Transient Neural Activation of Abstract Relations on an Incidental Analogy Task

Wei-Chun Wang, Liang-Tien Hsieh, Gowri Swamy, and Silvia A. Bunge

Abstract

■ Although a large proportion of the lexicon consists of abstract concepts, little is known about how they are represented by the brain. Here, we investigated how the mind represents relations shared between sets of mental representations that are superficially unrelated, such as car–engine and dog–tongue, but that nonetheless share a more general, abstract relation, such as whole–part. Participants saw a pair of words on each trial and were asked to indicate whether they could think of a relation between them. Importantly, they were not explicitly asked whether different word pairs shared the same relation, as in analogical reasoning tasks. We observed representational similarity for abstract relations in regions in the "conceptual hub" network, even when controlling for semantic relatedness between word pairs. By contrast, we did not observe representational

similarity in regions previously implicated in explicit analogical reasoning. A given relation was sometimes repeated across sequential word pairs, allowing us to test for behavioral and neural priming of abstract relations. Indeed, we observed faster RTs and greater representational similarity for primed than unprimed trials, suggesting that mental representations of abstract relations are transiently activated on this incidental analogy task. Finally, we found a significant correlation between behavioral and neural priming across participants. To our knowledge, this is the first study to investigate relational priming using functional neuroimaging and to show that neural representations are strengthened by relational priming. This research shows how abstract concepts can be brought to mind momentarily, even when not required for task performance. ■

INTRODUCTION

From an early age, we acquire abstract concepts by drawing analogy to concrete concepts (Gentner, 1983; Lakoff & Johnson, 1980). For example, we might learn in school that a genome contains instructions for making proteins, just as a cookbook contains recipes for making meals. Analogical reasoning, or the ability to find correspondences between the structures of distinct mental representations, is considered central to our capacity for learning and abstract thought (Hummel & Holyoak, 1997). Thus, understanding how the brain represents abstract concepts in the service of analogical reasoning and other forms of high-level cognition remains a significant question in cognitive neuroscience.

Although a large proportion of our lexicon consists of abstract concepts (Lupyan & Winter, 2018), little is known about how they are represented in the brain. Univariate fMRI studies show how abstract concepts are employed in cognitive processes such as analogical reasoning, but they are not particularly well suited for identifying mental representations of the abstract concepts themselves. That is, for regions activated during analogical reasoning,

it is not apparent whether these activations reflect processes that support reasoning or reflect the loci of stored representations. On the other hand, methods such as representational similarity analysis (RSA) allow researchers to identify multivoxel activation patterns associated with mental representations in the brain (Kriegeskorte, Mur, & Bandettini, 2008). RSA uses a multivariate approach to examine how activation patterns relate to similarity between items along a specified dimension. That is, if a given region represents a specific kind of information (e.g., abstract concepts), activation patterns between individual trials sharing that information should be more strongly correlated.

Multivariate methods such as RSA provide an ideal means for identifying mental representations of abstract concepts in the brain. Studies that use RSA to examine semantic memory suggest that representations of both concrete concepts (Wang, Wing, et al., 2018; Devereux, Clarke, Marouchos, & Tyler, 2013; Connolly et al., 2012) and abstract ones (Vargas & Just, 2020; Wang, Wu, et al., 2018) are distributed in regions associated with language and semantic processing (Wang, Brashier, Wing, Marsh, & Cabeza, 2018). These regions typically include the anterior, medial, and lateral temporal lobes; angular gyrus (AG); posterior cingulate cortex (PCC); precuneus; and several regions in pFC: the inferior frontal gyrus (IFG) and medial pFC (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Binder & Desai, 2011; Martin, 2007; also see Figure 1). This
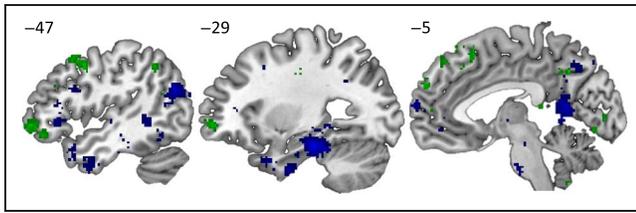
Figure 1. Regions implicated in semantic processing (blue) and analogical reasoning (green) based on Neurosynth meta-analyses (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011).

commonality may be because of the inherent difficulty of separating individual words that describe abstract concepts, such as "love," from concrete entities that they may be associated with, such as "mom" (Pecher, Boot, & Van Dantzig, 2011; Barsalou, 2008).

Thus, to approach the question of how abstract concepts are represented in the brain, we applied RSA to a different type of abstract concept. Drawing from the analogical reasoning literature, the current study examined representations of concepts that—at a broad level—describe the relation between concepts such as a car and an engine as well as a dog and a tail, which are semantically distant from each other but are both examples of whole–part relations. These types of abstract concepts are referred to in the reasoning literature as "second-order relations"; they are defined hierarchically as arising from the process of jointly considering and abstracting over first-order relations (such as car–engine)—that is, looking beyond the concrete features of the word pairs to more general correspondences (Hummel & Holyoak, 1997).

To isolate the neural substrates of analogical reasoning, fMRI studies typically examine univariate activation differences in second-order compared to first-order relational processing (e.g., Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2008; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; Bunge, Wendelken, Badre, & Wagner, 2005; Christoff et al., 2001). These studies most consistently implicate left rostrolateral pFC, inferior parietal lobule, and dorsolateral pFC in analogical reasoning (Hobeika, Diard-Detoeuf, Garcin, Levy, & Volle, 2016; Vendetti & Bunge, 2014; Krawczyk, 2012; see Figure 1) and suggest that left rostrolateral pFC and inferior parietal lobule, in particular, play a domain-general role in evaluating second-order relations (Hobeika et al., 2016; Wendelken, Chung, & Bunge, 2012). Critically, studies also report that activations associated with second-order relational processing are not because of differences in task difficulty (Vendetti & Bunge, 2014; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2010; Wendelken et al., 2008) or encoding demands (Wendelken & Bunge, 2010), further suggesting that regions such as rostrolateral pFC are sensitive to the level of relational abstraction that is needed during reasoning.

