The Primacy of Experience in Language Processing: Semantic Priming Is Driven Primarily by Experiential Similarity

Leonardo Fernandino[1,2]* and Lisa L. Conant[1]

[1] Department of Neurology, Medical College of Wisconsin.

[2] Department of Biomedical Engineering, Medical College of Wisconsin

*Correspondence to: lfernandino@mcw.edu

Abstract

The organization of the representational space underlying the neural coding of word meaning has long been a central question in cognitive science. Although there is general agreement that lexical semantic representations must make contact with sensory-motor and affective experiences in a non-arbitrary fashion, the nature of this relationship remains controversial. Many researchers have proposed that lexical concepts (i.e., word meanings) are represented primarily in terms of their experiential content, ultimately derived from sensory-motor and affective processes. However, the recent success of distributional language models in emulating human linguistic behavior has led to proposals that word co-occurrence information may play a more important role in the representation of concepts in the human brain. We investigated this issue by using representational similarity analysis (RSA) of semantic priming data. Participants performed a speeded lexical decision task in two sessions separated by approximately one week. All target words were presented once in each session, but each time they were preceded by a different prime word. Priming was computed for each target as the difference in RT between the two sessions. We evaluated eight models of semantic word representation in terms of their ability to predict the magnitude of the priming effect for each target: two based on experiential information, three based on distributional information, and three based on taxonomic information. Crucially, we used partial correlation RSA to account for intercorrelations between predictions from different models, which allowed us to assess, for the first time, the unique predictive power of experiential and distributional models. We found that semantic priming was driven primarily by experiential similarity between prime and target, with no evidence of an independent effect of distributional similarity. Furthermore, only the experiential models accounted for unique variance in priming after partialling out predictions from explicit similarity ratings. These results support experiential accounts of language representation and indicate that, despite their good performance at some linguistic tasks, distributional models do not encode the same kind of information used by the brain's semantic system.

Introduction

The nature of the relationship between language meaning and the physical reality experienced through the senses has been the subject of scholarly debate since the dawn of the Western intellectual tradition. From the writings of Plato and Aristotle through the works of other luminaries such as Descartes, Locke, Hume, Kant, and Wittgenstein, the question of how the meanings of words and sentences are implemented in the mind and brain has figured prominently, and it remains a central issue in the contemporary cognitive sciences (e.g., Barsalou, 1999; Chomsky, 1975; Fodor, 1980; Glenberg & Robertson, 2000; Jackendoff, 2002; Miller & Johnson-Laird, 1976; Rogers & McClelland, 2004; Skinner, 1957; see also the special issue of *Psychonomic Bulletin and Review*, *23*(4)). Currently, there is consensus among scholars that, aside from physiological processes taking place in sensory and motor organs, any representations or computations underlying language production and comprehension must be somehow implemented in terms of neurophysiological processes in the brain. The ongoing debate concerns the characterization of these language-related neural representations and processes in terms of the kinds of information they encode and the kinds of computations that are performed. Much of this debate has focused on whether, and to what extent, they encode information originating from sensory-motor processes in the course of our interactions with the world (i.e., experiential information), as opposed to information structures that are independent of the organization of the sensory-motor systems.

According to one prominent view, language meaning is represented in the brain as a symbolic system, in which the neural implementation of elementary units of meaning (i.e., lexical concepts) is arbitrarily related to the neural representations and processes involved in perception and action (Fodor, 1980; Newell, 1980; Pylyshyn, 1984; Simon & Newell, 1971). Symbolic-conceptual representations are specified exclusively in terms of other symbolic-conceptual representations, and their correspondence to experiential representations is given via arbitrary associations. The relationship between conceptual representations and sensory-motor representations is seen as analogous to the relationship between software and hardware in general-purpose computers, in which the former can be completely specified without any reference to the latter. In its original formulation this view has been largely discredited because it leads to circular definitions and begs the question of how symbolic representations map onto sensory-motor experience, also known as the symbol-grounding problem (Harnad, 1990; Searle, 1980).

In contrast to the symbolic view, many authors have argued that experiential information plays an essential role in semantic language processing. They propose that accessing the

meaning of a word involves the activation of neural networks storing an ensemble of elementary experiential representations associated with the word form in question (Barsalou, 1999; Damasio, 1989; Glenberg, 1997; Pulvermüller, 1999). These neural networks, and the representations they encode, are thought to overlap with those involved in processing sensory-motor information during perception and action, implying a substantial degree of continuity between sensory-motor and semantic systems. This view, often referred to as “grounded”, “embodied”, or “situated” semantics, is boosted by several lines of evidence pointing to specific interrelationships between language meaning and sensory-motor representations that are not predicted by the symbolic view (for reviews, see Barsalou et al., 2003; Binder & Desai, 2011; Hauk & Tschentscher, 2013; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012).

However, several authors have been skeptical of the idea that experiential information is an essential component of language meaning. It has been argued, for example, that experimental results taken to indicate activation of sensory-motor information during word comprehension may be driven, instead, by the ortho-phonological structure of the corresponding word forms or by their grammatical class (Bedny et al., 2008; Zubicaray et al., 2013). Another objection is based on the idea that the activation of sensory-motor representations by words and sentences may be epiphenomenal to semantic processing, simply reflecting spreading of activation from conceptual to perceptual and motor representations (Mahon, 2014; Mahon & Caramazza, 2008; Weiskopf, 2010). Some authors cite the existence of category-specific semantic deficits as evidence for a taxonomic, rather than experiential, organization of semantic memory (Caramazza & Mahon, 2003), while others have pointed to the absence of obvious semantic deficits in persons with congenital sensory-motor impairments as a challenge to embodied semantics (Bedny, 2020).

More recently, it has been proposed that statistical patterns of co-occurrence between word forms in natural language may be used by the brain to represent word meaning (Andrews et al., 2014; Burgess & Lund, 2000; Landauer & Dumais, 1997; Louwerse, 2008; Mandera et al., 2017). This proposal builds on the idea – known as the “distributional hypothesis” – that words with similar meanings tend to occur in similar linguistic contexts, that is, surrounded by similar sets of words. A corollary of this idea is that, given a large enough body of language samples, the degree of overall similarity in meaning between two words can be determined from their respective linguistic contexts (Harris, 1954; Sahlgren, 2008). With the advent of the internet, the availability of extremely large text corpora in digital form have enabled the emergence of high-performance implementations of this idea, known as distributional language models. These models have been shown to approach human-level performance on a number of semantic

language tasks (Mandera et al., 2017; Pereira et al., 2016), forming the basis of several artificial intelligence applications currently in use.

Although distributional structure has been touted as the solution to the problem of lexical semantic representation (Landauer & Dumais, 1997), it is now generally recognized that a representational system based exclusively on distributional information runs into the same logical pitfall as the original symbolic view, that is, the symbol-grounding problem. Thus, it has been proposed that lexical concepts are represented in terms of a hybrid experiential-distributional code, in which some concepts are directly grounded in experiential information while others are indirectly grounded via their statistical patterns of co-occurrence with experientially grounded concepts (Andrews et al., 2014; Banks et al., 2021; Barsalou et al., 2008; Louwerse, 2008)

In sum, the current effort to characterize the underlying nature of language meaning can be framed in terms of which information structures must be postulated in order to achieve a neuroscientifically plausible explanation of human semantic behavior and its associated patterns of neural activity. While taxonomic structure is immediately apparent in semantic word processing, there are important reasons (such as the symbol-grounding issue) to believe that lexical concepts are not encoded in terms of symbolic tokens, and that category-related effects are likely to be an emergent property of the underlying representational code. It is easy to see how taxonomic structure can emerge from the patterns of covariation of experiential features across lexical concepts, since items belonging to the same category tend to be more similar in terms of sensory-motor and affective features than items belonging to different categories (Connell et al., 2019). Likewise, taxonomic structure could also emerge from distributional structure, since items in a given category tend to appear in similar linguistic contexts, as demonstrated by distributional semantic models.

