Individual Differences in the Neural Localization of Relational Networks of Semantic Concepts

Katherine L. Alfred, Megan E. Hillis, and David J. M. Kraemer

Abstract

■ Semantic concepts relate to each other to varying degrees to form a network of zero-order relations, and these zero-order relations serve as input into networks of general relation types as well as higher order relations. Previous work has studied the neural mapping of semantic concepts across domains, although much work remains to be done to understand how the localization and structure of those architectures differ depending on various individual differences in attentional bias toward different content presentation formats. Using an item-wise model of semantic distance of zero-order relations (Word2vec) between stimuli (presented both in word and picture forms), we used representational similarity analysis to identify individual differences in the neural localization of semantic concepts and how those localization differences can be predicted by individual variance in the degree to which individuals attend to word information instead of pictures. Importantly, there were no reliable representations of this zero-order semantic relational network when looking at the full group, and it was only through considering individual differences that a stable localization difference became evident. These results indicate that individual differences in the degree to which a person habitually attends to word information instead of picture information substantially affects the neural localization of zero-order semantic representations. ■

INTRODUCTION

Relational reasoning, the ability to identify and apply patterns of relationships present in a situation to another similar situation, is a critical component of abstract thought. Much of the research on relational reasoning examines second-order relations (see Krawczyk, 2012, for a review), which refer to the way that first-order relationships between items relate to first-order relationships between other items, such as part–whole relationships (e.g., door–house) or similarity between category comembers (e.g., lemon–lime). A second-order relationship would be the relationship of the relationships between two pairs of words, such as saying that toe–foot and finger–hand are both part–whole relationships and are analogous. The first-order relationship is the specific relationship between items, such a toe–foot. Finally, at the most basic level, zero-order relationships are the way that all items relate semantically to each other when there is no explicit consideration of that relationship. Ultimately, first-order relationships are rooted in zero-order relationships, just as each subsequently higher level of relationships is rooted in the level below. Therefore, a better understanding of the neural basis of zero-order relationships (i.e., semantic similarity) will also inform an understanding of the representational structure underlying first-order and higher relationships.

Second-order reasoning is a critical feature of relational reasoning because the ability to observe patterns of relations among one set of items and compare those patterns of relationships to those observed between other sets of items provides the basis for transfer of skills from one problem to another, insight on how to approach novel situations, and abstract thought as a whole (Hummel & Holyoak, 1997). Second-order reasoning is theorized to function through the representation of conceptual networks throughout posterior sensory-based regions of the brain (Binder, Desai, Graves, & Conant, 2009; Martin, 2007; Patterson, Nestor, & Rogers, 2007; Thompson-Schill, 2003), and domain-general relational reasoning networks, including inferior parietal lobe and rostrolateral pFC, receive input during semantic retrieval and analogical reasoning and identify the second-order relations (e.g., analogies) when similar zero-order relations are co-active (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010; Green, Fugelsang, Kraemer, & Dunbar, 2008; Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2008; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; Bunge, Wendelken, Badre, & Wagner, 2005).

Here, we focus on the nature of the zero-order representations that serve as input into the relational reasoning process, including where these zero-order relational networks are localized in the brain and how that localization differs between individuals. The question of how concepts and

This article is part of a Special Focus, Relational Reasoning, deriving from a symposium at the 2019 annual meeting of the Cognitive Neuroscience Society, organized by Silvia Bunge and Keith Holyoak.
Dartmouth College

Journal of Cognitive Neuroscience 33:3, pp. 390–401
https://doi.org/10.1162/jocn_a_01657

the zero-order relationships between them serve as input into second-order relationships is still the subject of investigation. One possibility is that relational reasoning relies on directly activating the nodes and connections in the semantic network each time a second-order relationship is perceived. However, a recent study indicates that the brain may implement a more efficient solution, in which a network of general relation types (e.g., similar, contrast, part–whole) exists independently of the conceptual network and serves as the direct input to second-order reasoning (Chiang, Peng, Lu, Holyoak, & Monti, 2020). However, as the focus of relational reasoning studies has typically been on second-order relations or general relation types, these studies tend to use a relatively constrained set of potential relationships between items. To better understand the semantic networks that form the foundation of relational reasoning, as well as how zero-order reasoning inputs into higher-order reasoning, more work needs to be done to investigate how the foundational zero-order relations are represented, and how the neural instantiation of those representations can vary between individuals. Because concepts relate to each other in relational similarity space that reflects semantic distance (Mikolov, Chen, Corrado, & Dean, 2013), we can use representational similarity analysis (RSA; Kriegeskorte et al., 2008) to localize regions of the brain where the neural patterns of activity in response to these concepts are similar to that predicted semantic relational structure.

This study provides insight into the fine-grained patterns of semantic similarity and how representations of meaningful words and pictures can reliably and systematically vary between individuals. This analysis leverages the Word2vec model of semantic similarity (Mikolov et al., 2013) on the level of individual items. This process results in a searchlight RSA model that will localize patterns of neural activity where items like "wrench" and "screw" are represented similarly, and differently from items like "bicycle" and "fox." Importantly, whereas the Word2vec model captures fine-grained differences in the similarity of different items, the Word2vec model used in the searchlight RSA is agnostic to whether the item was originally presented to the participant as a word or a picture. By using this pure semantic model, similarities between items will not be confounded by the format of item presentation, and we will be better able to determine that any material-specific processing is because of individual differences in the format of conceptual representation. We aim to use this analysis to put a finer point on the results from our previous study (Alfred, Hayes, Pizzie, Cetron, & Kraemer, 2020), in which we identified stable patterns of individual variation to localized regions that are common between participants and which are the result of individual differences in representation because of differences in habitual attention bias to word versus picture representations of concepts.

For both our prior study (Alfred et al., 2020) and the current study, we use a measure of attentional bias as our primary measure of interest for individual variation. The way a stimulus is encoded is shaped largely by which of its features the viewer is attending. For example, when looking at directions to a new place, one person may remember the step-by-step directions, whereas another person may create a mental map of the area. The theory of material-specific encoding states that the format of a presented stimulus (e.g., the word "cat" vs. a picture of a cat) affects which areas of the brain are recruited during encoding and retrieval (Grady, McIntosh, Rajah, & Craik, 1998; Wagner et al., 1998). Earlier research has shown that there is a divide between cognition, which is primarily linguistically mediated (verbal cognition) and that which primarily operates on visuospatial representations (visual cognition), and that these are localized to different areas of the brain. For example, Milner, Corsi, and Leonard (1991) found that lesions in the left medial temporal lobe interfere with verbal memory, whereas lesions in the right temporal lobe interfere with memory for nonverbal material. Further research has concluded that language processing is associated with the left hemisphere, whereas visuospatial processing is associated with the right hemisphere (Golby et al., 2001; Gross, 1972; Milner, 1971). However, mental representations of almost any object encompass a large set of properties including its appearance; the word for the item; how it sounds, feels, or smells; and the context in which it is encountered (Martin, 2007; Lacey & Campbell, 2006; Malt, 1990). An individual's perception of the item depends on which of these properties they preferentially attend to. As a result, different individuals may encode verbal and visual stimuli differently relative to one another on the basis of habitual attention to word or picture representations regardless of the original format of presentation. Building on the results of our previous study, we predict that people who show an attentional bias toward words will encode stimuli in language-specific regions, such as the speech production network, despite the fact that half of the stimuli were only presented in picture form.