In addition to regions implicated in domain-general analogical reasoning across both semantic and visuospatial domains, domain-specific regions have also been identified. In particular, the left anterior IFG (Brodmann's areas 45 and

47) is activated for semantic analogical reasoning tasks (Whitaker, Vendetti, Wendelken, & Bunge, 2018; Hobeika et al., 2016; Wendelken et al., 2008; Bunge et al., 2005); its role in these tasks is thought to reflect the selective retrieval of task-relevant semantic relations.

To date, we are aware of only one study that reports RSA for abstract relations (Chiang, Peng, Lu, Holyoak, & Monti, 2020). In that study, Chiang and colleagues used BART, a computational model of relational similarity (Lu, Wu, & Holyoak, 2019), to assess brain activation patterns related to second-order relations during an analogical reasoning task. They found neural similarity for abstract relations in frontal and parietal ROIs drawn from the reasoning literature, including left rostrolateral pFC, inferior parietal lobule, and IFG. This finding suggests that regions implicated in both domain-general and domain-specific reasoning contain information about second-order relations. Furthermore, it lays a foundation for testing whether or how abstract relations are represented by the brain in the absence of an explicit analogical reasoning judgment, that is, when there is no requirement to abstract a second-order relation from two first-order ones.

Here, we first examined whether neural representations of abstract relations transcend concrete examples of these relations. In other words, we sought to test whether there are regions in the brain that show neural similarity for abstract relations—that is, relational similarity—even after controlling for the strength of semantic associations between word pairs. Prior neuroimaging research on semantic memory and on analogical reasoning yield different predictions (see Figure 1). On the basis of studies investigating semantic representations, one would predict relational similarity effects in the temporal lobe (anterior, lateral, and medial), AG, PCC, precuneus, left IFG, and/or medial pFC (Lambon Ralph et al., 2017; Binder & Desai, 2011; Martin, 2007). On the other hand, on the basis of studies investigating the cognitive process of relational reasoning, one would predict effects in rostrolateral pFC, left inferior parietal lobule, and/or dorsolateral pFC (Hobeika et al., 2016; Vendetti, Matlen, Richland, & Bunge, 2015; Krawczyk, 2012).

We sought to further elucidate the nature of abstract concept representations in the brain by drawing on behavioral work suggesting that abstract relations can be automatically and readily activated depending on task demands. Although teachers often explicitly draw on analogies to teach new concepts (Richland & Simms, 2015; Vendetti et al., 2015; Dunbar, 2001), learning also frequently occurs in informal settings and unstructured interactions outside the classroom (Dunst, Bruder, Trivette, Raab, & McLean, 2001; Brownell & Hendrickson, 1950). Thus, it is critical to test whether and how abstract concept representations can be activated.

There is some behavioral evidence that second-order relations can be primed, such as when participants are faster to recognize car–engine after seeing dog–tail (Green, Fugelsang, Kraemer, & Dunbar, 2008; Green, Fugelsang, &

Dunbar, 2006; Estes, 2003; Spellman, Holyoak, & Morrison, 2001). This little-studied phenomenon is referred to as "relational priming." Relational priming is typically found when participants intentionally and explicitly consider second-order relation between word pairs (Green et al., 2008; Green, Fugelsang, & Dunbar, 2006; Estes, 2003; Spellman et al., 2001). However, there is also evidence that relational priming can occur even when participants are not aware of shared relations between pairs. For example, Popov and Hristova (2015) reported priming of second-order relations in a lexical decision task (i.e., deciding whether both presented letter strings were words), without instructing participants to jointly consider relations among word pairs presented on different trials. Although the authors acknowledged potential methodological confounds with the stimuli—namely, greater semantic similarity for primed than unprimed pairs—priming was significant even after controlling for semantic similarity (also see Estes & Jones, 2006). This finding (Popov & Hristova, 2015) is consistent with evidence of spontaneous and unintentional activation of analogies during reading comprehension (Day & Gentner, 2007).

Taken together, these prior studies provide evidence of distinct mental representations of abstract concepts, which are separable from representations of concrete concepts and whose activation can be induced through priming. This work provides an impetus for examining the nature of abstract concept representations in the brain.

Here, using a different paradigm than prior studies examining relational priming, we examined whether abstract relations between word pairs can be primed without an explicit analogical reasoning judgment, that is, on an incidental analogy task in which participants were asked to make first-order relational judgments about a pair of words but sequentially presented word pairs sometimes shared an abstract relation. We hypothesized that sequential presentation of word pairs with a shared relation would yield relational priming because the relevant abstract concept would have just been activated. As a result, we predicted that participants would show relational priming at a behavioral level— that is, faster RTs for primed than unprimed pairs. Moreover, we also predicted relational priming at a neural level, that is, greater relational similarity for primed than unprimed pairs as a result of residual activation of the abstract relation when two examples of it have been presented sequentially.

Furthermore, we tested for an association between the magnitudes of behavioral and neural priming. Prior studies have, to our knowledge, only documented a brain–behavior association for repetition priming, that is, when the identical stimulus is repeated (e.g., Moore, Yi, & Chun, 2013). However, repetition priming and other forms of priming lead to different neural responses in the brain (Raposo, Moss, Stamatakis, & Tyler, 2006). Thus, it is an open question as to whether behavioral and neural indices of relational priming are associated. On the basis of our assumption that relational priming reflects the facilitated activation of

second-order relations, we predicted a correlation between behavioral and neural priming across participants. To our knowledge, this is the first study to investigate relational priming using functional neuroimaging.

The Current Study

In this study, participants saw a pair of words on each trial (e.g., car–engine) and indicated whether or not they could think of a relation between them (e.g., cars have engines). Word pairs were described by one of three types of abstract relations: category–exemplar (e.g., vehicle–boat), place–thing (e.g., school–eraser), and whole–part (e.g., dog–tail). Critically, we selected concrete nouns, as opposed to abstract words, as a stronger test of the hypothesis that representations of abstract concepts are distinct from the concrete concepts that comprise them. Each word pair was followed by a pair with the same (i.e., primed) or a different (i.e., unprimed) abstract relation. This task was designed to encourage incidental relational processing and priming of word pairs, without explicitly requiring an analogical judgment.