Because experiential similarity structure is also correlated with distributional similarity structure (i.e., items with similar experiential features tend to appear in similar linguistic contexts), studies investigating the representational structure of lexical concepts have produced results consistent with both experiential and distributional representations (Louwerse, 2008, 2011). One exception is a recent study that used representational similarity analyses (RSA) of fMRI activation patterns to compare experiential and distributional models in terms of how much unique variance they explained in the similarity structure of two large sets of nouns (Fernandino et al., 2022). RSA with partial correlations showed that a model based on 48 experiential features (Exp48) accounted for all the variance explained by either of the distributional models tested (word2vec and GloVe) plus an additional 17% of the explainable variance, indicating that

experiential information does contribute to the neural activation patterns underlying lexical concepts, with no evidence of distributional information.

Here, we investigated the extent to which lexical concepts reflect experiential and distributional information using an independent measure of representational similarity, namely, automatic semantic priming in lexical decision. Semantic priming has been used extensively to study the organization of semantic memory, and its magnitude has been shown to reflect the degree of semantic relatedness between the prime and the target according to Hyperspace Analogue to Language (HAL) vector representations (Günther et al., 2016a, 2016b; Heyman et al., 2016; Hutchison et al., 2008; Jones et al., 2006). We used RSA to evaluate and contrast 8 models of semantic representation, including two experiential (Exp48 and Lancaster), three distributional (word2vec, GloVe, and fastText) and three taxonomic (WordNet, Coarse Categorical, and Fine-Grained Categorical) models. Crucially, we took into account the fact that model-based similarity structures are intercorrelated, using partial correlation to evaluate the unique effects of experiential and distributional models. Since, in principle, priming can be driven by semantic similarity (e.g., mouse – rat) as well as by thematic/contextual association (e.g., mouse – cheese) (Hutchison, 2003; Lucas, 2000; Thompson-Schill et al., 1998), we also collected independent ratings of similarity and association for all prime-target pairs. These ratings were used to evaluate (1) the unique contributions of semantic similarity and thematic association to priming, (2) how well each factor predicts priming relative to the semantic models and (3) how well each model predicts priming when the effect of either type of relatedness is partialled out.

Methods

Semantic Priming

Participants

Thirty-one monolingual English speakers (19 females, ages 22-51, mean age = 31) participated in the study. All were right-handed according to the Edinburgh Handedness Scale, with no history of neurological or psychiatric disorders, and at least 12 years of formal education (mean = 16.5 years).

Stimuli

Targets consisted of 210 English nouns and 210 pronounceable pseudowords. Each target was associated with two primes (all real nouns). The task was designed to be performed in two sessions (on different days), such that the same words appeared as targets in both sessions, but each time preceded by a different prime. Targets and primes were selected from the list of

434 nouns for which experiential attribute ratings were available (Binder et al., 2016). Based on the results of Hutchison et al. (2008), who investigated the impact of several variables on item-level semantic priming, we defined a word's "priming susceptibility score" as its orthographic neighborhood size (Coltheart's N; henceforth, OrthN) z-score minus its log-transformed HAL frequency (LogHF) z-score. The 210 nouns with the highest priming susceptibility score were selected as targets (**Supplemental Table 1**). Of these, 157 were concrete and 53 were abstract concepts.

Primes were selected from the 420 nouns with the lowest priming susceptibility score (i.e., relatively high frequency and low OrthN). From this pool of potential primes, a Python script selected two primes for each target based on several criteria. Semantic similarity among words was estimated by averaging together the full pairwise cosine similarity matrices obtained from HAL (http://www.lingexp.uni-tuebingen.de/z2/LSAspaces) and from WordNet (Wu-Palmer distance; http://wordnet.princeton.edu). The word with the highest semantic similarity to the target that had not been assigned to a different target was selected as the "close" prime. The "distant" prime was then selected from the list of remaining prime candidates sorted by semantic similarity to the target in descending order. The algorithm searched for the prime candidate with lowest semantic similarity to the target that (a) had not been assigned to a different target, (b) had letter length, number of syllables, LogHF, and OrthN were similar to the close prime, and that (c) was no less then *k* words from the end of the similarity-ranked list. For each new target, *k* was made incrementally larger to select progressively more distant primes, thus providing a wide range of distances from the target. Therefore, rather than selecting primes that were either "related" or "unrelated" to the target (as is common in semantic priming studies), we selected prime pairs whose difference between their similarity to the target (i.e., cosine[target, close prime] – cosine[target, distant prime]) varied continuously across a range (0 to 0.78). None of the primes were strongly associated to their respective targets (all prime-target pairs had a forward association strength lower than 0.1 according to the University of South Florida Free Association Norms).

"Close" and "distant" primes were evenly distributed across two testing sessions, A and B, so that each session contained 105 trials with close primes and 105 trials with distant primes. Mean cosine similarity between prime and target was similar for the two sessions (session A: 0.400 (SD = 0.219); session B: 0.401 (SD = 0.206); p = .98, *t* test, two-tailed). Priming was defined as the difference in standardized response times to the same target between the two sessions (session order was counterbalanced across participants). The primes in the two sessions were matched according to 24 lexical variables, including number of letters, number of

phonemes, frequency, orthographic neighborhood size, bigram and trigram frequency, and age of acquisition (**Supplemental Table 2**).

Pseudowords were created with the MCWord database (https://www.neuro.mcw.edu/mcword) and were matched to the real noun targets in number of letters, bigram and trigram frequency, and orthographic neighborhood size.

Procedures

Testing was performed over two sessions, approximately one week apart (range: 4-10 days). Each target word was presented once in each session, each time preceded by a different prime. The priming effect for a given target word was computed as the difference in z-transformed response times between the two sessions. Each trial started with a central fixation cross (duration jittered 1-2 s), followed by the prime (150 ms), a mask (a sequence of hash marks matching the prime in length, 50 ms), and the target (2 s). The prime was presented in lowercase and the target in uppercase letters. Trials were presented in a different pseudorandomized order for each participant, and the order of presentation of the two primes for a given target was counterbalanced across participants.

Stimuli were presented in light gray with a dark background on the center of a computer screen located 80 cm in front of the participant. Participants were instructed to ignore the prime and make a speeded lexical decision on the target, indicating their response by pressing one of two keys on a response pad with their right index and middle fingers. They were instructed to respond as fast as possible without making mistakes. Stimulus presentation and response recording were performed with E-prime 2.0 software running on a Windows computer and a PST Serial Response Box (Psychology Software Tools, Inc.). For each session, the task was divided into 8 blocks, each lasting 5 minutes. At the beginning of the first session, participants provided informed consent and filled a health history questionnaire. They performed a short practice session (12 trials) immediately before the actual task on both sessions. None of the stimuli included in the experiment appeared in the practice session.

Similarity and association ratings

Participants

Participants were recruited anonymously through the online crowdsourcing service Amazon Mechanical Turk (www.mturk.com) and were financially compensated. They were required to have completed at least 5,000 previous surveys on the Mechanical Turk platform with at least a 95% acceptance rate, and to have account addresses in the United States. All participants

reported that they were native speakers of English. Twenty-three participants completed the semantic similarity survey, and 25 participants completed the thematic association survey.

Stimuli

The stimuli were the same 210 word triplets (i.e., a target and its two respective primes) included in the priming task.

Procedures

Semantic similarity. In each trial, the three words in a triplet were presented simultaneously in a triangular arrangement (**Supplemental Figure 1, left panel**). The word previously used as a target in the lexical decision task appeared at the top center position and its two corresponding primes appeared at the two bottom positions on each side of the screen. A horizontal sliding scale was displayed on the center of the screen. Participants were instructed to position the slider at the point along the scale that best reflected how similar the top word was to the two bottom words. If the top word was equally similar (or equally dissimilar) to the two bottom words, the slider should be positioned in the center of the scale. If the top word was slightly more similar to one bottom word (say, the one on the left) than to the other, the slider should be positioned slightly off center in the direction of the more similar word. The larger the difference in similarity between the two word pairs (i.e., top word-left word similarity versus top word-right word similarity), the further away from the center the slider should be placed. The instructions noted that similarity should be judged based on the extent to which the two concepts had similar properties or could be considered as the same “kind of thing” (e.g., “mouse” and “rat” should be considered highly similar); concepts that often appear together (i.e., thematically or contextually associated, such as “mouse” and “cheese”) but do not have similar properties should not be considered similar.

The task was self-paced, and participants responded by moving the cursor and clicking on a point along the horizontal scale bar on the screen. A circular marker appeared on the scale to indicate the location clicked. Participants could change the position of the marker as many times as desired before clicking on a button labeled “Next” to conclude the trial and start the next one. The response (i.e., the position of the slider) was recorded as a continuous variable, with the center of the scale corresponding to 0, the leftmost position corresponding to -10, the rightmost position corresponding to 10.