METHODS

This study is a novel analysis of the data set originally used in Alfred et al. (2020) and Hayes, Alfred, Pizzie, Cetron, and Kraemer (2020). The previous two studies focused on the similarity in the neural representation of studied word and picture stimuli (compared to abstract pictures and pseudowords) and differences in white and gray matter structure associated with verbal attentional bias. In this study, we aim to use the same data set to examine the neural representation of zero-order semantic relations. Compared to the study of Alfred et al. (2020), which utilized a basic level dissimilarity matrix (DM) to localize regions of the cortex that represented meaningful words and pictures similarly to each other and differently from pseudowords and abstract pictures, this study leverages a rich dissimilarity structure calculated with Word2vec to identify a specific zero-order relational structure.

Participants

Twenty-eight (16 female, M_{age} = 20.7 years) undergraduate and graduate students at Dartmouth College participated in this study. All participants were right-handed native English speakers with normal or corrected-to-normal vision, with no history of neurological or psychiatric disorders. All participants provided informed written consent and were compensated with a choice of cash or course credit for their participation. This study has been approved by the Dartmouth Committee for the Protection of Human Subjects. The data from two participants had significant problems going through the analysis pipeline (described below) and could not be analyzed, leaving data from a total of 26 participants for analysis.

Verbal Attentional Bias Task

We designed a novel behavioral task to measure the degree to which participants attended to visual and verbal information. This task was administered as part of a battery of measures of visual and verbal cognitive ability, as well as the Verbalizer-Visualizer Questionnaire measure of cognitive style. In each trial, a card suit symbol was displayed with an accompanying text label, and participants were asked to press a key to identify whether they were being shown club, spade, or heart (Figure 1A). The experimenters gave the participants the specific instructions to "Please respond as quickly and as accurately as you can. Use three fingers on J, K, L pressed by the index, middle, and ring fingers respectively. 'J' corresponds to club. 'K' corresponds to heart. 'L' corresponds to spade (in alphabetical order). Open the door and get me when the experiment is complete."

The experiment consisted of 192 trials, where 144 (75%) presented congruent information—that is, the text labels matched the symbols shown. In 48 (25%) of the trials, however, the picture and the text label had conflicting information (e.g., a picture of a club with text that says "spade"). Participants were not informed that there may be any incongruency between the picture and word presented, and all practice trials with the experimenter present were congruent. The experimenter then left the room, so when encountering incongruent trials, the participant would need to determine which piece of information was most salient. A measure called "word attentional bias" was calculated as the percentage of incongruent trials for which the participant pressed the key corresponding to the verbal label (Figure 1B). Each of the three suits was the target in an equal number of times, and the location of the text was counterbalanced for presentation above and below the picture. The center of the screen was between the picture and the text. This task was completed in the same session as the other behavioral measures, within a week after the initial fMRI session.

Word and Picture Intentional Encoding Task (fMRI Task)

During the fMRI session, participants were presented with sequences of items to study for a later test. The items were

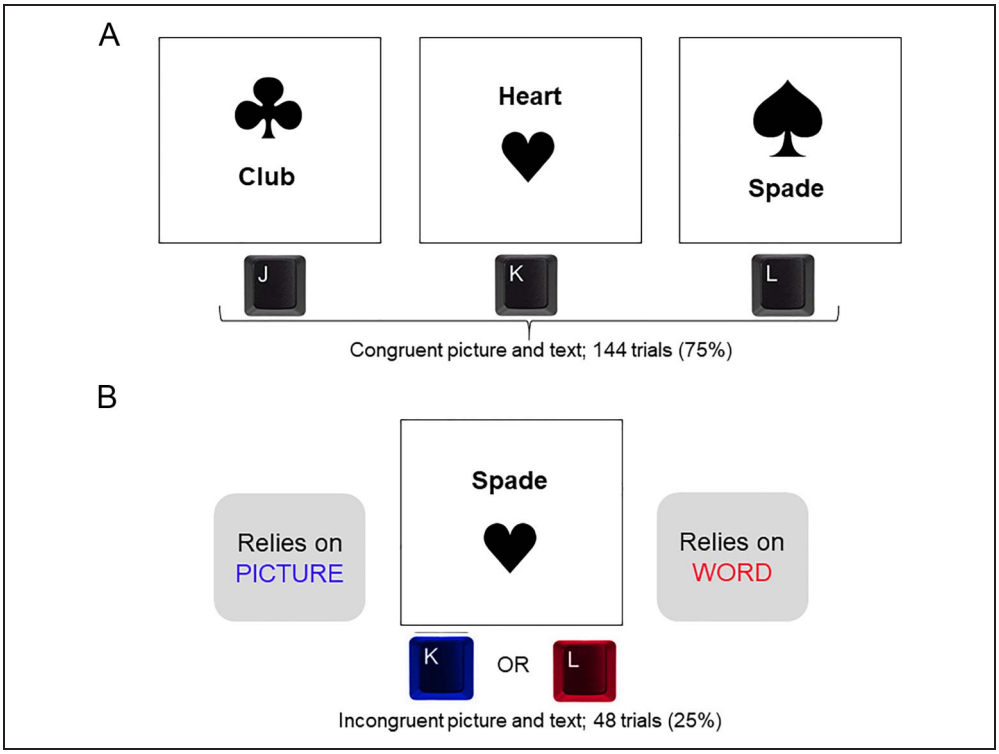

Figure 1. Structure of attentional bias task. (A) Participants were instructed to press J for club, K for heart, and L for spade, and to respond as quickly and accurately as possible. Most trials (75%) presented congruent word and picture information. (B) Some trials (25%) unexpectedly presented a word and picture that were incongruent. Participants had to rapidly decide to select the key corresponding to the picture (in this case, responding K for heart) or word (in this case, responding L for spade). Figure used with permission from Alfred et al. (2020).