To examine how abstract concepts are represented in the brain, we used this novel paradigm involving sequential relational judgments to measure RSA for abstract relations. To ensure that neural similarity was not driven in some way by semantic similarity between word pairs, we controlled for the semantic distance between them. Next, we tested for relational priming at both the behavioral and neural levels. Finally, to assess whether the observed neural similarity is behaviorally meaningful, we tested for a relationship between behavioral and neural priming across individuals.

METHODS

Participants

Participants for the behavioral experiment comprised 25 young adults (aged 18–22 years; 16 women) recruited from the University of California, Berkeley. Behavioral participants received course credit through the Psychology Department's Research Participation Program. Three other behavioral participants were excluded because of outlier performance on the task (see Behavioral Results section). An additional 18 right-handed young adults (aged 18–28 years; 11 women) participated in the fMRI experiment. fMRI participants were recruited through ads posted online through Facebook groups associated with the University of California, Berkeley, and were paid for their participation.

Both behavioral and fMRI participants were native English speakers. In addition, fMRI participants were prescreened to be free of neurological disorders and MRI contraindications. All participants gave their informed consent to participate in the study, which was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley.

Task Overview

Participants were told that the experiment examined language processing. They were instructed to think about the relation between words within a pair (i.e., first-order relations, in the parlance of analogical reasoning). No reference was made to any shared relations across word pairs (i.e., second-order relations). Task instructions included one example of each of the three relation types, as well as one example of a user–tool relation (i.e., carpenter–hammer), that was not shown in the experiment. This latter example was included to minimize awareness that only three relation types were being tested and, as a result, reduce behavior that may discourage relational processing (e.g., labeling or rehearsing). Participants were told to press the left key if they could think of a relation between the words and to press the right key if they could not think of a relation.

Upon confirmation that they understood the task instructions, participants completed a practice run followed by four experimental runs of the relational judgment task. Afterward, they completed a posttest questionnaire that assessed the extent to which participants were aware of the experimental manipulations and strategies that they used to determine the relatedness of words. These responses are available in the Supplementary Materials.[1]

Task Stimuli and Study Design

Stimuli consisted of 288 semantically related word pairs and 24 semantically unrelated word pairs. Related word pairs were created by an author or a research assistant on this study. Each word pair was a first-order relation that served as an example of one of three abstract relations (96 pairs per relation) selected from Bejar, Chaffin, and Embretson (1991), consisting of category–exemplar (e.g., dessert–brownie), place–thing (e.g., temple–priest), and whole–part (e.g., book–page) relations. To demonstrate that abstract relations are measurable irrespective of the concrete concepts they are composed of, all words were selected to be concrete nouns. Category–exemplar and whole–part pairs were selected to be half living–living (e.g., dog–tail) and half nonliving–nonliving (e.g., tool–hammer). Place–thing pairs were selected to be half place–living (e.g., barn–cow) and half place–nonliving (e.g., bathroom–mirror). On the basis of these criteria, at least one author reviewed and approved each pair as an acceptable example of its corresponding relation type.

Unrelated word pairs (e.g., mascara–spoon) were included as catch trials, to reduce behavior that may discourage relational processing and to identify participants who were not performing the task as intended. Unrelated pairs were created by G. S. and approved by the first author as having no obvious associations. A full list of the stimuli can be found in the Supplementary Materials.

On each trial, a word pair was shown for 2 sec, followed by a variable duration fixation cross. For behavioral participants,

this fixation period ranged from 900 to 1100 msec to improve response inhibition and selection (Wodka, Simmonds, Mahone, & Mostofsky, 2009). For fMRI participants, this fixation period ranged from 1 to 6 sec to efficiently space out trials for event-related fMRI analyses. To examine relational priming, one third of related pairs were primed (i.e., preceded by a pair from the same relation type), whereas the remaining two thirds were unprimed. Trial timing and sequences were counterbalanced across participants and optimized for fMRI using optseq2 (surfer.nmr.mgh.harvard.edu/optseq). This counterbalancing of stimulus conditions also reduced the potential susceptibility of RSA to design-related confounds (Cai, Schuck, Pillow, & Niv, 2019).

Image Acquisition and Preprocessing

fMRI Data Collection

Images were collected on a 3-T Siemens MRI scanner at the Brain Imaging Center on the campus of University of California, Berkeley. Using a 32-channel head coil, functional volumes were acquired using an EPI sequence (matrix = 64×64, 34 descending axial slices, 3.5-mm isotropic voxels, repetition time = 2000 msec, echo time = 22 msec, flip angle = 50°). High-resolution structural volumes were collected using a 3-D, T1-weighted gradient-echo inversion pulse sequence (matrix = 240×256, 160 descending axial slices, 1-mm isotropic voxels, repetition time = 2300 msec, echo time = 2.98 msec, flip angle = 9°).

Preprocessing

Data were preprocessed with SPM12 (www.fil.ion.ucl.ac.uk/spm). After discarding the first two volumes of each run to reduce magnetic field inhomogeneity, functional volumes for each participant were slice-time corrected, realigned, and coregistered to their respective structural volumes. The structural volumes were then segmented into separate gray and white matter volumes to normalize the functional and structural volumes into Montreal Neurological Institute space. The functional volumes were not spatially smoothed to preserve fine-grained pattern information (Kriegeskorte et al., 2008).

Data Modeling and Analysis

Single-trial Beta Estimation

Each trial was modeled in a separate general linear model using the least squares separate approach (i.e., one regressor for the trial of interest and one regressor for all other trials), an optimal method for estimating single-trial activation patterns in event-related fMRI designs (Mumford, Turner, Ashby, & Poldrack, 2012). Additional covariates of no interest consisted of six motion parameters estimated during realignment, baseline and session effects, global mean, and motion outliers from the Artifact Detection

Toolbox (www.nitrc.org/projects/artifact_detect). This yielded a single-trial beta volume for each trial for each participant that was used to conduct the RSA.

Representational Dissimilarity Matrix Construction

With a view to measuring relational similarity at the neural level, we created representational dissimilarity matrices (RDMs) that represent the dissimilarity for all pairwise combinations of trials (i.e., word pairs). Each cell represents the value [1 − similarity] for a pair of trials. An RDM is typically created for each predictor (in this case, relation type) and for the dependent measure (e.g., multivoxel activation patterns). As schematized in Figure 2B, we first constructed an RDM of predicted relational similarity between word pairs based on whether they shared an overarching relation (top row).