Thematic association. In each trial, two horizontal scales were displayed on the screen, one on top and the other on the bottom (**Supplemental Figure 1, right panel**). To the left of

each scale, two words were presented simultaneously, one being a target in the lexical decision task, the other being one of its two primes. The same target word appeared in both word pairs. For each word pair, participants were instructed to click on a point along the corresponding scale indicating how closely the two words are associated with each other. The instructions noted that association should be judged based on the extent to which the two concepts appear together, not based on whether they have similar properties (e.g., “cake” and “candle” should be considered as strongly associated); concepts that have similar properties but do not often occur in the same context (e.g., “dagger” and “scalpel”) should be considered as weakly or not at all associated. Responses were recorded as a number between 0 (not associated) and 10 (extremely associated). The task was self-paced and both word pairs needed to be rated before participants could proceed to the next trial. Both rating tasks were programmed in Psychopy 3 (Peirce, 2007) and hosted on Pavlovia.org.

Data cleaning. Data collected via online crowdsourcing, in which subjects participate anonymously, are more likely to include non-compliant participants. To identify poor quality data (i.e., data from participants who did not follow the task instructions), we computed the Pearson correlation between the ratings from each participant and the group median ratings.

Semantic models

Experiential model. The *Exp48* model is based on the relative importance of 48 different features of phenomenological experience to the meaning of a given word. It is based on the experiential feature norms obtained via online crowdsourcing by (Binder et al., 2016). Each model dimension encodes the relative importance of an experiential feature according to ratings on a Likert-type scale. It consists of 48 dimensions that can be mapped onto independently established neurocognitive processes (i.e., processes operationalized independently of semantic tasks). Exp48 encompasses perceptual, motor, spatial, temporal, causal, valuative, and goal-related dimensions (**Supplemental Table 3**). The feature ratings stand, roughly, for the extent to which a given functional brain system (e.g., the visual motion system) is activated when a concept is retrieved.

The *Lancaster* model is based on the Lancaster Sensorimotor Norms (Lynott et al., 2019). Similarly to Exp48, it quantifies the importance of each feature to the meaning of a word. However, it yields a much coarser representational space than Exp48, consisting of 9 dimensions corresponding to major sensory and motor modalities: vision, hearing, touch, taste, smell, interoception, hand actions, mouth actions, and foot actions. These norms are also derived from ratings collected via online crowdsourcing.

Taxonomic models. *WordNet* (Miller et al., 1990) is the largest and most influential database of taxonomic information for lexical concepts. It is organized as a knowledge graph in which words are grouped into sets of synonyms ("synsets"), each expressing a distinct concept. Synsets are interconnected according to conceptual-semantic relations. Our WordNet model is based on the superordinate-subordinate relation (hypernymy-hyponymy), which links more general synsets (e.g., "vehicle") to increasingly specific ones (e.g., "car" and "sedan"). This hierarchical structure is represented as a tree, and all noun hierarchies ultimately go up to the root node ("entity"). We used the Natural Language Toolkit (NLTK 3.4.5; https://www.nltk.org) to compute semantic similarity between prime and target according to three different methods (path length, Leacock-Chodorow, and Wu-Palmer), and the measure providing the highest correlation between priming predictions and the priming data (Wu-Palmer) was used in subsequent analyses.

We also evaluated two ad hoc models based on categorical membership. Unlike WordNet, these models computed semantic similarity based only on the categories and subcategories included in the stimulus set. Model Categorical A was coarser, consisting of 10 categories [subcategories]: Abstract [Mental State, Other Abstract], Event, Animate Object, Inanimate Object [Artifact, Food, Other Inanimate], and Place. Model Categorical B was more fine-grained, consisting of 21 categories [subcategories]: Abstract [Mental State, Other Abstract], Event [Social Event, Physical Event, Time Period], Animate Object [Human, Animal, Body Part], Inanimate Object [Artifact [Tool, Building Part, Furniture/Appliance, Instrument, Vehicle, Other Artifact], Food, Other Inanimate], and Place [Natural, Manmade].

Distributional models. *Word2vec* (Goldberg & Levy, 2014; Mikolov et al., 2013) is a distributional model that, rather than directly computing word co-occurrence frequencies, uses a 3-layer neural network trained to predict a word based on the words preceding and following it. We used the 300-dimensional word vectors trained on the Google News dataset (approximately 100 billion words) based on the continuous skipgram algorithm and distributed by Google (https://code.google.com/archive/p/word2vec). In contrast to word2vec, which uses only local information, *GloVe* (Pennington et al., 2014) – short for Global Vectors – is based on the ratio of co-occurrence probabilities between pairs of words across the entire corpus. We used the 300-dimensional word vectors trained on Common Crawl (840 billion words) and made available by the authors (https://nlp.stanford.edu/projects/glove). In a comparative evaluation of distributional semantic models (Pereira et al., 2016), word2vec and GloVe were the two top performing models in predicting human behavior across a variety of semantic tasks. Finally, fastText is a more recently developed model based on a continuous-bag-of-words (CBOW)

algorithm that has been shown to outperform both word2vec and GloVe in standard benchmark tests of natural language processing (Mikolov et al., 2017). It contains several improvements over the original word2vec model (e.g., position-dependent weighting and use of subword information). We used the pre-trained vectors provided by the authors (https://fasttext.cc/docs/en/english-vectors.html), which were trained with subword information on Common Crawl (600 billion words).

Data Analysis

Analyses were conducted in Python using custom scripts. Only trials that received correct responses were included in the analyses. For each testing session, we computed the mean and standard deviation of response times (RTs) for correct trials. Trials with RTs more than 5 standard deviations away from the session mean for each participant were considered extreme values and excluded from further analyses.

Priming was computed for each target as the difference in standardized response times between the two sessions. For all semantic models other than WordNet, semantic similarity was computed as the cosine between the vector representations of the two words. Priming predictions were computed for each model as P = similarity(Target, Prime1) – similarity(Target, Prime2). The correspondence between model predictions and observed priming (i.e., RSA correlation) was computed via Spearman correlations. Correlations were tested for significance across participants (subject-wise analysis) using the Wilcoxon Signed Rank test and across trials (item-wise analysis) using the Mantel test. We also ran separate analyses for concrete and abstract target words to explore a possible effect of target concreteness on the results.

The noise ceiling of a dataset corresponds to the highest RSA correlation that any model could achieve given the amount of noise in the data (Nili et al., 2014). We computed the upper-bound noise ceiling estimate as the mean, across all participants, of the Spearman correlations between priming magnitude for each participant and the group mean. The lower-bound noise ceiling estimate was computed as the mean, across all participants, of the Spearman correlations between priming magnitude for each participant and the mean priming magnitude across all *other* participants.

Data reliability was also assessed via split-half correlations with 10,000 iterations. On each iteration, the subject sample was randomly split into 2 halves (15 and 16 subjects each) and the average item-level priming magnitudes were calculated for each half separately. The correlation between priming magnitudes of the two halves was computed and the Spearman–Brown formula was applied to the result. We report the average correlation across all iterations.

Partial correlations were used to evaluate the extent to which a model predicted the observed priming magnitudes once the predictions of a different model were taken into account. This provided a measure of how much unique information about word similarity patterns (as measured via priming) was encoded in each model.

We also used Spearman correlation to assess the extent to which the subjective similarity and association ratings predicted semantic priming. We used partial correlation to verify whether each representational model predicted priming above chance level while controlling for similarity or association rating.