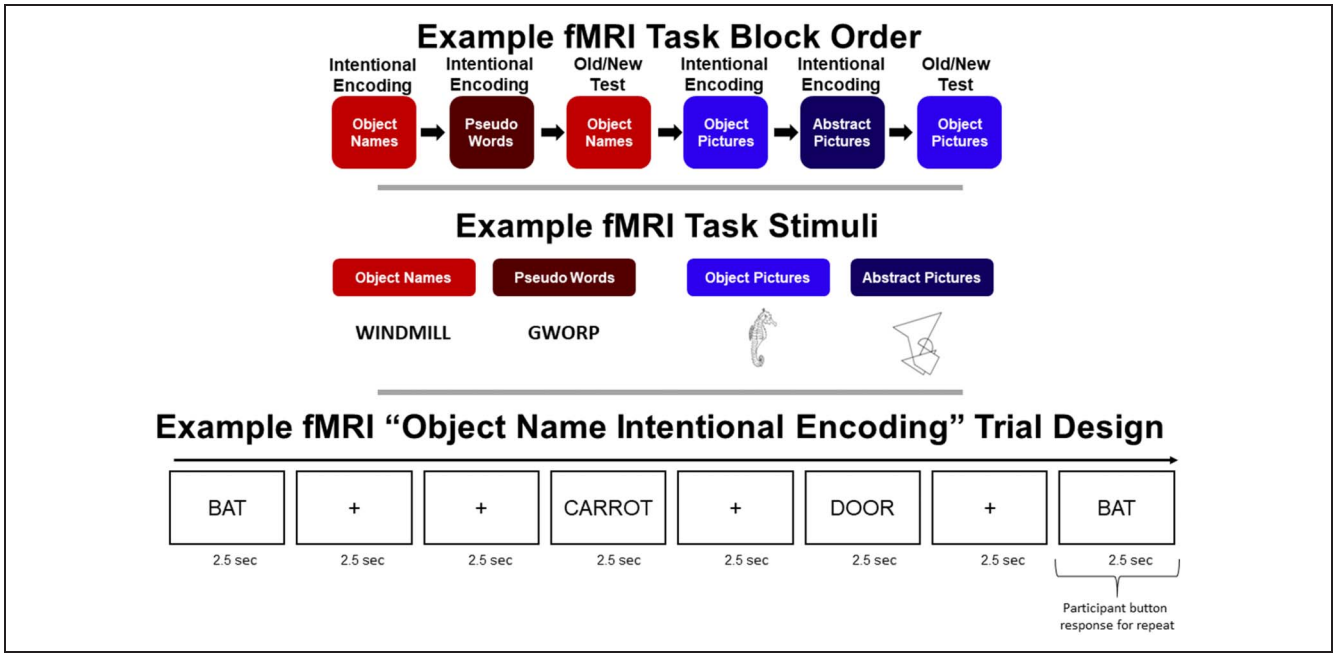

Figure 2. Overview of fMRI session and stimuli. Participants were instructed to study a block of items with object names, such as "windmill," followed by a block of pseudowords, such as "gworp." After studying pseudowords, participants were tested on the object names they had studied earlier. This procedure was repeated with object pictures and abstract pictures, with a test on the object pictures. Whether participants saw the object name + pseudoword block first or the object picture + abstract picture block first was counterbalanced between participants. Figure used with permission from Alfred et al. (2020).

presented in lists that were blocked by content type—object names, pseudowords, object pictures, and abstract pictures—designed to measure neural activity during intentional encoding (Figure 2). Participants were instructed to pay attention to the stimuli for a later test, specifically being told, "In this section you will see a series of words. Pay attention to each word and try to remember it—your memory for these words will be tested later. You should also pay attention to when an item has been presented more than once. When you see a word appear for the second time, press the button with your right index finger. Otherwise do not press a button." The same instructions were used for the object name and pseudoword conditions. In conditions with object pictures and abstract pictures, the instructions remained the same except that "word" was changed to "picture" to match the stimulus type being presented. Participants completed 4 study blocks and two test blocks, and were only tested on object names and object pictures.

During study blocks, while participants studied the items for the test, they simultaneously watched for repeated items, and made a button response during the study blocks only if they identified a repeated item. After memorizing a list of English object names and a list of pronounceable pseudowords, participants took a test on the object names they had studied. Tests contained 120 trials, and participants indicated if an item was studied or new, as well as if they had a "high" or "low" level of confidence, or if they were making a "guess." Half of the trials contained one of the 60 items studied in the word block, and 60 items were new. Whereas the participants took the test in the scanner, anatomical and diffusor tensor imaging sequences were collected instead of functional scans. After the object name test, participants studied a set of object pictures and a set of abstract pictures and then took a test on the object pictures. Participants were not given tests on either the pseudowords or abstract pictures. Each block contained 60 items that would appear on the test for that block (presented 2.5 sec each), six repeat items that were shown twice (total of 12 presentations, 2.5 sec each), and variable-duration fixation crosses (72 fixation periods, 2.5 sec each, with up to three fixation periods in a row) all interleaved together. In the object picture block, the critical items were readily nameable black line drawings from the Snodgrass item set (Snodgrass & Vanderwart, 1980). In the object name block, the words used were the names of pictures from the object picture block, although no participant saw the set of object names that labeled the object pictures presented to them. Abstract pictures were black line drawings with both straight and curved lines, but did not resemble any object. Pseudowords were drawn from the Deacon, Dynowska, Ritter, and Grose-Fifer (2004) set of pronounceable nonwords that did not have English roots. Repeat items to check for continued attention were present during all four blocks and were composed of the same type of stimuli but were not used in the test. Although it is possible participants could have become aware that they were not going to be tested on abstract picture/pseudoword blocks after not being tested on the presentation of the first of those two blocks, this did not affect participant engagement in the encoding task, as evidenced by performance on the repeat detection task.

Regardless of which of the two abstract blocks participants were exposed to first, participants had nearly identical signal detection rates for repeated items, indicating engagement in the encoding process, especially given most of the repeated trials had a large number of intervening items between the initial presentation and the repeat. Participants showed virtually identical signal detection rates regardless of if the object name + pseudoword block was presented before or after the object picture + abstract picture block ($d'_{\text{first-pres.}}$: $M = 2.48$, $SD = 0.78$; $d'_{\text{second-pres.}}$: $M = 2.56$, $SD = 0.83$), $t(26.5) = -0.24$, $p = .81$; and similarly showed no difference in signal detection for abstract pictures ($d'_{\textit{first-pres.}}$: $M = 1.47$, $SD = 0.90$; $d'_{\text{second-pres.}}$: $M = 1.55$, $SD = 0.64$), $t(25.3) = -0.28$, $p = .79$. These data were not used for any further individual differences analyses because of the limited number of repeat trials per participant. As mentioned above, with the exception of attention check repeat trials, none of the items were repeated between conditions (i.e., a word studied in one block would not be the name of a picture studied in a later block), and the lists were counterbalanced between participants. The fMRI session lasted approximately 2.5 hr (Figure 2).

MRI Scanner Information

Scans took place at the Dartmouth Brain Imaging Center with a Phillips 3T Achieva Intera using a 32-channel sense head coil. For the functional runs, there were four runs of 150 volumes per run for 600 functional (T2*-weighted) volumes with a repetition time of 2.5 sec. The functional scans were a gradient-echo EPI with 42 transverse slices at 3 mm per slice. Echo time was 35, and flip angle was 90°. The scan image acquisition order used was Philips interleaved.

Univariate Functional Imaging Analysis

Neural data were preprocessed for analysis using the FSL tools for motion correction and registration (MCFLIRT: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MCFLIRT; Jenkinson, Bannister, Brady, & Smith, 2002). After preprocessing, each participant's neural response to each item was then modeled using a gamma-convolved hemodynamic response function after onset of the display of the items during the study blocks and were smoothed with a 5-mm FWHM Gaussian kernel. Regressor covariance estimates output by FSL during univariate analysis confirmed that these portions of the trial were statistically separable (no two regressors were correlated at greater than 0.25) because of the jittered-duration fixation periods inserted between each trial. Beta values used in the RSA (described below) were calculated from the contrast of studied item (separated by study block by a run-level regressor to account for the blocked design) compared to jittered fixation baseline. Participant anatomical data for the searchlight RSA were prepared from 1-mm T1-weighted images using FreeSurfer and PyMVPA's *prepafni-surf* function (Fischl, 2012).

Searchlight RSA

We used a surface-based searchlight mapping technique with a 5-voxel radius and with white matter excluded (Oosterhof, Wiestler, Downing, & Diedrichsen, 2011) to produce a whole-brain map for each participant (registered to Montreal Neurological Institute space). The resultant value at each surface node reflected the z-scored cosine similarity between local neural DM (untargeted neural dissimilarity between all of the items) and the target similarity structure (the a priori Word2vec model). The a priori Word2vec semantic DM was created to probe for semantic similarity with a model that is sensitive to fine-grained semantic differences between words but does not account for differences between whether the item was presented in word or picture form. Specifically, the DM for the searchlight RSA was created using the Word2vec (Mikolov et al., 2013) dissimilarity between each pairwise combination of the items viewed by participants in the object name and object picture blocks, with the diagonal values discarded and not used in further analyses (Figure 3). One item was not able to be modeled by Word2vec ("rollerskate") and was discarded from further analysis, leaving 119 total items in the analysis.

