity differences invariant between the brain data and the linguistic data) for the set of words {Berlin, Rome, Warsaw, north, south, east, west, France, Greece, Poland} from 60 single-trial classifications for each participant.**  
doi:10.1371/journal.pone.0065366.g018

stimuli, such as words in their sentential context. But the methods they use to compare their representational dissimilarity matrices differ from the methods we use to compare the invariant partial orders of our representational structures. Kriegeskorte and others argue against a linear match between dissimilarity matrices; instead, like us, they use the Spearman rank correlation coefficient [50]. However, we go a fundamentally important step further and, for those similarity matrices that are significantly strongly correlated, we construct the partial orders that are invariant with respect to the brain data and the experimental data. These invariant structures give a detailed assessment of where the similarities between the brain data and the linguistic data actually lie. These invariant structures will be invaluable in future work, such as when we take into account more information in the experimental setup, for example, the participants' responses – namely whether they judged a statement true or false and whether they were right or wrong. Formulating and testing complex hypotheses about brain activity become possible with the methods presented in this paper.

In considering how the brain as a system processes language, the idea of an associative network has long held appeal. The work on associative networks by Collins and Loftus [51] and the investigation of the semantics of associative networks by Woods



**Figure 19. Significant structural similarities (partial orders of similarity differences invariant between the brain data and the linguistic data) for the set of words {Madrid, Rome, Vienna, north, south, east, west, Spain, Italy, Austria} from 60 single-trial classifications for each participant.**  
doi:10.1371/journal.pone.0065366.g019

[52] are early contributions. Associative networks for ambiguity resolution have received a lot of attention, beginning with Hirst [53]. Associative networks for information retrieval and web search have gained increasing attention, starting with Crestani [54]. Activation and spreading activation, the principal operations proposed for associative networks, have been characterized in various ways over time, but no specific method or algorithm has yet emerged as definitive. What is still missing from discussions of associative networks in language processing is a coherent, general account of how associative networks are used in complex tasks such as answering a question or determining the truth or falsity of an assertion. In this paper, associative networks such as those in Figure 10 and Figure 11 have been presented as the brain's dynamic representation of a set of words during the response to a sentence such as *Rome is not the capital of Spain*. Such associative networks are not static, but are configurations that change under different stimulus conditions and under the operations of activation and spreading activation. However, these changes are not unconstrained. The associative networks are approximately invariant across language users and across different occasions on which the words are used. What this paper suggests is that the associative networks underlying sentence comprehension, particularly for the task of determining the truth or falsity of an assertion, are possibly more like the associative networks of WordNet than those obtained by statistical methods from large corpora.

WordNet is far from the end of the story, however. Binary, ternary and quaternary relations are generally missing from WordNet, but are common in ordinary language. The sentence *Rome is a capital* may be represented with some ease using WordNet and its *is\_a* relation, and the truth or falsity of the assertion determined by the simple test: is there an *is\_a* relation between *Rome* and *capital*? But *Rome is the capital of Italy*, expressing a binary relation, has no straightforward representation in WordNet, and none that permits its truth or falsity to be determined directly. Similarly for ternary relations such as those expressed in the sentence *Sue gave Bill a book* or, using our geography words, *Paris is more north of Rome than west of Berlin*. Associative networks that include binary and higher-order relations, along with definitions of

activation and spreading activation for each order, are still needed. Large enough or focused enough corpora could potentially provide such associations. We noted earlier the work of Kelly and others who used parsed data to identify three-part associations involving the well-known taxonomic relations of *is-a* and *part-of* in Wikipedia and the British National Corpus [20]. The recognition of higher-order relations, such as those in *Paris is the capital of France* and *Paris is further west of Moscow than Berlin is*, would be a natural progression from that work. The methods of relative structural similarity developed in this paper can be used to assess the extent to which any network of semantic relations is structurally similar to brain data collected during a language task.

### Supporting Information

**Figure S1** a – j: WordNet-based and LSA-based semantic similarities and brain conditional probability estimates for (a) *London*; (b) *Moscow*; (c) *Paris*; (d) *north*; (e) *south*; (f) *east*; (g) *west*; (h) *Germany*; (i) *Poland*; and (j) *Russia*, relative to *London*, *Moscow*, *Paris*, *north*, *south*, *east*, *west*, *Germany*, *Poland*, and *Russia*. Data taken from Figure 1, Figure 2, and Figure 4. (TIFF)

**Material S1** Trigger sentences and timings. Table S1: Timings for trigger sentences. Table S2: Timings for the presentation of the individual words in each of the 48 trigger sentences. (DOCX)

**Material S2** WordNet entries. (DOCX)

**Material S3** Five Wordnet-based measures of similarity. (DOCX)

### Author Contributions

Conceived and designed the experiments: PS. Performed the experiments: PS. Analyzed the data: CC PS. Contributed reagents/materials/analysis tools: MP. Wrote the paper: CC.