To create neural RDMs for each participant (Figure 2B, center row), we extracted single-trial beta values—that is, patterns of activation across trials—for voxels in each searchlight sphere or in each ROI (described below). Then, we calculated the Pearson's correlation of the activation patterns between each pair of trials, yielding an RDM of neural pattern similarity for each searchlight sphere and ROI.

As a control, we also created an RDM of predicted semantic similarity based on Euclidean distance using word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Word2vec is a two-layer neural net model that uses co-occurrence statistics to compute word embedding vectors. To calculate semantic similarity between word pairs, we created an averaged word2vec vector for the two words in each pair before computing the Euclidian distance between pairs. This semantic RDM was then used to control for semantic similarity in a partial correlation analysis with the relational and neural RDMs (Figure 2B, bottom row).

RSA

We used the multivariate fMRI analysis toolbox (github.com/brg015/mfMRI_v2) to calculate RSA for second-order relations. For each participant, we calculated Spearman's correlations between the relational RDM and neural RDMs for voxels centered on each searchlight sphere (radius = 5 voxels) across all voxels in the brain—yielding a whole-brain RSA correlation map for second-order relations. These correlation maps were then Fisher transformed and analyzed with a one-sample t test; we corrected for multiple comparisons ($p < .05$) with 3dClustSim (Version 18.0.11; afni.nimh.nih.gov/) using an uncorrected threshold of $p < .001$ and a cluster extent of 30 voxels (for a discussion of cluster-level corrections, see Slotnick, 2017).

For follow-up analyses, we created ROIs from clusters exhibiting significant RSA for relations. Each cluster was treated as a separate ROI, except for the left temporal lobe cluster (Figure 4), which was split into medial fusiform gyrus (FG)/medial temporal lobe and lateral inferior temporal gyrus ROIs based on the Harvard Oxford Cortical Atlas (Desikan et al., 2006). We also interrogated a single ROI composed of all voxels across the whole set of clusters. The first ROI analysis examined whether RSA for relations remained significant after controlling for semantic similarity. For each ROI in each participant, we computed partial Spearman's correlations between the relational and neural RDMs while controlling for the semantic RDM.

The second ROI analysis examined whether RSA for relations differed between primed and unprimed trials.
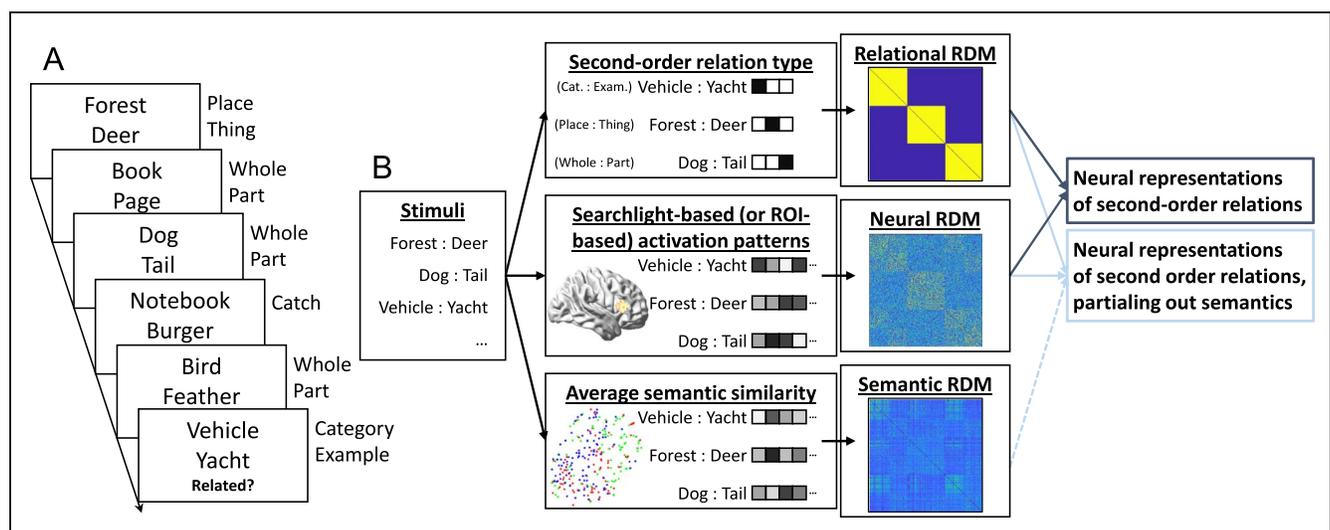

Figure 2. (A) Schematic of experimental design. (B) Schematic of RSA. RDMs were created based on second-order relation types and multivoxel activation patterns to assess neural representations of abstract concepts. In addition, as a secondary analysis, semantic similarity, based on the Euclidean distance between averaged word2vec vectors between word pairs, was partialed out to confirm that RSA for second-order relations was not driven by semantic relatedness.

Here, we created separate RDMs of relational, neural, and semantic similarity for primed and unprimed trials to examine priming effects. Correlations were Fisher-transformed and submitted to one- and paired-sample t tests ($p < .05$ false discovery rate [FDR]-corrected for multiple comparisons). Then, for each ROI, we computed the Pearson's correlation between the RT and RSA priming effects across participants to examine the relationship between behavioral and neural priming. In addition, for significant correlations (after $p < .05$ FDR correction), we used the Robust Correlation Toolbox (Pernet, Wilcox, & Rousselet, 2013) to compute bootstrapped Pearson's skipped correlations (95% confidence intervals [CIs]; $p < .05$), to account for potential outliers (Wilcox, 2004).

Finally, to provide a more complete view of our findings, we followed up with an exploratory pattern similarity analysis—a variant of RSA that has also been used to probe neural representations with fMRI (Dimsdale-Zucker & Ranganath, 2018). This analysis provided very similar and converging results (see Supplementary Materials).

RESULTS

Behavioral Results

After excluding three outlier participants from the behavioral study who responded two SDs below the average proportion of "related" responses to related pairs, we examined accuracy in the relational judgment task. The behavioral and fMRI groups performed well above chance (50%) for both related and unrelated pairs (Table 1). Subsequent analyses focused on accurate responses to related pairs, that is, trials on which participants correctly indicated that two words were related.