The local neural DM for each participant at each surface node was computed using z-scored cosine similarity between activity patterns for all possible pairings of the 119 items (7021 pairwise dissimilarities, total). Activity patterns were defined by the (2 mm^3) voxel-wise estimated hemodynamic responses from general linear model analysis of the functional data, outlined in detail above. These analyses were performed using Python and PyMVPA (www.pymvpa.org; Hanke et al., 2009), SciPy (scipiy.org), and NumPy (numpy.scipy.org). At each searchlight location, we calculated the z-scored cosine similarity between the resultant DMs with the semantic item model DM, yielding a whole-brain map for each participant. To correct for multiple comparisons and determine the likelihood that the observed correlations could have occurred because of chance, we conducted a permutation test to compare our observed results to a null distribution of 10,000 random permutations of the modeled semantic DM (Winkler et al., 2016). The values displayed in our RSA results were calculated as the number of times the actual observed average mean-centered cosine similarity exceeded the average mean-centered cosine similarity at a given searchlight across participants for permuted observations, divided by 10,000 (Figure 4).

Whole-Brain Correlation with Attentional Bias

Using the whole-brain permutation-corrected z-map RSA results for each participant as input, we correlated these results with each participant's individual word attentional bias score, which was calculated by subtracting the percentage of trials that a participant responded to pictures from the percentage of the time the participant responded

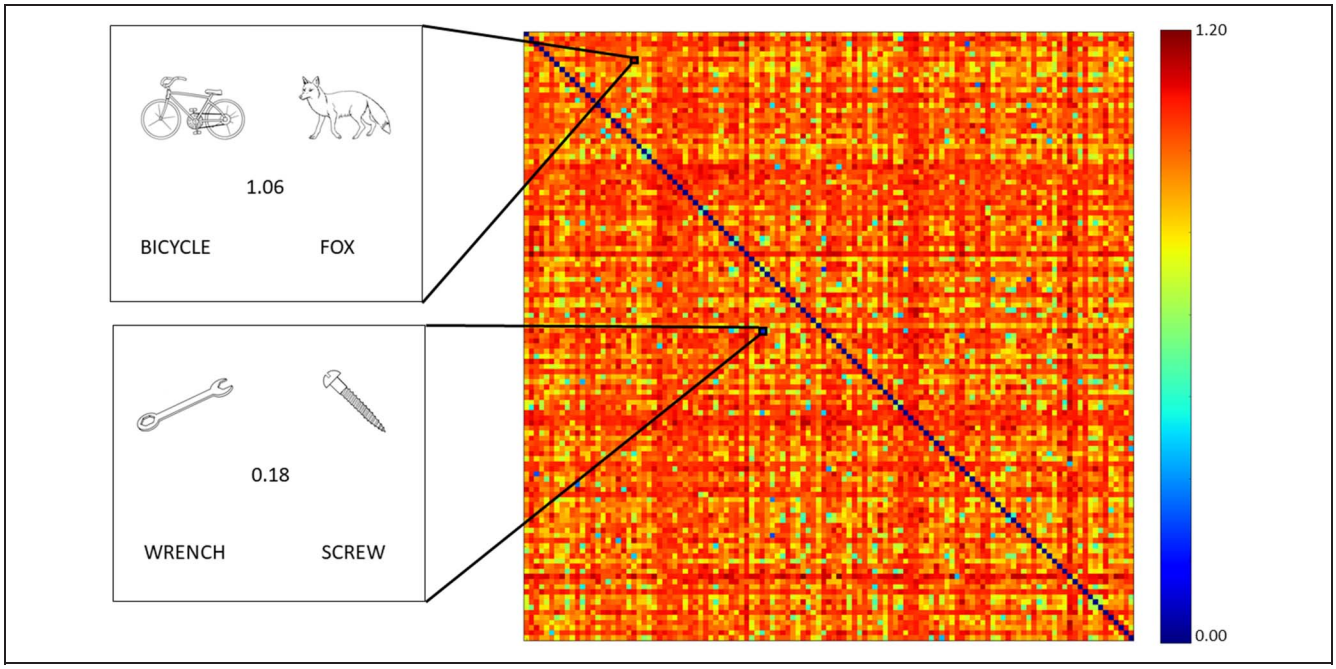

Figure 3. Semantic model of dissimilarity of zero-order relations, using Word2vec. Each item had its pairwise semantic dissimilarity calculated using Word2vec, and this model of semantic dissimilarity was used in the searchlight RSA with participant neural dissimilarity for each item.

to the word. The resulting score ranged from 1 (responded only to words) to −1 (responded only to pictures). A participant with a score of 0.8 would respond to the word 80% of the time, a participant with a score of −0.6 would respond to the picture 60% of the time, and a participant with a score of 0 would respond to words and pictures at equal rates. Any trials where participants provided invalid responses (such as responding "club" to a trial with a picture of a heart and the word "spade") were discarded and not included in the calculation of this score. After correlation with RSA maps, this resulted in a single surface map where the value at each node was the correlation between how biased the participant was toward stimuli presented as a word and the permutation-corrected z-score of how closely the pattern of neural activity at that node reflected the patterns of semantic relationships between items. In summary, the correlation map identifies neural regions where the patterns of neural activity match the patterns of semantic similarity better for participants who attend to words more than pictures. This correlation map was permutation corrected to account for the large number of correlations in a similar way to how the individual participant RSA correlations were corrected. We generated a distribution of 1000 potential correlations at each node

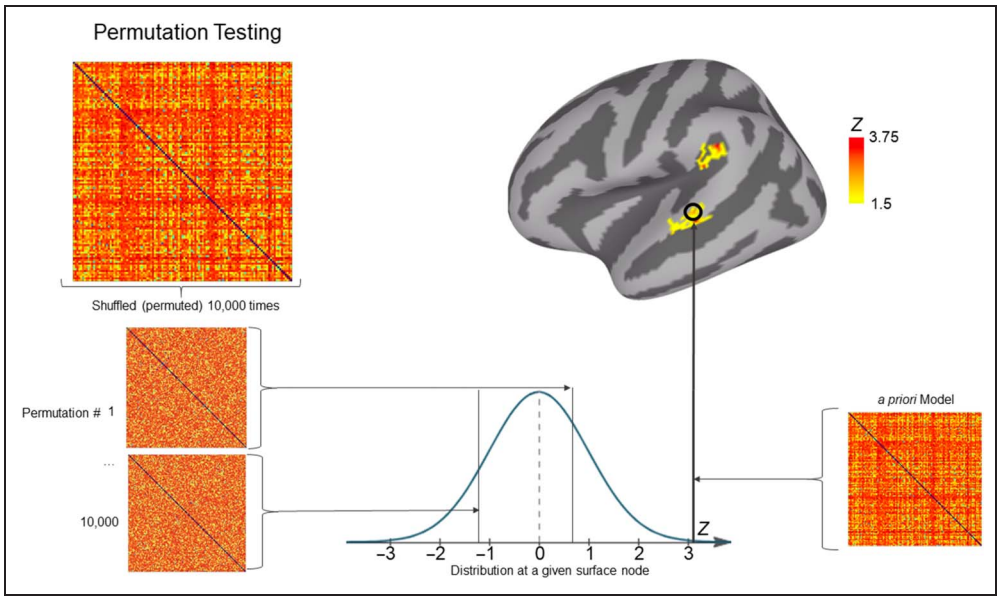

Figure 4. Permutation correction process. The a priori Word2vec model of semantic similarity is first shuffled (or permuted) 10,000 times to create a distribution of possible correlations at each surface node. Because this distribution is based off of the actual data, no assumptions about the distribution need to be made. We then calculate how much greater the actual observed similarity was compared to the average similarity, and that resulting z value is displayed on our later RSA surface maps.

by shuffling the ordering of the word attentional bias score vector and calculated the z score of the likelihood that the correlation at that node was because of chance. The final surface map shows the permutation corrected z scores, which represent the strength of the patterns of neural activity similar to the Word2vec model of semantic similarity for participants who attend to word information relative to participants who attend to picture information.