Results

Lexical decision. Group mean accuracies (Acc) and response times (RTs) in the lexical decision task are presented in **Table 2**. The classic RT advantage for words over pseudowords was observed ($p = 2 \times 10^{-9}$, paired-samples t-test, two-tailed). RTs were also faster in session 2 compared to session 1 for both words and pseudowords (both $p < .0004$, paired-samples t-test, two-tailed), indicating that performance on the task improved with practice even though sessions were separated by several days. Importantly, the assignment of specific primes to the first or second session was counterbalanced across participants, so that the effect of session order was cancelled out in the computation of priming. For two trials in which the target happened to be a very low frequency word ("folly" and "perjury"), RT data was either missing (due to incorrect response) or had values greater than 5 standard deviations from the session mean for about half of the participants (17 and 14 participants, respectively). Therefore, these trials were not included in the priming RSA analyses. Including those trials did not change the RSA results. Split-half reliability of the magnitude of the priming effect was .52, which is high for semantic priming in lexical decision (Heyman et al., 2018). The difference in prime-target orthographic Levenshtein distance between the two sessions did not correlate with priming ($\rho = .00$), neither did differences in the other 23 lexical variables listed in **Supplemental Table 2**, except for concreteness ($\rho = .18$, uncorrected $p = .022$), ruling out the possibility that priming magnitude was driven by ortho-phonological factors.

	Session 1		Session 2		Overall	
	Word	Pseudoword	Word	Pseudoword	Word	Pseudoword
Acc	0.97 (.006)	0.95 (.008)	0.97 (.005)	0.96 (.006)	0.97 (.004)	0.96 (.005)
RT (ms)	631 (14)	722 (20)	598 (15)	662 (19)	614 (10)	692 (14)

Table 2. Group mean accuracy (Acc) and response time (RT) in the lexical decision task. Number in parenthesis is the standard error of the mean.

Semantic similarity rating. The median correlation between the ratings from each participant and the group median ratings was r = .92. For 20 of the 23 participants, correlations were between .86 and .96. The correlation was somewhat lower for one participant (.54). For two participants, correlations were exceptionally low (.11 and .26) indicating poor compliance with task instructions; these two participants were thus excluded from further analyses. Split-half correlation for the remaining 21 participants was .98, indicating excellent reliability.

Thematic association rating. The median correlation between the ratings from each participant and the group median ratings was r = .86. Four participants were excluded due to exceptionally low correlation with the group median ratings (r < .06) indicating poor compliance with task instructions. For the remaining 21 participants, correlations ranged from .39 to .93. Split-half correlation was .94, again showing excellent reliability.

As shown in **Figure 1, Supplemental Figure 1**, and **Supplemental Data 1**, all semantic representation models predicted priming significantly above chance, in subject-wise analyses (SW; all mean $\rho > .13$, all FDR-corrected $p < 10^{-7}$, one-sample Wilcoxon signed rank test, one-tailed) as well as in item-wise analyses (IW; all $\rho > .47$, all FDR-corrected $p < 10^{-10}$, permutation test) analyses. RSA correlations approached the lower-bound estimate of the noise ceiling (0.147) for all models, indicating that differences in performance between models may have been masked by a ceiling effect imposed by the level of noise present in the priming data. The same was true for the explicit ratings of subjective semantic similarity and thematic association. The Exp48 model was the only one whose RSA correlation numerically surpassed the lower-bound noise ceiling estimate or the predictions derived from the semantic similarity ratings, although there were no significant differences between models (SW: all FDR-corrected p > .9, Wilcoxon signed rank test, two-tailed; IW: all FDR-corrected p > .6, permutation test). In the item-wise analysis, Exp48 accounted for 27% of the variance in the data ($\rho = .52$). The overall pattern of model performances was similar across concrete and abstract targets (**Supplemental Figure 2**), although correlations for abstract targets had generally lower values and larger variances, possibly as a result of the smaller number of trials (53 abstract versus 157 concrete).

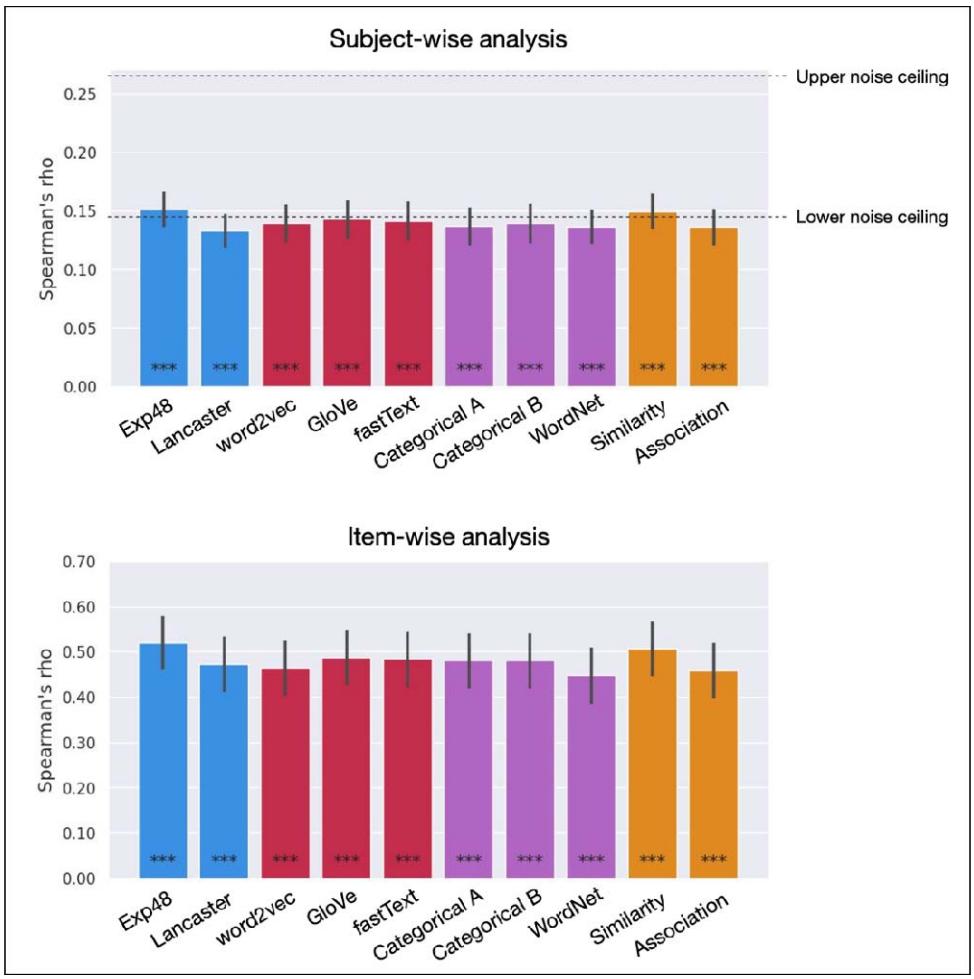

Figure 1. RSA correlations for experiential (blue), distributional (red), and taxonomic (purple) models of semantic representation and for explicit ratings of semantic similarity and thematic association (orange). *** FDR-corrected $p < 10^{-10}$.

Crucially, partial correlation analyses revealed that the experiential models explained unique variance in the priming data that could not be accounted for by taxonomic and distributional semantics models. Pairwise partial correlations evaluated the unique effect of the experiential models (blue bars in **Figure 2, top panel**) while controlling for each of the other models, one at a time, and, conversely, the unique effects of each taxonomic (purple) and distributional (red) model while controlling for each of the experiential models (see **Supplemental Data 2** and **3**). All partial correlations were highly significant for Exp48 (all FDR-corrected $p \leq .005$) while none of the taxonomic or distributional models correlated with priming when Exp48 was partialled out (all FDR-corrected $p > .14$). In other words, none of the other models explained significant variance beyond that explained by Exp48, while Exp48 explained significant unique variance not

explained by any other models. This is strong evidence that lexical concepts encode experiential information.

Interestingly, given its relatively small number of features, the Lancaster sensory-motor model also explained unique variance in priming after controlling for any of the distributional and taxonomic models (blue bars in **Figure 2, bottom panel**). This demonstrates that the sensory-motor structure of concepts can be detected in priming even when probed with a coarse set of sensory-motor features.

Separate analyses for concrete and abstract targets suggested an overall advantage for the experiential models with both kinds of concepts, although not all partial correlations for those models reached significance, likely due to insufficient power (**Supplemental Figure 3**, blue bars). When the experiential models were partialled out, there was a non-significant trend toward stronger partial correlations for taxonomic models (purple bars) relative to distributional models (red bars) with concrete targets, while abstract targets showed a trend in the opposite direction. This pattern is consistent with the fact that taxonomic structure is more clearly defined for concrete than for abstract concepts, and that distributional models capture other types of information (beside taxonomic relationships) that may be more important for abstract lexical concepts.