We then examined RTs to test for a relational priming effect—that is, to test whether, for correct "related" decisions to related pairs, participants responded more quickly if the trial was preceded by a pair with the same relation than a different relation. A mixed ANOVA was conducted with Repetition of the second-order relation (i.e., primed and unprimed) as a within-participant factor and Group (i.e., behavioral and fMRI participants) as a between-participant factor (Figure 3).

Table 1. Accuracy in the Relational Judgment Task

Accurate Responses	Behavioral Participants	fMRI Participants
"Related" to related word pairs	0.978 (0.003)	0.987 (0.002)
"Unrelated" to catch trials	0.767 (0.034)	0.725 (0.059)[a]

SEM in parentheses.

[a]The first six fMRI participants were not presented with unrelated catch trials; as a result, this calculation is based on the remaining 12 participants. Accuracy on the "related" word pairs for these six participants ($M = 98\%$, $SD = 2\%$) was within the range of the other fMRI participants.
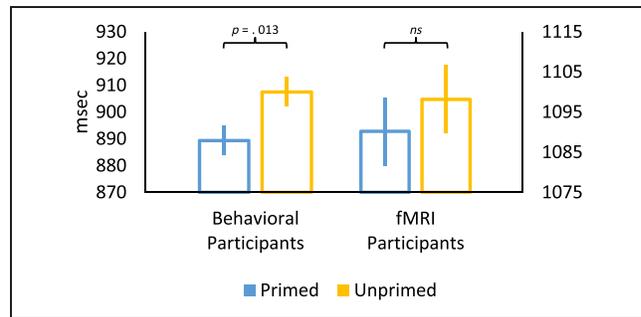

Figure 3. RTs for primed and unprimed second-order relations for behavioral and fMRI participants. Error bars reflect 95% within-participant CIs.

As predicted, RTs were faster for primed than unprimed word pairs, $F(1, 41) = 5.80$, $p < .05$, $\eta_p^2 = .12$, providing evidence of relational priming. Although the priming effect was not significant when considering fMRI participants alone (see Figure 3), the interaction between repetition and group was not significant, $F(1, 41 = 0.92$, $p = .34$, $\eta_p^2 = .02$, indicating that priming in fMRI participants trended in the same direction as behavioral participants. The only group difference that was significant is that RTs were slower for the fMRI participants than the behavioral ones, $F(1, 41) = 8.02$, $p < .005$, $\eta_p^2 = .16$, consistent with prior studies showing worse performance in the MRI environment (e.g., van Maanen, Forstmann, Keuken, Wagenmakers, & Heathcote, 2016). A follow-up analysis that excluded outlier RTs for each participant (defined as ± 3 median absolute deviations; see Leys, Ley, Klein, Bernard, & Licata, 2013) also revealed significant effects of relational priming, $F(1, 41) = 4.47$, $p < .05$, $\eta_p^2 = .10$, and group, $F(1, 41) = 6.01$, $p < .05$, $\eta_p^2 = .13$, and no significant Repetition × Group interaction, $F(1, 41) = 0.04$, $p = .85$, $\eta_p^2 = .001$. These results provide evidence for relational priming across pairs of trials, although participants were not asked to relate word pairs to each other.

Self-report Questionnaire

After the study was completed, participants reported on what they thought the goal of the study was, what strategies they had adopted, and what types of relations they had noticed (see Supplementary Materials Section II). Their answers to the question of strategy choice suggest that they did not explicitly categorize word pairs according to abstract relations. Participants who reported trying to verbalize a relation said they tried to form a sentence such as "A hammer *is a type of* tool." Thus, they tended to think of relations in a way that was stimulus bound. Furthermore, only 3 of 18 participants retrospectively named or described all three of the broad, overarching relations incorporated into our study design. Others named or described category–exemplar relations and either whole–part or place–thing relations but not both, suggesting that there may be conceptual overlap between these (e.g., a page is both part of a book and can

be found in a book). Importantly, none of the participants reported attempting to relate the current word pair to the previous one; that is, they did not treat this as an analogy task. Additional discussion and the full text of the fMRI participants' responses are presented in Supplementary Materials Section II. To summarize, participants' self-reports suggest that they performed the incidental analogy task as intended, considering first-order relations between the concrete nouns in each word pair.

fMRI Results

RSA for Second-Order Relations

We examined RSA for abstract relations in a whole-brain searchlight analysis. As illustrated in Figure 4, significant clusters were found in the left inferior and middle temporal gyri, bilateral FG, parahippocampal gyrus, right precuneus extending into PCC, and left IFG (see also Table 2). Voxels in these regions showed activation patterns that were more similar among examples of the same relation than different ones.

We then conducted ROI analyses on these clusters to test whether RSA for relations remained significant after partialing out semantic similarity, as measured with word2vec (see Methods). As shown in Table 2 (panel on the right, "All Trials" column), RSA for relations in each of these regions was still significant after controlling for semantic similarity. However, as noted below, the left IFG did not show as robust an effect as the other regions; for the most part, the regions exhibiting RSA were located in posterior regions.

Neural Priming Effects

Next, we examined whether repetition of abstract relations across consecutive trials (i.e., relational priming) affected the magnitude of RSA for relations within the significant clusters identified from the whole-brain analysis (Figure 4).

Given behavioral evidence of relational priming, we predicted that RSA would be greater between primed trial pairs compared to between unprimed trial pairs. Indeed, effects were significantly greater for primed than unprimed trials in many of the regions identified from the correlation with the whole-brain RDM (Table 2). In fact, the RSA effects were driven almost entirely by the primed trials. Specifically, left-lateralized inferior temporal gyrus, lateral occipital cortex/AG and middle temporal gyrus, and right precuneus/PCC showed RSA on primed trials but not unprimed ones. Only the left FG/parahippocampal gyrus showed significant RSA on unprimed trials, albeit numerically weaker than on primed trials. The left IFG and right FG/parahippocampal gyrus showed weak RSA effects that were not significant when subdivided by priming condition. Thus, neural overlap within relational categories was driven in large part by primed word pairs.