NeuroSynth "Speech Production" Correlational Analysis

Based on the results of our previous studies (Alfred et al., 2020; Hayes et al., 2020) and the fact that many participants in the previous study reported a strong preference for subvocal repetition strategies (e.g., "Repeated the names of the items to yourself") compared to other types of verbal strategies (e.g., "Constructed a story based on the items") in the memory task, we wanted to determine if the strength of a participant's word attentional bias score predicted the strength of the similarity between the Word2vec model of semantic similarity and patterns of neural activity in speech production regions. Rather than use an anatomically defined region, we used the NeuroSynth (www.neurosynth.org; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) term-based association test meta-analytic map for "speech production." This association test meta-analytic map identifies brain regions that are selectively active in studies about speech production by comparing the rates of being identified as an active region in studies mentioning "speech production" (86 studies) to the rates of being identified as an active region in all other studies (14,371 total studies). The "speech production" association test map is thresholded at false discovery rate-corrected $p <$ .01. This speech production map was binarized and used to mask each participant's permutation-corrected RSA z-map. Inside the mask, each participant's average z score was calculated, so that each participant had one value corresponding to the average degree of similarity that their patterns of neural activity in that mask corresponded with the Word2vec model. These values were correlated with each participant's word attentional bias score to calculate the degree to which the strength of the semantic representation correlated with the degree to which each participant was attentionally biased toward words over pictures.

RESULTS

Attentional Bias Task

The purpose of this task was to measure how frequently participants selectively attended to information provided by words instead of pictures. For each participant, we calculated a word attentional bias score, which was the rate that participants responded based on the word in the incongruent trials instead of the picture, or the percentage of trials that a participant responded to pictures subtracted from the percentage of the time the participant responded to the word, resulting in a score that ranged from 1 (responded only to words) to −1 (responded only to pictures). were split on whether they preferentially responded to words or pictures, although participants tended to be internally consistent across trials (Figure 1). Many participants had a clear preference for either words or pictures, and even the three participants with the attentional bias scores closest to 0 still chose their preferred content type between 10% and 20% more than the other content type (Figure 5).

Word attentional bias was significantly negatively correlated with higher accuracy during the picture memory test,

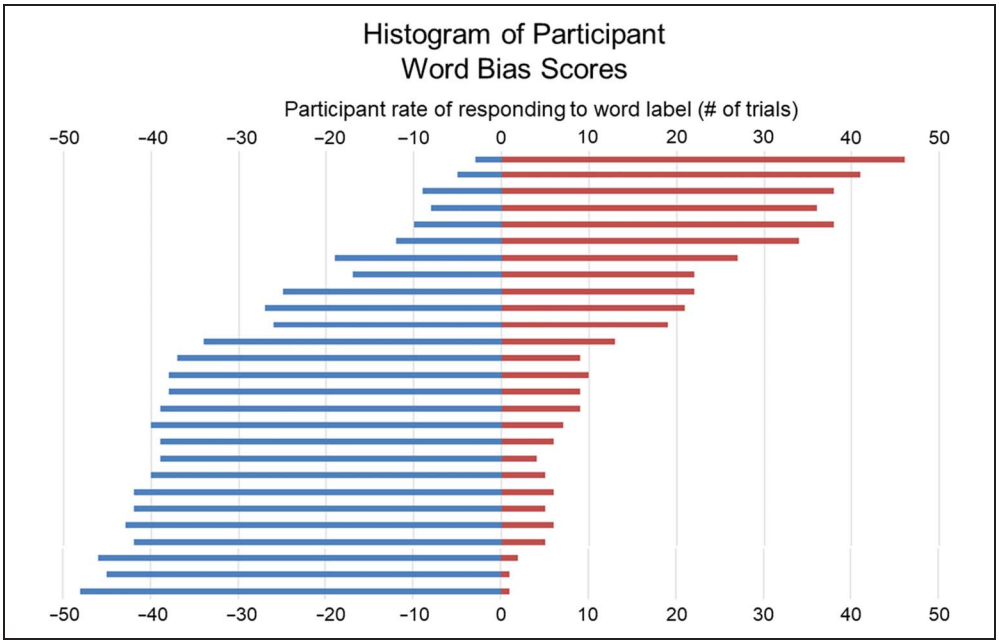

Figure 5. Distribution of attentional bias scores. The majority of participants showed a clear attentional bias toward either word or picture information. Although three participants seemed to be split between the two, they still preferred one content type between 10% and 20% more than the other.

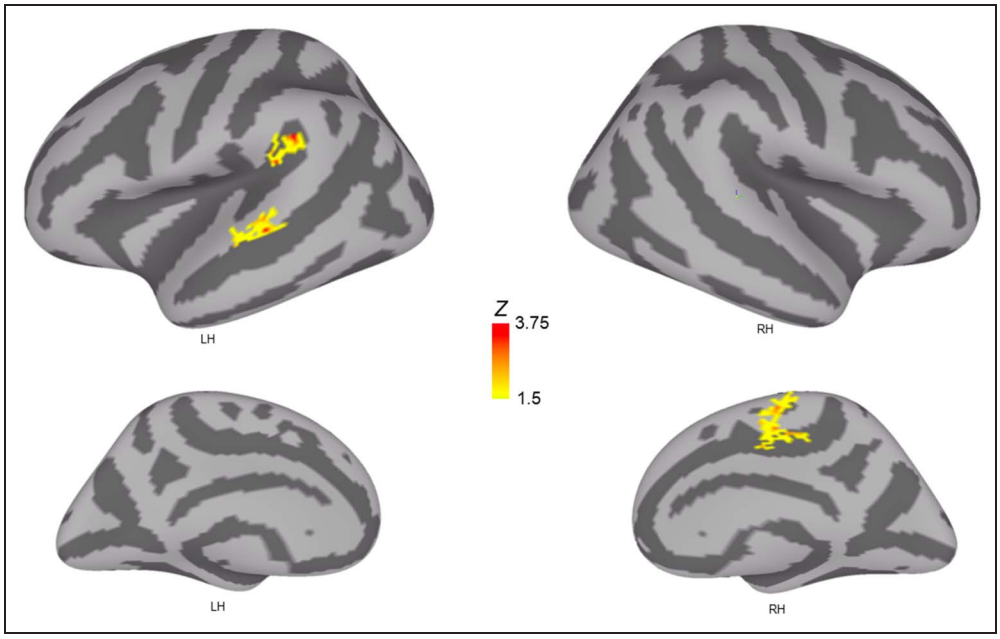

Figure 6. Brain regions where word attentional bias scores predict strength of semantic representation. In the displayed clusters, participants with a stronger attentional bias toward words show patterns of neural activity that more closely reflect the semantic structure of zero-order relations predicted by Word2vec than participants with attentional bias toward pictures.