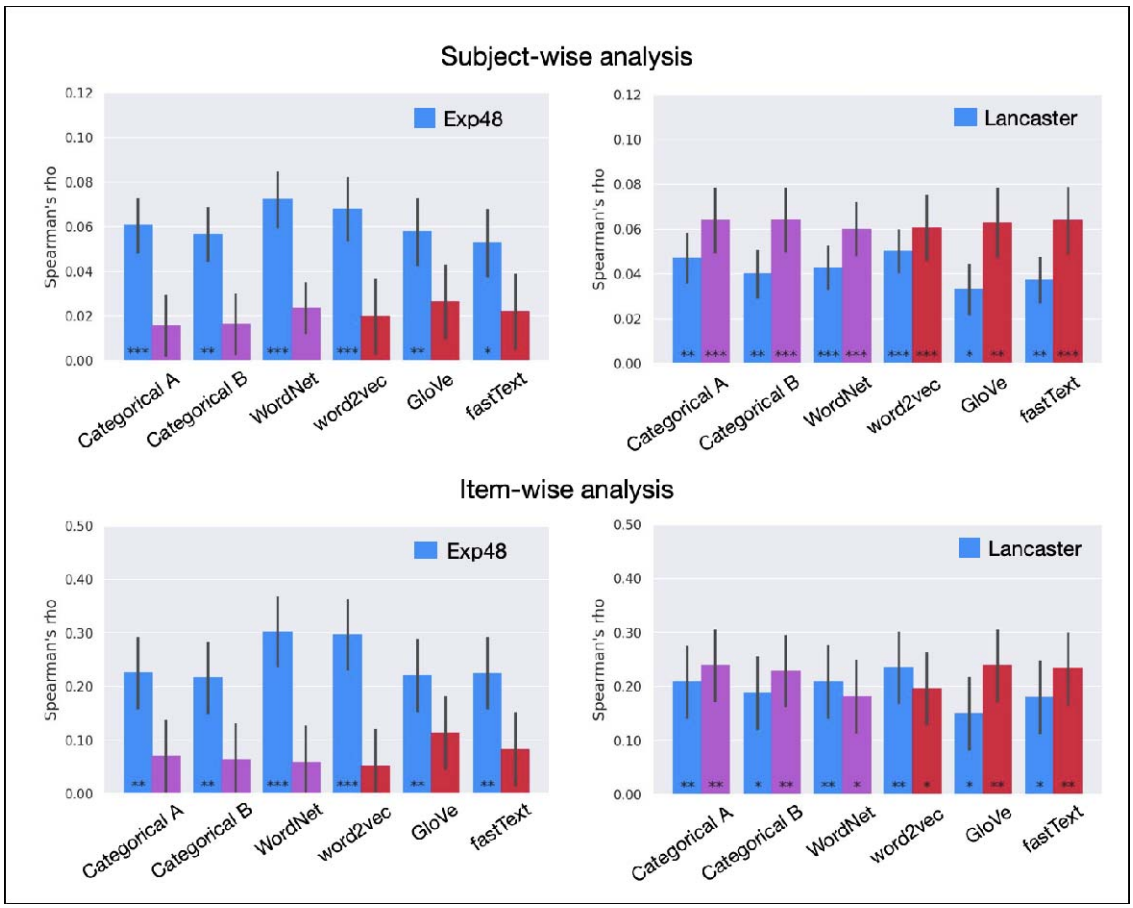

Figure 2. Pairwise partial correlations evaluating the amount of unique variance explained by the experiential models (blue) while controlling for each of the other models, and the amount of unique variance explained by the distributional (red) and taxonomic (purple) models while controlling for each experiential model. ***FDR-corrected p < .0005; **FDR-corrected p < .005; *FDR-corrected p < .01.

Exp48 predicted priming above chance level even after partialling out the variance explained by all taxonomic and distributional models simultaneously (SW: mean ρ = .038, FDR-corrected p = .025, Wilcoxon signed rank test, one-tailed; IW: ρ = .158, FDR-corrected p = .022, permutation test), again providing strong evidence that experiential information is independently encoded in lexical semantic representations and plays a central role in semantic word processing (**Supplemental Figure 5**). The Lancaster model predicted priming after simultaneously partialling out the variance explained by all 3 distributional models (SW: mean ρ = .0293, FDR-corrected p = .026, Wilcoxon signed rank test, one-tailed; IW: ρ = .147, FDR-corrected p = .034, permutation test) or by all 3 taxonomic models (SW: mean ρ = .029, FDR-corrected p = .012, Wilcoxon signed rank test, one-tailed; IW: ρ = .152, FDR-corrected p = .028,

permutation test), although correlations did not reach significance when all 6 models were partialled out simultaneously.

We then assessed whether each semantic model predicted priming above chance level after partialling out the variance explained by the explicit semantic similarity and thematic association ratings. Only Exp48 and Lancaster reached uncorrected significance after partialling out semantic similarity (SW: p = .014 and p = .027, respectively, Wilcoxon signed rank test; IW: p = .011 and p = .048, respectively, permutation test), although they did not survive correction for multiple tests (**Figure 3**, left panels). Partial correlations did not approach significance for any of the distributional models (all uncorrected p > .28). All models predicted priming after partialling out the thematic association ratings (all FDR-corrected p < .034; **Figure 3**, right panels), indicating that their performance in the main analysis was primarily driven by the effect of feature-based similarity on priming.

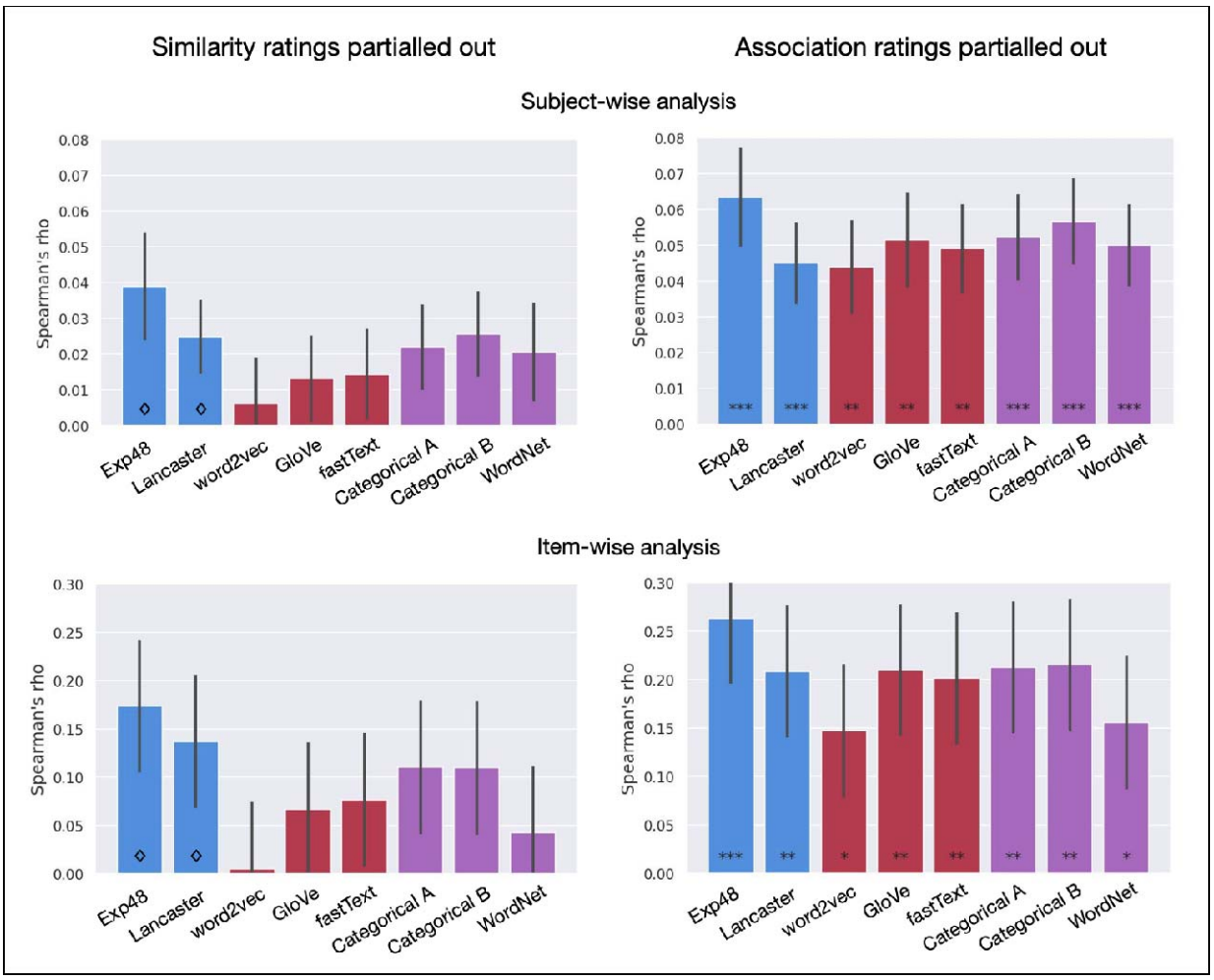

Figure 3. RSA correlations with the priming data after partialling out the similarity (left panels) or association (right panels) ratings. ◊ indicates uncorrected $p < .05$. ***FDR-corrected $p < .001$; **FDR-corrected $p < .005$; *FDR-corrected $p < .05$.