Finally, under the assumption that representations of second-order relations are distributed throughout the brain, we sought to test whether treating all clusters as a single ROI would yield/show similar, if not stronger, effects than each of the clusters individually. Indeed, we found stronger effects for the single, distributed ROI (Figure 4, right, and Table 2, bottom row). Both primed and unprimed trials exhibited RSA within the combined cluster, but the effect was stronger for primed trials. All in all, RSA associated with second-order relation types was distributed across predominantly posterior brain regions and was stronger when primed.

Brain–Behavior Correlations

Finally, we examined whether there was a relationship between behavioral responses to and neural representations of abstract relations. Toward this goal, we correlated behavioral priming (i.e., unprimed − primed RTs, where more positive numbers indicate a greater behavioral

Figure 4. Regions exhibiting significant RSA for second-order relations in a whole-brain searchlight analysis, corrected for multiple comparisons. See Table 2 for abbreviations. The right panel shows the RSA effects for all clusters treated as a single ROI. LOC = lateral occipital complex.
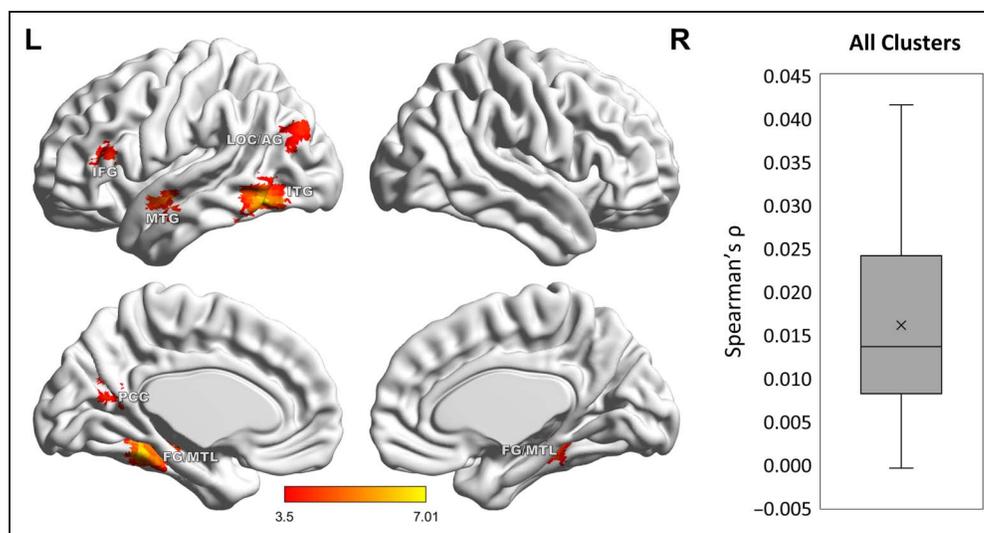
Table 2. Peak Voxels in Clusters Exhibiting Significant RSA for Second-Order Relations in a Whole-Brain Searchlight Analysis (Left Panel)

| | | MNI | | | | | | ROI-based Partial ρ^b | | | |
| | | | | | | | | All Trials | Primed Trials | Unprimed Trials | Primed > Unprimed |
k	BA^a	x	y	z	Z	Region	ρ				
233	37	−38	−52	−10	5.48	L FG	0.013	0.012^c	0.015^c	0.008^c	0.007
	36	−27	−42	−7	5.20	L parahippocampal gyrus (MTL)					
236	37/20	−48	−56	−10	5.42	L inferior temporal gyrus	0.011	0.011^c	0.020^c	0.006	0.014^c
91	19	−34	−84	28	3.88	L lateral occipital cortex	0.009	0.008^c	0.021^c	0.003	0.018^c
	39	−30	−77	38	3.29	L AG					
88	21	−55	−10	−14	4.69	L middle temporal gyrus	0.009	0.008^c	0.015^c	0.003	0.012^c
50	36	32	−35	−14	4.02	R parahippocampal gyrus (MTL)	0.006	0.005^c	0.006	0.002	0.004
	37	26	−38	−24	3.96	R FG					
47	23	−16	−63	14	3.77	Precuneus, PCC	0.009	0.008^c	0.017^c	0.004	0.013^c
40	46	−48	28	14	4.01	L inferior frontal gyrus	0.006	0.005^c	0.009	0.001	0.008
785	–	–	–	–	–	All clusters	0.016	0.015^c	0.031^c	0.008^c	0.023^c

MNI = Montreal Neurological Institute; MTL = medial temporal lobe.

[a]BA indicates approximate Brodmann's area. Several clusters spanned two regions, which are reported on separate rows.

[b]Spearman's correlations controlling for the effect of word2vec semantic similarity in ROI-based analyses (right panel).

[c]Significant ROI-based correlation at $p < .05$ following FDR correction.

priming effect) and neural priming (i.e., primed − unprimed RS, where more positive numbers indicate greater RSA for primed relations) across participants. Correlations were not significant between behavioral and neural priming in any of the individual clusters ($ps > .10$, uncorrected). However, there was a significant relationship when all voxels from the set of clusters were combined into a single ROI ($r = .63$; $p = .005$, uncorrected; $p < .05$, FDR corrected). That is, participants who showed more behavioral priming also showed more neural priming across a distributed set of brain regions that exhibit RSA for second-order relations (Figure 5A). Controlling for semantic similarity reduced the strength of this correlation only slightly ($r = .61$; $p = .008$, uncorrected; $p = .06$, FDR corrected). To reduce the influence of potential outliers, we computed a bootstrapped Pearson's skipped correlation, which ignores outliers, resamples the data, and computes correlation values and 95% CIs (Wilcox, 2004). This analysis also yielded a significant correlation ($r = .65$, 95% CIs [0.23, 0.86]; Figure 5B). Thus, individual differences in behavioral priming were correlated with individual differences in neural priming.
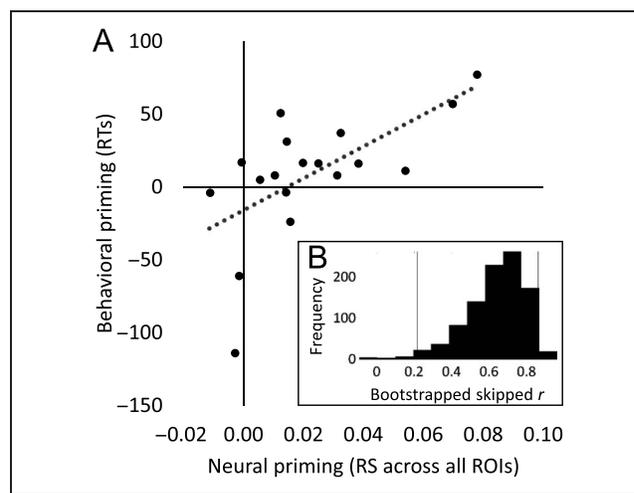

Figure 5. Across-participant (A) Pearson's correlation and (B) bootstrapped Pearson's skipped correlation between the RT and the RSA priming effects. ITG = inferior temporal gyrus; L = left; MTG = middle temporal gyrus; MTL = medial temporal lobe; R = right.