$r(24) = -.46, p = .018$, Cohen's $d = 1.04$. The comparable verbal–visual cognitive style subtraction score did not correlate significantly with the picture memory test, $r(24) = -.18$, $p = .38$, Cohen's $d = 0.36$. Furthermore, this measure of attentional bias has been shown to be able to predict behavioral outcomes that were not predicted by any other measure (for details, see the Supplementary Material of Alfred et al., 2020).

Searchlight RSA

We ran a whole-brain surface searchlight RSA on the individual participant level to understand where in the brain patterns of neural activity best reflect the structure of semantic similarity (as defined by the Word2vec model). To identify potential brain surface regions where all participants had patterns of brain activity that correlated well with patterns of semantic similarity between items, we created one map that was the average map of all participants. There were no significant clusters (area ≥ 136 mm^2, $p < .05$, calculated through AFNI's surface-based implementation of ClustSim bootstrapped cluster extent threshold, *slowsurfclustsim*), indicating that there were no brain regions that reliably represented the semantic structure predicted by Word2vec across all participants.

Correlation between RSA and Word Attentional Bias

To check for effects of individual differences in the localization of the representation of the structure of the zero-order semantic relations, we created a whole-brain map that correlated each participant's permutation-corrected z score at each node with that participant's word attentional bias score. This resultant map shows which brain regions have a higher degree of similarity between the patterns of neural activity and the modeled semantic structure predicted by the degree to which an individual selectively attends to information presented as a word instead of a picture. The RSA and word attentional bias map was permutation corrected for multiple comparisons, similarly to how participant data were permutation corrected for the original searchlight RSA. Using the same bootstrapped cluster extent threshold (area ≥ 136 mm^2), we identified three significant surface clusters (see Figure 6 and Table 1). Two of the three clusters were left-lateralized, in the supramarginal (SMG) and superior temporal sulcus, which have previously been found to be associated with language processing (Sliwinska, Khadilkar, Campbell-Ratcliffe, Quevenco, & Devlin, 2012; Stoeckel, Gough, Watkins, & Devlin, 2009; Xu, Kemeny, Park, Frattali, & Braun, 2005; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003). The cluster in the left SMG is similar to a cluster identified by Kraemer, Rosenberg, and Thompson-Schill (2009) as being more active in participants with a verbal cognitive style (as defined by the Visualizer-Verbalizer Questionnaire; Kirby, Moore, & Schofield, 1988). Furthermore, Kraemer, Hamilton, Messing, DeSantis, and Thompson-Schill (2014) found that rapid TMS applied to the left SMG during a task that required applying verbal labels to pictures affected the performance of participants with a higher verbal cognitive style compared to participants with a visual cognitive style. These results indicate that the left SMG in particular is not only important for language cognition, but that this region is particularly important for individuals who prefer verbal material and may be using a strategy that involves verbal labeling of visual items.

We also calculated the same correlation map with the word attentional bias scores flipped (calculated as *% of incongruent trials participant responded to picture − %*

Table 1. Significant Clusters from Permutation-Corrected Whole-Brain Correlation with RSA

Region	*Average Permuted z*	*Maximum Permuted z*	*Cluster Size (mm^2)*	*Peak Coordinates (x, y, z), MNI*
Left SMG	2.185	3.593	160.06	−56, −36, 0
Right Middle Cingulate	2.158	3.212	385.29	8, −20, 62
Left STS	2.151	3.553	194.46	−52, −42, 30

MNI = Montreal Neurological Institute; STS = superior temporal sulcus.

of incongruent trials participant responded to word) to check for regions where patterns of activity that better reflect semantic structure are more likely in participants who are attentionally biased toward pictures. There were no clusters that were significant at the bootstrapped extent (area ≥ 136 mm^2). The closest three clusters were 83, 76, and 59 mm^2, and were located in the left middle orbitofrontal cortex, left superior occipital gyrus, and right precentral gyrus.

Correlation within Speech Production Regions

The NeuroSynth "speech production" association test meta-analytic map was used to define which regions were considered to be related to speech production (for more details about the speech production map, see Methods section). This speech production map was used to mask each participant's individual permutation-corrected RSA z-map. We then calculated the average z in the speech production map for each participant and correlated that value with participant word attentional bias scores. We found that, inside the NeuroSynth speech production network, the average z values from the permutation-corrected semantic Word2vec RSA correlated with participant word attentional bias scores, $r(24) = .41$, $p = .038$, Cohen's $d = 0.9$ (Figure 7). This is consistent with our prior findings (Alfred et al., 2020) and indicates that individuals who demonstrated a stronger attentional bias toward words have stronger representations of zero-order semantic relations in speech production regions.

DISCUSSION

Overall, these results indicate that there is a significant degree of reliable individual variation along the dimension of attentional bias and that measure predicts different localization of zero-order semantic relations as well as the strength of that representation. Furthermore, individual differences are critical to being able to identify the brain regions where the pattern of neural activity reflects the zero-order semantic structure between those items. When the analysis was run across all participants, there were no regions that significantly represented the zero-order semantic structure. However, when testing for the correlation between the zero-order semantic structure and word attentional bias, several brain regions showed that word attentional bias modulated the strength of the representation of zero-order relationships in these regions. (as shown in Figure 6). No regions were found to show activity correlating with picture bias. However, it is possible

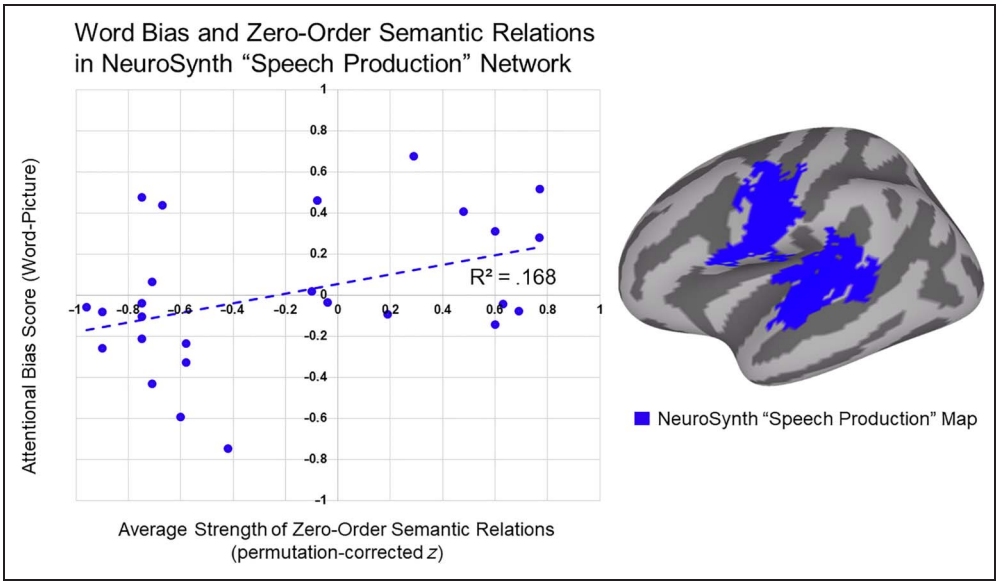

Figure 7. Word attentional bias score significantly correlates with the strength of the zero-order relations of semantic representation in speech production regions. Participant average permutation-corrected z scores from the RSA with the Word2vec semantic model were predicted by participants' word attentional bias scores, indicated stronger representation of semantic information in speech production regions for participants who selectively attend to word information.