### References

- Miller GA (1995) WordNet: A lexical database for English. *Communications of the ACM* 38: 39–41.
- Fellbaum C, editor (1998) WordNet: An electronic lexical database. Cambridge, MA: MIT Press.
- Landauer TK, Dumais ST (1997) A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104: 211–240.
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R (1990) Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41: 391–407.
- Suppes P, Han B, Epelboim J, Lu Z-L (1999a) Invariance between subjects of brain wave representations of language. *Proceedings of the National Academy of Sciences* 96: 12953–12958.
- Suppes P, Lu Z-L, Han B (1997) Brain wave recognition of words. *Proceedings of the National Academy of Sciences* 95: 14965–14969.
- Suppes P, Han B, Lu Z-L (1998) Brain-wave recognition of sentences. *Proceedings of the National Academy of Sciences* 95: 15861–15866.
- Suppes P, Han B (2000) Brain-wave representation of words by superposition of a few sine waves. *Proceedings of the National Academy of Sciences* 97: 8738–8743.
- Wong DK, Perreau-Guimaraes M, Uy ET, Suppes P (2004) Classification of individual trials based on the best independent component of EEG-recorded sentences. *Neurocomputing* 61: 479–484.
- Wong DK, Uy ET, Perreau-Guimaraes M, Yang W, Suppes P (2006) Interpretation of perceptron weights as constructed time series for EEG classification. *Neurocomputing* 70: 373–383.
- Perreau-Guimaraes M, Wong DK, Uy ET, Grosenick L, Suppes P (2007) Single-trial classification of MEG recordings. *IEEE Transactions on Biomedical Engineering* 54: 436–443.
- Suppes P, Perreau-Guimaraes M, Wong DK (2009) Partial Orders of Similarity Differences Invariant Between EEG-Recorded Brain and Perceptual Representations of Language. *Neural Computation* 21: 3228–3269.
- Wong DK, Grosenick L, Uy ET, Perreau-Guimaraes M, Carvalhaes CG, et al. (2008) Quantifying inter-subject agreement in brain-imaging analyses. *NeuroImage* 39: 1051–1063.
- Suppes P, Han B, Epelboim J, Lu Z-L (1999b) Invariance of brain-wave representations of simple visual images and their names. *Proceedings of the National Academy of Sciences* 96: 14658–14663.
- Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, et al. (2008) Predicting human brain activity associated with the meanings of nouns. *Science* 320: 1191–1195.
- Brants T, Franz A (2006) "Web 1T 5-gram Version 1." Linguistic Data Consortium, Philadelphia. Available: <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>. Accessed 2013 May 6.
- Just MA, Cherkassky VL, Aryal S, Mitchell TM (2010) A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE* 5: e8622. doi:10.1371/journal.pone.0008622.
- Pereira F, Botvinick M, Detre G (2010) Learning semantic features for fMRI data from definitional text. In: *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neuro Linguistics*, June 2010, Los Angeles, USA, Association for Computational Linguistics 1–9. Available: <http://www.aclweb.org/anthology/W10-0601>.
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *The Journal of Machine Learning Research* 3: 993–1022.
- Kelly C, Devereux B, Korhonen A (2010) Acquiring human-like feature-based conceptual representations from Corpora. In *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neuro Linguistics*, June 2010, Los Angeles: Association for Computational Linguistics, 61–69.
- Devereux B, Kelly C, Korhonen A (2010) Using fMRI activation to conceptual stimuli to evaluate methods for extracting conceptual representations from