Finally, we evaluated the extent to which priming was driven by semantic similarity versus contextual association between prime and target. The partial correlation between semantic similarity and priming while controlling for contextual association was significant (SW: $\rho = 0.06$, $p = 0.0001$, Wilcoxon signed-rank test; IW: $\rho = 0.21$, $p = 0.003$, permutation test), but the partial correlation for contextual association while controlling for semantic similarity was not (SW: $\rho = -0.002$, $p = 0.78$, Wilcoxon signed-rank test; IW: $\rho = -0.021$, $p = 0.76$). This indicates that priming was driven by feature-based similarity, with little or no contribution of contextual association relationships.

Discussion

We set out to evaluate the extent to which a behavioral measure of semantic relatedness between words – semantic priming – reflects experiential and distributional similarity. While previous studies had investigated whether particular sensory-motor or affective features play a role in semantic word processing, the degree to which experiential information explains semantic behavior was still unknown. We used RSA to assess the combined effect of several experiential dimensions on the similarity structure of lexical concepts as revealed by semantic priming. We evaluated two experiential models, one based on coarse dimensions (Lancaster) and one based on both coarse and fine-grained dimensions (Exp48), relative to three of the top performing distributional models and three taxonomic models. The results showed a remarkably robust association between priming and experiential similarity, with both experiential models accounting for unique variance in the data after regressing out the variance explained by the distributional and taxonomic models (**Figure 2 and Supplemental Figure 5**) and the variance explained by semantic similarity and thematic association ratings (**Figure 3**). These results were found in subject-wise as well as in item-wise analyses, indicating that they are not specific to our stimulus set. Furthermore, they could not be explained by ortho-phonological factors (since none of those factors was correlated with priming) nor by grammatical class (since all the stimuli were nouns).

The overall trend toward Exp48 performing better than the Lancaster model, although not statistically significant, is consistent with the idea that lexical concepts encode information about

fine-grained experiential features (and consistent with the fMRI results of Fernandino et al., 2022). While experiential structure can still be detected at a coarse level (i.e., the relative importance of information originating from each of the major sensory-motor modalities, as indexed by the Lancaster model), information about the relative importance of more specific systems, such as those underlying the processing of color, shape, texture, motion, and reward value, among others, seems to play an important role in semantic word processing.

Much of the literature on embodied semantics has focused on whether modality-specific brain systems play a role in the representation of word meaning. However, it is important to note that seemingly elementary features of sensory-motor experience actually integrate information across multiple sensory-motor systems. The sense of space, for example, is a representational system integrating inputs from the visual, motor, tactile, proprioceptive, vestibular, and auditory systems, and is subserved by a network of multimodal neural structures such as the hippocampus, the entorhinal cortex, and the posterior parietal cortex (Baumann & Mattingley, 2014; Graziano et al., 1994; Graziano et al., 1997; Sanders et al., 2015). Graspability and manipulability also integrate information across visual, tactile, proprioceptive, and motor systems, and appear to rely most strongly on multimodal cortical areas such as the posterior middle temporal gyrus, the anterior supramarginal gyrus, the anterior intraparietal sulcus, and the ventral precentral sulcus (Jastorff et al., 2010; Peeters et al., 2009; Reynaud et al., 2019). The sense of causality, another ubiquitous aspect of human experience, is still poorly understood from a neurobiological perspective, but it is likely built upon experiential primitives that integrate sensory, motor, spatial, and temporal information into causal event schemas (Leshinskaya et al., 2021; Pelt et al., 2016; Pulvermüller, 2018; Rakison & Krogh, 2012). Therefore, while modality-specific effects on semantic language processing have provided evidence that sensory-motor systems contribute to semantic representation, many (if not most) sensory-motor features of experience combine information from multiple modalities.

Exp48 includes experiential features at several levels of coarseness and modality specificity. Coarse modality-specific features include Vision, Audition, and Touch; fine-grained modality-specific features include Color, Texture, High/Low Pitch, and Temperature. Multimodal features include Near (i.e., spatial proximity to the body), Manipulability, Caused, and Consequential. Shape, although listed as a visual feature in Binder et al. (2016), most likely results from the integration of visual, tactile, and motor information in neurologically typical individuals, because congenitally blind individuals appear to develop similar shape representations, particularly for manipulable artifacts (Bi et al., 2016; Kim et al., 2019; Peelen et al., 2014). Although the present

data does not allow us to specify the extent to which each of the features included in the Exp48 and Lancaster models affects semantic priming, it does suggest that at least some of the more fine-grained and multimodal features in Exp48 contribute to the phenomenon in a substantial way.

We also found that Exp48 accounted for all of the variance that was explained by any of the distributional and taxonomic models (**Figure 2**, left panels). In other words, although the non-experiential model predictions were correlated with priming magnitude when assessed in isolation (**Figure 1**), partial correlation analyses revealed that distributional and taxonomic similarity only predicted priming to the extent that they corresponded to experiential similarity. Therefore, we found no evidence for the hypothesis that distributional information contributes to the representation of lexical concepts. This finding, however, may be related to technical limitations of the particular distributional models tested; it remains possible that a future, more sophisticated distributional model could predict unique priming variance after accounting for the similarity structure predicted by Exp48. On the other hand, it is important to note that while the Exp48 and Lancaster models are based exclusively on theoretical principles, with no consideration of decoding or predictive performance, the architectures and parameters of the distributional models evaluated here are driven by performance optimization in word prediction via supervised learning (Mikolov et al., 2013, 2017; Pennington et al., 2014). It stands to reason that an experiential model optimized for semantic decoding (for example, via optimization of feature weights) would perform substantially more accurately than Exp48 and, potentially, more accurately than any possible distributional model.

It has been proposed that distributional information may play a more important role in the representation of abstract concepts than in the representation of concrete concepts . Although most target words in the present study (157 words) were relatively concrete (median concreteness = 4.9, IQR = 0.2), 51 target words were relatively abstract (median concreteness = 2.5, IQR = 1.4), such as “role”, “motive”, and “fate”. When trials with concrete and abstract targets were analyzed separately, there was no indication that distributional models performed better on the latter, whether in absolute terms or relative to the performance of the experiential models (**Supplemental Figure 3**). Partial correlations showed that neither Exp48 nor any of the distributional models explained significant unique variance for abstract targets, although the Lancaster model did after controlling for word2vec and fastText (subject-wise analyses only; **Supplemental Figure 4**). Of course, these results should be interpreted with caution, given the relatively low number of abstract trials.

Partialling out the predictions of the semantic similarity ratings or the contextual association ratings indicated that, for all models, RSA correlations with priming were primarily driven by semantic (i.e., feature-based) similarity (**Figure 3**). The two ratings were strongly correlated with each other (r = .90), but partial correlations showed that only one – semantic similarity – accounted for unique variance in the priming data that could not be explained by the other.