that such regions exist, despite the fact that our analysis did not detect them. Furthermore, research by Blazhenkova and Kozhevnikov (2009) has suggested that rather than a single, unified visual cognitive style, some individuals might respond preferentially to object imagery and spatial imagery. The present task was not designed to assay that distinction, so it is possible that individual differences in nonverbal attentional bias are not captured here. These results are also consistent with the results of our earlier study, Alfred et al. (2020), which identified regions reflecting coarse semantic similarity (where activity patterns during the object names condition and object pictures condition were similar to each other, but dissimilar to pseudowords and abstract pictures conditions) that differed based on word attentional bias. That study showed that individual biases toward verbal or visual processing affected the encoding of semantically meaningful content regardless of the format (word or picture) in which that content was presented. Participants with strong implicit bias toward visual cognition gave responses that corresponded to the pictorial information presented, whereas those who responded preferentially to presented words were said to have a verbal bias. At the condition level, these attentional biases predicted differences in behavioral performance for a picture-only memory test and gave rise to distinct patterns of neural activation in response to both words and pictures (e.g., participants with a verbal attentional bias had neural patterns of activity that were similar for both words and pictures in verbally-associated regions). Importantly, these individual differences were only observed for semantically unambiguous and meaningful content (i.e., highly imageable words and easily namable pictures) and not for easily pronounceable pseudowords or abstract pictures. Searchlight RSA revealed that brain areas associated with content-general semantic encoding were more similar between participants, consistent with previous research (Frankland & Greene, 2015; Shinkareva, Malave, Mason, Mitchell, & Just, 2011; Binder, Westbury, McKiernan, Possing, & Medler, 2005; Thompson-Schill, 2003), whereas individual differences were reflected in material specific areas (such as the speech production network). Particularly, this study expands on the results of Shinkareva et al. (2011), who were able to use multivariate pattern analysis to decode object category from cross-format stimuli (e.g., identify a picture's category based on the viewing of the word that labeled that picture). Through our approach in this study, we were able to similarly identify cross-format semantic structure, even with participants only seeing a single presentation of each item in one format rather than seeing the item in both word and picture form. The results of the current study also provide a significant advance over our earlier study by using Word2vec to model the pairwise zero-order relational semantic similarity between items, regardless of whether the item was originally presented as a word or a picture, instead of a condition-level analysis that could only probe for regions that represented meaningful content in two formats (word and picture) similarly to each other while controlling for basic stimulus features. Despite the significant changes in the analysis approach, both studies highlight the importance of considering individual differences in the representation of zero-order semantic relations.

Other research has suggested that the dimensions of verbal and visual cognition represent an important domain of individual differences in cognition (for a review, see Alfred & Kraemer, 2017). Prior research has shown that individuals differ in the extent to which they prefer to attend to information in each of these dimensions (Miller et al., 2009; Kirchhoff & Buckner, 2006; Casasanto et al., 2002; Miller et al., 2002). For example, Zarnhofer et al. (2013) found that individuals who have a verbal cognitive bias show increased activation in language-associated areas of the left hemisphere (e.g., the left angular gyrus) during mental calculation. Neural differences in information encoding and retrieval that are biased toward either verbal or visual cognition are stable across time (Miller, Donovan, Bennett, Aminoff, & Mayer, 2012; Miller et al., 2009), and correlate with self-reported verbal or visual habits of thought (Kraemer et al., 2009, 2014, Miller et al., 2012; Hsu, Kraemer, Oliver, Schlichting, & Thompson-Schill, 2011; Kirchhoff & Buckner, 2006). Given the extensive body of work that has continued to demonstrate that there are high degrees of reliable variation between individuals in the processing of zero-order semantic relations, more studies should include measures of individual differences when attempting to model the structure of these representations.

Acknowledgments

The authors would additionally like to thank Daniel J. Harris and Carissa A. Crawford for their contributions toward designing the attentional bias measure and collection of behavioral and neural data for this study.

Reprint requests should be sent to Katherine L. Alfred, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, or via e-mail: katherine.l.alfred.gr@dartmouth.edu.

Funding Information

The authors would like to thank the National Science Foundation (DRL-1661088 to D. J. M. K.) for funding this research.

REFERENCES

Alfred, K. L., Hayes, J. C., Pizzie, R. G., Cetron, J. S., & Kraemer, D. J. M. (2020). Individual differences in encoded neural representations within cortical speech production network. *Brain Research*, *1726*, 146483. **DOI:** https://doi.org/10.1016/j.brainres.2019.146483, **PMID:** 31585067

Alfred, K. L., & Kraemer, D. J. M. (2017). Verbal and visual cognition: Individual differences in the lab, in the brain, and in the classroom. *Developmental Neuropsychology*,

42, 507–520. **DOI:** https://doi.org/10.1080/87565641.2017.1401075, **PMID:** 29505308

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*, 2767–2796. **DOI:** https://doi.org/10.1093/cercor/bhp055, **PMID:** 19329570, **PMCID:** PMC2774390

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*, 905–917. **DOI:** https://doi.org/10.1162/0898929054021102, **PMID:** 16021798

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*, 638–663. **DOI:** https://doi.org/10.1002/acp.1473

Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, *15*, 239–249. **DOI:** https://doi.org/10.1093/cercor/bhh126, **PMID:** 15238433

Casasanto, D. J., Killgore, W. D. S., Maldjian, J. A., Glosser, G., Alsop, D. C., Cooke, A. M., et al. (2002). Neural correlates of successful and unsuccessful verbal memory encoding. *Brain and Language*, *80*, 287–295. **DOI:** https://doi.org/10.1006/brln.2001.2584, **PMID:** 11896642

Chiang, J. N., Peng, Y., Lu, H., Holyoak, K. J., & Monti, M. M. (2020). Distributed code for semantic relations predicts neural similarity during analogical reasoning. *Journal of Cognitive Neuroscience*, 1–13. **DOI:** https://doi.org/10.1162/jocn_a_01620, **PMID:** 32762520

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*, 1193–1201. **DOI:** https://doi.org/10.1093/brain/awg104, **PMID:** 12690058

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*, 60–74. **DOI:** https://doi.org/10.1111/1469-8986.00120, **PMID:** 14693001

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*, 774–781. **DOI:** https://doi.org/10.1016/j.neuroimage.2012.01.021, **PMID:** 22248573, **PMCID:** PMC3685476

Frankland, S. M., & Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 11732–11737. **DOI:** https://doi.org/10.1073/pnas.1421236112, **PMID:** 26305927, **PMCID:** PMC4577152

Golby, A. J., Poldrack, R. A., Brewer, J. B., Spencer, D., Desmond, J. E., Aron, A. P., et al. (2001). Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding. *Brain*, *124*, 1841–1854. **DOI:** https://doi.org/10.1093/brain/124.9.1841, **PMID:** 11522586

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. M. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 2703–2708. **DOI:** https://doi.org/10.1073/pnas.95.5.2703, **PMID:** 9482951, **PMCID:** PMC19469