DISCUSSION

In the current study, we addressed two questions critical to our understanding of abstract concepts. First, is there overlap in the neural representation of word pairs that share an abstract relation but that are not semantically related to one another? Second, can abstract relations be automatically activated? If so, are there observable behavioral and neural indices of relational priming, and are they related to each other? Our study provided support in favor of each of these questions, as described below.

Representation of Abstract Concepts

In a whole-brain searchlight analysis, we found RSA for abstract relations in left-lateralized medial and lateral temporal lobe, AG, and IFG as well as precuneus and PCC. These regions are typically associated with semantic memory and processing (Lambon Ralph et al., 2017; Binder & Desai, 2011; Martin, 2007). However, our findings remained significant after partialing out semantic similarity, suggesting that our effects were not driven by semantic relatedness. Rather, our data are consistent with the proposal that these regions reflect a "conceptual hub" network that represents abstract concepts (Binder, 2016). Binder posits that the neuroanatomical characteristics of these regions allow them to serve as high-level convergence zones where multiple inputs are integrated to form supramodal conjunctive representations, such as of abstract concepts. This hypothesis is consistent with both our results and results from other multivariate studies that have reported abstract representations in lateral temporal, parietal, and pFC regions (Vargas & Just, 2020; Wang, Wu, et al., 2018). Together, this suggests that these high-level regions within the semantic memory network may be critical for the representation of abstract concepts.

On the other hand, unlike Chiang et al. (2020), we did not find RSA for second-order relations in regions associated with domain-general reasoning (i.e., rostrolateral pFC, inferior parietal lobule, and dorsolateral pFC) and found weak effects in anterior IFG, a region associated with semantic reasoning (Whitaker et al., 2018; Hobeika et al., 2016; Wendelken et al., 2008; Bunge et al., 2005). Although the left anterior IFG was identified in the whole-brain RSA map, it showed weaker RSA than other left-hemisphere regions and did not show a significant priming effect. There are, however, a few notable differences between Chiang et al.'s study and ours. First, they used a continuous measure of similarity derived from a computational model (Lu et al., 2019). In fact, in a secondary multi-voxel pattern analysis, they reported significant classification of second-order relation types in many of the same pFC, temporal, and parietal regions that we observed. Second, they utilized an explicit analogical reasoning task—that is, an explicit comparison of two word pairs—while we utilized an incidental task. These differences

lead us to believe that regions associated with analogical reasoning may be brought online to represent second-order relations only in the presence of an explicit reasoning judgment.

Potentially consistent with this hypothesis, which has yet to be tested directly, a recent study by Alfred, Connolly, and Kraemer (2018) examined RSA for newly learned mental models of spatial information (i.e., height) in a transitive inference task (a form of reasoning; Wendelken & Bunge, 2010). They found that activation patterns in the right inferior parietal lobule, IFG, and precuneus were related to differences in height. In a follow-up study, the authors showed that activation patterns in the right inferior parietal lobule, left IFG, and left anterior pFC were related to differences across three types of content (i.e., height, price, and an abstract nonsense descriptor; Alfred, Connolly, Cetron, & Kraemer, 2020). This set of findings further suggests that, in the presence of an explicit reasoning judgment, neural representations of abstract concepts can be detected in regions implicated in reasoning such as inferior parietal lobule. Future follow-up work should experimentally manipulate task orientation to test qualitative differences in how abstract concepts are represented in the brain. For example, prior studies show that task demands affect the strength of RSA (Wang, Brashier, et al., 2018); thus, we would predict that RSA for abstract relations would be significantly greater during an explicit analogical reasoning task compared to a semantic judgment task of the kind used here.

Behavioral and Neural Priming of Abstract Relations

As predicted, we observed priming of abstract relations, whereby RTs were faster for word pairs when the previous word pair had the same second-order relation than a different one. This behavioral result is consistent with prior work showing relational priming during a lexical decision task (Popov & Hristova, 2015), suggesting that abstract relations can be brought to mind in the absence of an explicit reasoning task. In our study, behavioral priming was observed although only three participants correctly identified all three abstract relations at the end of the study, suggesting that they could be activated as a result of "bottom–up" spreading of activation in a distributed semantic network.

Moreover, we found neural indices of relational priming; that is, the strength of neural similarity as a function of relational category was greater for primed than unprimed trials within the regions identified by the whole-brain analysis. Although there has been prior work examining how neural representations are affected by repetition priming, the current study is, to our knowledge, the first to show that neural representations are strengthened by relational priming. Moreover, strength of neural priming was correlated across participants with strength of behavioral

priming. Thus, the relational similarity effect measured by multivariate fMRI analysis is behaviorally meaningful.

In fact, neural similarity was driven almost entirely by the primed trials. There are at least two possible explanations for this effect. The first is that activation of relations at the highest level of abstraction (e.g., "bird–robin is a category–exemplar relation"), as opposed to a stimulus-bound relation (e.g., "a robin is a type of bird"), is more likely when participants have just responded to a semantically distant word pair from the same relational category. The second possible explanation is that, when there is a switch between relations, there may be residual activation of the preceding relation. In this case, there would be a mixture of active populations: one representing the preceding relationship and another representing the current relationship, thereby reducing relational similarity between unprimed pairs.