Previous studies have indicated that sensory-motor features of lexical concepts can be activated to a variable extent during semantic language processing, depending on the demands of the task (Dam et al., 2014; Goh et al., 2016; Raposo et al., 2009; Tousignant & Pexman, 2012). These findings raise the question of whether experiential features can be considered an essential aspect of lexical concepts rather than mere "post-semantic" representations recruited by task-related, strategic mechanisms (Weiskopf, 2010). Supporting the former interpretation, an event-related potentials (ERP) study by Muraki et al. (2019) suggests that experiential information is initially activated (within 200 ms of word onset) regardless of task-related factors and subsequently modulated by top-down processes according to task demands. The present study provides additional evidence for this view by showing that experiential features are quickly and reliably activated during lexical decision, a task that makes no explicit demands on semantic processing. While semantic content is implicitly activated and affects response latencies, lexical decision does not favor any particular kind of information associated with the word form, providing an unbiased assessment of the information content of lexical concepts. Furthermore, prime and target words were presented in a backward-masked, short SOA procedure that emphasizes the effect of semantic features that are activated within 200 ms of the prime onset. Therefore, our results indicate that experiential features are automatically activated during semantic word processing.

The idea that lexical concepts possess a "core" set of features that are automatically activated regardless of the context has been called into question by Lebois, Wilson-Mendehall, and Barsalou (2014). They reviewed several studies demonstrating that a variety of features considered central to the meanings of words can be modulated by contextual factors such as task demands and the composition of the stimulus set. The authors concluded that semantic feature activation is always dynamic and context-dependent, and that distinctions between core and peripheral features are therefore meaningless. Although we agree with Lebois et al.'s rejection of the idea that "context-independent automaticity is the gold standard of semantic processing" with "dynamic context-dependent features being relatively irrelevant", we disagree with their dichotomous characterization of automaticity and context-dependency. The reason is

that any phenomenon that is typically considered automatic is also susceptible to modulation by contextual factors. Consider, for example, the pattelar reflex: the leg extension is an automatic response to the mechanical stimulus delivered to the patellar tendon, yet the latency and magnitude of the response can be modulated by a variety of contextual factors. To say that a particular semantic feature is automatically activated by a word form means that, under normal conditions, activation of the orthographic or phonological word form representation is a sufficient condition for the activation of that feature; it does not mean that the degree to which the feature is activated cannot be modulated by contextual factors. Furthermore, different semantic features are not all equally susceptible to task-related factors. It is clear that, among the semantic features that can be associated with a particular word form, there are some that appear to be more consistently activated across different contexts (and across speakers of the language) than others. Hence, we do see a meaningful distinction between features that show relatively little sensitivity to contextual factors and those that are strongly context-dependent. Because the activation of semantic representations in our paradigm – particularly those associated with the prime – is completely incidental to the task and occurs at a relatively short time scale, our results suggest that experiential features are fundamental constituents of word meaning.

The results reported here are consistent with a hierarchical model of word meaning in which lexical concepts consist of fuzzy sets of experiential features encoded at various levels of representation throughout the cortex (Fernandino et al., 2016a). At the bottom of the hierarchy, primary sensory-motor, interoceptive, and limbic/paralimbic areas encode simple modality-specific features. At the next level, specialized networks encode more elaborate features (e.g., 3D shape, spatial relationships, manipulability, edibility, sequential structure, causal relationships, etc.) based on information extracted from lower-level areas and processed according to innate connectivity constraints and ecological/pragmatic needs. At the top level, a set of heteromodal cortical hubs integrate information across those networks to encode complex, flexible assemblages of experiential features that can become associated with an arbitrary word form (i.e., a phonological/orthographic representation). During semantic word processing, the word form representation initially activates a set of high-level experiential features stored in the heteromodal hubs, which, in turn, activate lower-level features stored in other multimodal areas in a context-dependent manner. These lower-level features can, in turn, activate simpler features stored in early sensory-motor and affective areas depending on task demands.

We identify these heteromodal hubs with the angular gyrus, the anterior lateral temporal cortex, the parahippocampal cortex, the precuneus/posterior cingulate cortex, and the medial prefrontal cortex. Collectively known as the “default mode network” (DMN), these areas have been shown to be the ones that are most distantly connected to primary sensory and motor areas, and they appear to integrate information across all modalities (Fernandino et al., 2016a; Fernandino et al., 2022; Margulies et al., 2016; Sepulcre et al., 2012; Tong et al., 2022). They are also highly interconnected with each other (Buckner et al., 2008) and have been shown to be jointly activated during semantic language tasks (Binder et al., 1999, 2009; Wirth et al., 2011) and during other tasks that depend primarily on information retrieved from long-term memory, such as remembering the past, thinking about the future, and daydreaming (Christoff et al., 2009; Mason et al., 2007; McKiernan et al., 2003; Spreng & Grady, 2010). We had previously shown that individual words could be decoded from fMRI activation patterns in these areas using a multimodal sensory-motor model, but not with a model based on ortho-phonological features of the corresponding word forms (Fernandino, 2016b). Furthermore, the similarity structure of fMRI activation patterns in these regions predicts the semantic similarity structure of both object and event nouns, and it does so significantly more accurately when semantic similarity is estimated based on experiential features than when it is based on taxonomic or distributional information (Fernandino et al., 2022; Tong et al., 2022).

In agreement with this model, the present study provides robust behavioral evidence that multimodal experiential information is automatically activated during semantic word processing. Further research is required to determine in more detail how different experiential features contribute to the representation of lexical concepts and how their activation is modulated by contextual factors. Future studies should also investigate how combinations of features map onto the heteromodal cortical hubs, as well as the activation time course of the hubs during word comprehension.

Acknowledgements

The authors would like to thank Jeffrey Binder for helpful discussions and invaluable support, and Taylor Kalmus for help with data collection. This work was supported by an NIDCD grant R01DC016622.

References

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. *Cognitive Science*, *45*(10), e13055. https://doi.org/10.1111/cogs.13055

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577-609-discussion 610-60. https://doi.org/10.1017/s0140525x99002149

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and Simulation in Conceptual Processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). https://doi.org/https://doi.org/10.1093/acprof:oso/9780199217274.003.0013

Barsalou, L. W., Simmons, W. K., Barbey, A., & Wilson, C. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.

Baumann, O., & Mattingley, J. B. (2014). Dissociable roles of the hippocampus and parietal cortex in processing of coordinate and categorical spatial information. *Frontiers in Human Neuroscience*, *8*, 1–6. https://doi.org/10.3389/fnhum.2014.00072/abstract

Bedny, M. (2020). The contribution of sensorimotor experience to the mind and brain. In *The Cognitive Neurosciences VI* (6th ed.). MIT Press.

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: the case of action verbs. *The Journal of Neuroscience*, *28*(44), 11347–11353. https://doi.org/10.1523/jneurosci.3039-08.2008

Bi, Y., Wang, X., & Caramazza, A. (2016). Object Domain and Modality in the Ventral Visual Pathway. *Trends in Cognitive Sciences*, *20*(4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002

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

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience*, *11*(1), 80–93. https://doi.org/10.1162/089892999563265

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38. https://doi.org/10.1196/annals.1440.011

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (1st ed., pp. 117–156). Lawrence Erlbaum Associates.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*(8), 354–361. https://doi.org/10.1016/s1364-6613(03)00159-1

Chomsky, N. (1975). *Reflections on Language*. Pantheon Books.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*(21), 8719–8724. https://doi.org/10.1073/pnas.0900234106

Connell, L., Brand, J., Carney, J., Brysbaert, M., & Lynott, D. (2019). Go big and go grounded: Categorical structure emerges spontaneously from the latent structure of sensorimotor experience. In G. A, S. C, & F. C (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (p. 3434).