Green, A. E., Fugelsang, J. A., Kraemer, D. J. M., & Dunbar, K. N. (2008). The micro-category account of analogy. *Cognition*, *106*, 1004–1016. **DOI:** https://doi.org/10.1016/j.cognition.2007.03.015, **PMID:** 17511980

Green, A. E., Fugelsang, J. A., Kraemer, D. J. M., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, *1096*, 125–137. **DOI:** https://doi.org/10.1016/j.brainres.2006.04.024, **PMID:** 16750818

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2010). Connecting long distance: Semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex*, *20*, 70–76. **DOI:** https://doi.org/10.1093/cercor/bhp081, **PMID:** 19383937

Gross, M. M. (1972). Hemispheric specialization for processing of visually presented verbal and spatial stimuli. *Perception & Psychophysics*, *12*, 357–363. **DOI:** https://doi.org/10.3758/BF03207222

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Fründ, I., Rieger, J. W., et al. (2009). PyMVPA: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, *3*, 3. **DOI:** https://doi.org/10.3389/neuro.11.003.2009, **PMID:** 19212459, **PMCID:** PMC2638552

Hayes, J. C., Alfred, K. L., Pizzie, R. G., Cetron, J. S., & Kraemer, D. J. M. (2020). Individual differences in white and grey matter structure associated with verbal habits of thought. *Brain Research*, *1742*, 146890. **DOI:** https://doi.org/10.1016/j.brainres.2020.146890, **PMID:** 32439344

Hsu, N. S., Kraemer, D. J. M., Oliver, R. T., Schlichting, M. L., & Thompson-Schill, S. L. (2011). Color, context, and cognitive style: Variations in color knowledge retrieval as a function of task and subject variables. *Journal of Cognitive Neuroscience*, *23*, 2544–2557. **DOI:** https://doi.org/10.1162/jocn.2011.21619, **PMID:** 21265605

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466. **DOI:** https://doi.org/10.1037/0033-295X.104.3.427

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841. **DOI:** https://doi.org/10.1016/s1053-8119(02)91132-8, **PMID:** 12377157

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, *13*, 169–184. **DOI:** https://doi.org/10.1016/0361-476X(88)90017-3

Kirchhoff, B. A., & Buckner, R. L. (2006). Functional-anatomic correlates of individual differences in memory. *Neuron*, *51*, 263–274. **DOI:** https://doi.org/10.1016/j.neuron.2006.06.006, **PMID:** 16846860

Kraemer, D. J. M., Hamilton, R. H., Messing, S. B., DeSantis, J. H., & Thompson-Schill, S. L. (2014). Cognitive style, cortical stimulation, and the conversion hypothesis. *Frontiers in Human Neuroscience*, *8*, 15. **DOI:** https://doi.org/10.3389/fnhum.2014.00015, **PMID:** 24523687, **PMCID:** PMC3905265

Kraemer, D. J. M., Rosenberg, L. M., & Thompson-Schill, S. L. (2009). The neural correlates of visual and verbal cognitive styles. *Journal of Neuroscience*, *29*, 3792–3798. **DOI:** https://doi.org/10.1523/JNEUROSCI.4635-08.2009, **PMID:** 19321775, **PMCID:** PMC2697032

Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, *1428*, 13–23. **DOI:** https://doi.org/10.1016/j.brainres.2010.11.080, **PMID:** 21129363

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141. **DOI:** https://doi.org/10.1016/j.neuron.2008.10.043, **PMID:** 19109916, **PMCID:** PMC3143574

Lacey, S., & Campbell, C. (2006). Mental representation in visual/haptic crossmodal memory: Evidence from interference effects. *Quarterly Journal of Experimental Psychology*, *59*, 361–376. **DOI:** https://doi.org/10.1080/17470210500173232, **PMID:** 16618639

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, *29*, 289–315. **DOI:** https://doi.org/10.1016/0749-596X(90)90002-H

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. **DOI:** https://doi.org/10.1146/annurev.psych.57.102904.190143, **PMID:** 16968210

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, M. B., Donovan, C.-L., Bennett, C. M., Aminoff, E. M., & Mayer, R. E. (2012). Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. *Neuroimage*, *59*, 83–93. **DOI:** https://doi.org/10.1016/j.neuroimage.2011.05.060, **PMID:** 21651986

Miller, M. B., Donovan, C.-L., Van Horn, J. D., German, E., Sokol-Hessner, P., & Wolford, G. L. (2009). Unique and persistent individual patterns of brain activity across different memory retrieval tasks. *Neuroimage*, *48*, 625–635. **DOI:** https://doi.org/10.1016/j.neuroimage.2009.06.033, **PMID:** 19540922, **PMCID:** PMC2763594

Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., et al. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, *14*, 1200–1214. **DOI:** https://doi.org/10.1162/089892902760807203, **PMID:** 12495526

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277. **DOI:** https://doi.org/10.1093/oxfordjournals.bmb.a070866, **PMID:** 4937273

Milner, B., Corsi, P., & Leonard, G. (1991). Frontal-lobe contribution to recency judgements. *Neuropsychologia*, *29*, 601–618. **DOI:** https://doi.org/10.1016/0028-3932(91)90013-x, **PMID:** 1944864

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, *56*, 593–600. **DOI:** https://doi.org/10.1016/j.neuroimage.2010.04.270, **PMID:** 20621701

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987. **DOI:** https://doi.org/10.1038/nrn2277, **PMID:** 18026167

Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M., & Just, M. A. (2011). Commonality of neural representations of words and pictures. *Neuroimage*, *54*, 2418–2425. **DOI:** https://doi.org/10.1016/j.neuroimage.2010.10.042, **PMID:** 20974270

Sliwinska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J. T. (2012). Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology*, *3*, 161. **DOI:** https://doi.org/10.3389/fpsyg.2012.00161, **PMID:** 22654779, **PMCID:** PMC3361019

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. **DOI:** https://doi.org/10.1037/0278-7393.6.2.174, **PMID:** 7373248

Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex*, *45*, 1091–1096. **DOI:** https://doi.org/10.1016/j.cortex.2008.12.004, **PMID:** 19232583, **PMCID:** PMC2726132

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia*, *41*, 280–292. **DOI:** https://doi.org/10.1016/s0028-3932(02)00161-6, **PMID:** 12457754

Wagner, A. D., Poldrack, R. A., Eldridge, L. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Material-specific lateralization of prefrontal activation during episodic encoding and retrieval. *NeuroReport*, *9*, 3711–3717. **DOI:** https://doi.org/10.1097/00001756-199811160-00026, **PMID:** 9858384

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to??": Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience*, *20*, 682–693. **DOI:** https://doi.org/10.1162/jocn.2008.20055, **PMID:** 18052787

Winkler, A. M., Webster, M. A., Brooks, J. C., Tracey, I., Smith, S. M., & Nichols, T. E. (2016). Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping*, *37*, 1486–1511. **DOI:** https://doi.org/10.1002/hbm.23115, **PMID:** 26848101, **PMCID:** PMC4783210

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *Neuroimage*, *25*, 1002–1015. **DOI:** https://doi.org/10.1016/j.neuroimage.2004.12.013, **PMID:** 15809000

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670. **DOI:** https://doi.org/10.1038/nmeth.1635, **PMID:** 21706013, **PMCID:** PMC3146590

Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Ninaus, M., et al. (2013). Individual differences in solving arithmetic word problems. *Behavioral and Brain Functions*, *9*, 28. **DOI:** https://doi.org/10.1186/1744-9081-9-28, **PMID:** 23883107, **PMCID:** PMC3728072