- corpora. In: Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics, June 2010. Los Angeles: Association for Computational Linguistics 70–78.
22. Murphy B, Baroni M, Poesio M (2009) EEG responds to conceptual stimuli and corpus semantics. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–7 August 2009. c 2009 ACL and AFNLP 619–627.
  23. Murphy B, Poesio M (2010) Detecting semantic category in simultaneous EEG/MEG recordings. In: First workshop on computational neurolinguistics, NAACL HLT 2010. Los Angeles: Association for Computational Linguistics 36–44.
  24. Murphy B, Poesio M, Bovolo F, Bruzzone L, Dalponte M, et al. (2011) EEG decoding of semantic category reveals distributed representations for single concepts. *Brain Language* 117: 12–22. doi: 10.1016/j.bandl.2010.09.013.
  25. Murphy B, Talukdar P, Mitchell T (2012) Selecting corpus-semantic models for neurolinguistic decoding. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval '12). Association for Computational Linguistics, Stroudsburg, PA, USA 114–123.
  26. Baroni M, Murphy B, Barbu E, Poesio M (2010) Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science - COGSCI* 34: 222–254. doi:10.1111/j.1551-6709.2009.01068.x.
  27. Jelodar AB, Alizadeh M, Khadivi S (2010) WordNet based features for predicting brain activity associated with meanings of noun. In: Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics, Los Angeles, Association for Computational Linguistics 18–26.
  28. Chan AM, Halgren E, Marinkovic K, Cash KS (2011) Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage* 54: 3028–3039.
  29. Sudre G, Pomerleau D, Palatucci M, Wehbe L, Fyshe A, et al. (2012) Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage* 62: 451–463.
  30. Huth AH, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76: 1210–1224.
  31. Miller GA, Nicely P (1955) An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27: 338–352.
  32. Kutas M, Hillyard SA (1984) Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307: 161–163.
  33. Borovsky A, Elman JL, Kutas M (2012) Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development* 8: 278–302.
  34. Kutas M, Federmeier KD (2011) Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology* 62: 621–647.
  35. Brennan J, Pykkänen L (2012) The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage* 60: 1139–1148.
  36. Suppes P (1974) The axiomatic method in the empirical sciences. In: Henkin L, et al., editors. Proceedings of the Tarski Symposium, Proceedings of Symposia in Pure Mathematics, 25. Providence, RI: American Mathematical Society. 465–479.
  37. Suppes P (1972) Axiomatic set theory. New York: Dover. 267 p.
  38. Luce RD (1956) Semiorders and a theory of utility discrimination. *Econometrica* 24: 178–191.
  39. Scott D, Suppes P (1958) Foundational aspects of theories of measurement. *The Journal of Symbolic Logic* 23: 113–128.
  40. Princeton University "About WordNet." WordNet. Available: <http://wordnet.princeton.edu>. Accessed 2013 May 3.
  41. Budanitsky A, Hirst G (2001) Semantic distance in WordNet: An experimental application-oriented evaluation of five measures. In Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources, Pittsburgh.
  42. Similarity." Available: <http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi>. Accessed 2013 May 3.
  43. Patwardhan S, Michelizzi J (2004) WordNet: Similarity - Measuring the Relatedness of Concepts. In Proceedings of the Nineteenth National Conference on Artificial Intelligence - Intelligent Systems Demonstration (AAAI-04).
  44. Ballatore A, Wilson DC, Bertolotto M (2012) The similarity jury: Combining expert judgements on geographic concepts. In: Castano, S et al., editors. ER Workshops 2012, LNCS 7518, 231–240. doi:10.1007/978-3-642-33999-8\_29.
  45. Landauer TK, Foltz PW, Laham D (1998) Introduction to Latent Semantic Analysis. *Discourse Processes* 25: 259–284.
  46. Karp RM (1972) Reducibility among combinatorial problems. In: Miller RE, Thatcher JW, editors. Complexity of Computer Computations, Proc. Symposium. IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y. New York: Plenum, 85–103.
  47. Matsunaga T, Yonemori C, Tomita E, Muramatsu M (2009) Clique-based data mining for related genes in a biomedical database. *BMC Bioinformatics* 10: 205 1–9 doi:10.1186/1471-2105-10-205.
  48. Martin P (2003) Correction and extension of WordNet 1.7. ICCS 2003, 11th International Conference on Conceptual Structures, Springer Verlag, LNAI 2746, July 21–25, 2003, Dresden, Germany, 160–173.
  49. Lund K, Burgess C (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers* 28: 203–208.
  50. Kriegeskorte N, Mur M, Bandettini PA (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi: 10.3389/neuro.06.004.2008.
  51. Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. *Psychological Review* 82: 407–428.
  52. Woods W (1975) What's in a link: Foundations for semantic networks. In: Bobrow D, Collins A, editors. Representation and understanding. New York: Academic Press 35–82.
  53. Hirst G (1988) Resolving lexical ambiguity computationally with spreading activation and polaroid words. In: Small SL, Cottrell GW, Tanenhaus MK, editors. Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence. San Mateo, CA: Morgan Kaufmann Publishers. 73–107.
  54. Crestani F (1997) Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review* 11: 453–482.