Dam, W. O. van, Brazil, I. A., Bekkering, H., & Rueschemeyer, S.-A. (2014). Flexibility in Embodied Language Processing: Context Effects in Lexical Access. *Topics in Cognitive Science*, *6*(3), 407–424. https://doi.org/10.1111/tops.12100

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-x

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

Fernandino, L., Humphries, C. J., Conant, L. L., Seidenberg, M. S., & Binder, J. R. (2016). Heteromodal cortical areas encode sensory-motor features of word meaning. *The Journal of Neuroscience*, *36*(38), 9763–9769. https://doi.org/10.1523/jneurosci.4095-15.2016

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

Fodor, J. A. (1980). *The Language of Thought*. Harvard University Press. https://doi.org/10.1093/0198236360.003.0001

Glenberg, A. (1997). What memory is for. *Behavioral And Brain Sciences*, *20*, 1–55. http://journals.cambridge.org/abstract_S0140525X97000010

Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, *43*(3), 379–401. https://doi.org/10.1006/jmla.2000.2714

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L.-C. (2016). Semantic Richness Effects in Spoken Word Recognition: A Lexical Decision and Semantic Categorization Megastudy. *Frontiers in Psychology*, *7*(723), 802–810. https://doi.org/10.3389/fpsyg.2016.00976

Goldberg, Y., & Levy, O. (2014). word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv.Org*, *cs.CL*. arXiv.org

Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Coding the Locations of Objects in the Dark. *Science*, *277*(5323), 239–241. https://doi.org/10.1126/science.277.5323.239

Graziano, M., Yap, G., & Gross, C. (1994). Coding of visual space by premotor neurons. *Science*, *266*(5187), 1054–1057. https://doi.org/10.1126/science.7973661

Günther, F., Dudschig, C., & Kaup, B. (2016a). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology*, *69*(4), 626–653. https://doi.org/10.1080/17470218.2015.1038280

Günther, F., Dudschig, C., & Kaup, B. (2016b). Predicting Lexical Priming Effects from Distributional Semantic Similarities: A Replication with Extension. *Frontiers in Psychology*, *7*, 1646. https://doi.org/10.3389/fpsyg.2016.01646

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1–3), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harris, Z. S. (1954). Distributional Structure. *Word*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hauk, O., & Tschentscher, N. (2013). The Body of Evidence: What Can Neuroscience Tell Us about Embodied Semantics? *Frontiers in Psychology*, *4*, 50. https://doi.org/10.3389/fpsyg.2013.00050

Heyman, T., Bruninx, A., Hutchison, K. A., & Storms, G. (2018). The (un)reliability of item-level semantic priming effects. *Behavior Research Methods*, *50*(6), 2173–2183. https://doi.org/10.3758/s13428-018-1040-9

Heyman, T., Hutchison, K. A., & Storms, G. (2016). Uncovering underlying processes of semantic priming by correlating item-level effects. *Psychonomic Bulletin & Review*, *23*(2), 540–547. https://doi.org/10.3758/s13423-015-0932-2

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/bf03196544

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *The Quarterly Journal of Experimental Psychology*, *61*(7), 1036–1066. https://doi.org/10.1080/17470210701438111

Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., & Orban, G. A. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *Journal of Neurophysiology*, *104*(1), 128–140. https://doi.org/10.1152/jn.00254.2010

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003

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

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. *Proceedings of the National Academy of Sciences*, *116*(23), 11213–11222. https://doi.org/10.1073/pnas.1900952116

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 211–240. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.184.4759&rep=rep1&type=pdf

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2014). Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. *Cognitive Science*. https://doi.org/10.1111/cogs.12174

Leshinskaya, A., Bajaj, M., & Thompson-Schill, S. L. (2021). Novel objects with causal event schemas elicit selective responses in tool- and hand-selective lateral occipitotemporal cortex. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac442

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*(4), 838–844. https://doi.org/10.3758/pbr.15.4.838

Louwerse, M. M. (2011). Symbol Interdependency in Symbolic and Embodied Cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2019). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *26*(3), 6288. https://doi.org/10.3758/s13428-019-01316-z

Mahon, B. Z. (2014). What is embodied about cognition? *Language, Cognition and Neuroscience*, *ahead-of-print*, 1–10. https://doi.org/10.1080/23273798.2014.987791

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris*, *102*(1–3), 59–70. https://doi.org/10.1016/j.jphysparis.2008.03.004

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12574–12579. https://doi.org/10.1073/pnas.1608282113

Mason, M. F., Norton, M. I., Horn, J. D. V., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science*, *315*(5810), 393–395. https://doi.org/10.1126/science.1131295

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A Parametric Manipulation of Factors Affecting Task-induced Deactivation in Functional Neuroimaging. *Journal of Cognitive Neuroscience*, *15*(3), 394–408. https://doi.org/10.1162/089892903321593117

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

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv.Org*, *cs.CL*.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in Pre-Training Distributed Word Representations. *arXiv*. https://doi.org/10.48550/arxiv.1712.09405

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, *3*(4), 235–244. https://doi.org/10.1093/ijl/3.4.235

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Harvard University Press.

Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2019). Mapping semantic space: property norms and semantic richness. *Cognitive Processing*, *44*(3), 1028–13. https://doi.org/10.1007/s10339-019-00933-y

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*(2), 135–183. https://doi.org/10.1016/s0364-0213(80)80015-2

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. *The Journal of Neuroscience*, *34*(1), 163–170. https://doi.org/10.1523/jneurosci.1114-13.2014

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *The Journal of Neuroscience*, *29*(37), 11523–11539. https://doi.org/10.1523/jneurosci.2040-09.2009

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pelt, S. van, Heil, L., Kwisthout, J., Ondobaka, S., Rooij, I. van, & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience*, *11*(6), 973–980. https://doi.org/10.1093/scan/nsw017

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Empirical Methods in Natural Language Processing*, 1532–1543.

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*(3–4), 175–190. https://doi.org/10.1080/02643294.2016.1176907

Pulvermuller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–279. https://doi.org/10.1017/s0140525x9900182x

Pulvermüller, F. (2018). The case of CAUSE: neurobiological mechanisms for grounding an abstract concept. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170129. https://doi.org/10.1098/rstb.2017.0129

Pylyshyn, Z. (1984). *Computation and cognition: toward a foundation for cognitive science*. MIT Press.

Rakison, D. H., & Krogh, L. (2012). Does causal action facilitate causal perception in infants younger than 6 months of age? *Developmental Science*, *15*(1), 43–53. https://doi.org/10.1111/j.1467-7687.2011.01096.x

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia*, *47*(2), 388–396. https://doi.org/10.1016/j.neuropsychologia.2008.09.017

Reynaud, E., Navarro, J., Lesourd, M., & Osiurak, F. (2019). To Watch is to Work: a Review of NeuroImaging Data on Tool Use Observation Network. *Neuropsychology Review*, *29*(4), 484–497. https://doi.org/10.1007/s11065-019-09418-3

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition*. MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Sahlgren, M. (2008). The distributional hypothesis. *Rivista Di Linguistica*, *20*(1), 33–53.

Sanders, H., Rennó-Costa, C., Idiart, M., & Lisman, J. (2015). Grid Cells and Place Cells: An Integrated View of their Navigational and Memory Function. *Trends in Neurosciences*, 1–13. https://doi.org/10.1016/j.tins.2015.10.004

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/s0140525x00005756

Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *The Journal of Neuroscience*, *32*(31), 10649–10661. https://doi.org/10.1523/jneurosci.0759-12.2012

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145–159. https://doi.org/10.1037/h0030806

Skinner, B. F. (1957). *Verbal behavior*. Copley Publishing Group. https://doi.org/10.1037/11256-000

Spreng, R. N., & Grady, C. L. (2010). Patterns of Brain Activity Supporting Autobiographical Memory, Prospection, and Theory of Mind, and Their Relationship to the Default Mode Network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of Semantic and Associative Relatedness on Automatic Priming. *Journal of Memory and Language*, *38*(4), 440–458. https://doi.org/10.1006/jmla.1997.2559

Tong, J.-Q., Binder, J. R., Humphries, C. J., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A distributed network for multimodal experiential representation of concepts. *The Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/jneurosci.1243-21.2022

Tousignant, C., & Pexman, P. M. (2012). Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. *Frontiers in Human Neuroscience*, *6*, 53. https://doi.org/10.3389/fnhum.2012.00053

Weiskopf, D. A. (2010). Embodied cognition and linguistic comprehension. *Studies in History and Philosophy of Science Part A*, *41*(3), 294–304. https://doi.org/10.1016/j.shpsa.2010.07.005

Wirth, M., Jann, K., Dierks, T., Federspiel, A., Wiest, R., & Horn, H. (2011). Semantic memory involvement in the default mode network: A functional neuroimaging study using independent component analysis. *NeuroImage*, *54*(4), 3057–3066. https://doi.org/10.1016/j.neuroimage.2010.10.039

Zubicaray, G. de, Arciuli, J., & McMahon, K. (2013). Putting an “End” to the Motor Cortex Representations of Action Words. *Journal of Cognitive Neuroscience*, *25*(11), 1957–1974. https://doi.org/10.1162/jocn_a_00437