In either case, mental representations of abstract relations appear to be transiently activated, as they do not persist across intervening trials. Thus, we hypothesize—based on this finding as well as the fact that participants did not explicitly relate word pairs to one another but were nonetheless faster to respond on primed trials—that activation of relations at the highest level of abstraction on this task was driven (at least in large part) by transient, "bottom–up" spreading activation in a distributed semantic network. We predict, by contrast, that relational similarity would be detected for both primed and unprimed word pairs on an explicit analogical task, in which abstract representations must be consciously brought to mind.

Of note, we found that neural priming was numerically greater in the combined ROI than in any single one. In addition, only the combined ROI exhibited a significant correlation with behavioral priming across participants. This finding indicates that behavior reflects the pooled input from all regions that contain information regarding second-order relations. In addition, our data are consistent with the idea that a distributed network, rather than a single region, represents abstract concepts (Binder, 2016).

Several limitations of the current study are factors that likely would have worked against us, reducing our estimate of relational similarity in the brain. The sample size was relatively small, and we observed substantial individual differences in behavioral or neural priming; indeed, several participants exhibited no priming. Furthermore, the first six fMRI participants were not given catch trials—although there is no evidence that they performed the task differently from the rest of the sample. Finally, our stimulus set has not been independently validated, and there may be overlap between whole–part and place–thing relations, which would have reduced the degree of correspondence between our relational RDM and the fMRI data. The fact that we nevertheless detected neural similarity, behavioral priming, neural priming, and a relation between behavioral and neural priming is promising—and warrants a larger-scale and more thorough investigation.

Conclusion

This study provides novel insights into how a form of abstract concept is represented in the brain. First, it reveals that abstract relations can be represented in a way that transcends unrelated, concrete concepts. Second, it localizes these representations to the semantic memory network rather than the relational reasoning network. Third, it suggests that they can be brought to mind momentarily without explicit effort. Finally, it shows a relationship between behavioral priming and neural priming across the semantic memory network. More broadly, this work serves as an important foundation for future work testing the nature of these representations during high-level cognitive tasks. Furthermore, it could provide insights into which approaches are most likely to succeed or fail to optimize reasoning in the service of learning or pointing to best practices for teaching novel concepts, for example, in science and engineering courses by indicating when and how explicitly drawing analogies is helpful (Begolli & Richland, 2016; Vendetti et al., 2015; Gick & Holyoak, 1983). Knowing how abstract concepts can be readily brought to mind will allow researchers to better understand the ability to reason and the ability to learn new concepts in both structured and unstructured environments, both as children and as lifelong learners.

Reprint requests should be sent to Silvia A. Bunge, Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94720, or via e-mail: sbunge@berkeley.edu.

Note

1. Supplementary materials for this paper can be retrieved here: https://www.researchgate.net/publication/344637805 _WangBunge_JOCN_SupplementaryMaterialspdf.

Author Contributions

Gowri Swamy: Investigation; Writing - original draft.

REFERENCES

Alfred, K. L., Connolly, A. C., Cetron, J. S., & Kraemer, D. J. M. (2020). Mental models use common neural spatial structure for spatial and abstract content. *Communications Biology*, *3*, 17. **DOI:** https://doi.org/10.1038/s42003-019-0740-8, **PMID:** 31925291, **PMCID:** PMC6952387

Alfred, K. L., Connolly, A. C., & Kraemer, D. J. M. (2018). Putting the pieces together: Generating a novel representational space through deductive reasoning. *Neuroimage*, *183*, 99–111. **DOI:** https://doi.org/10.1016/j.neuroimage.2018.07.062, **PMID:** 30081195

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. **DOI:** https://doi.org/10.1146/annurev .psych.59.103006.093639, **PMID:** 17705682

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*, 194–213. **DOI:** https://doi.org/10.1037/edu0000056

Bejar, I. I., Chaffin, R., & Embretson, S. (1991). *Cognitive and psychometric analysis of analogical problem solving*. New York: Springer. **DOI:** https://doi.org/10.1007/978-1-4613-9690-1

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*, 1096–1108. **DOI:** https://doi.org/10.3758/s13423-015-0909-1, **PMID:** 27294428

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*, 527–536. **DOI:** https://doi.org/10.1016/j.tics.2011.10.001, **PMID:** 22001867, **PMCID:** PMC3350748

Brownell, W. A., & Hendrickson, G. (1950). How children learn information, concepts, and generalizations. In N. B. Henry (Ed.), *The forty-ninth yearbook of the National Society for the Study of Education, Part 1: Learning and instruction* (pp. 92–128). Chicago: University of Chicago Press. **DOI:** https://doi.org/10.1037/11789-004

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

Cai, M. B., Schuck, N. W., Pillow, J. W., & Niv, Y. (2019). Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. *PLoS Computational Biology*, *15*, e1006299. **DOI:** https://doi.org/10.1371/journal.pcbi.1006299, **PMID:** 31125335, **PMCID:** PMC6553797

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

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, *14*, 1136–1149. **DOI:** https://doi.org/10.1006/nimg.2001.0922, **PMID:** 11697945

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., et al. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*, 2608–2618. **DOI:** https://doi.org/10.1523/JNEUROSCI.5547-11.2012, **PMID:** 22357845, **PMCID:** PMC3532035

Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, *35*, 39–49. **DOI:** https://doi.org/10.3758/BF03195940, **PMID:** 17533878

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980. **DOI:** https://doi.org/10.1016/j.neuroimage.2006.01.021, **PMID:** 16530430

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*, 18906–18916. **DOI:** https://doi.org/10.1523/JNEUROSCI.3809-13.2013, **PMID:** 24285896, **PMCID:** PMC3852350

Dimsdale-Zucker, H. R., & Ranganath, C. (2018). Representational similarity analyses: A practical guide for functional MRI applications. In D. Manahan-Vaughan (Ed.), *Handbook of behavioral neuroscience* (Vol. 28, pp. 509–525). San Diego, CA: Elsevier. **DOI:** https://doi.org/10.1016/B978-0-12-812028-6.00027-6

Dunbar, K. N. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 313–334). Cambridge, MA: MIT Press.

Dunst, C. J., Bruder, M. B., Trivette, C. M., Raab, M., & McLean, M. (2001). Natural learning opportunities for infants, toddlers, and preschoolers. *Young Exceptional Children*, *4*, 18–25. **DOI:** https://doi.org/10.1177/109625060100400303

Estes, Z. (2003). Attributive and relational processes in nominal combination. *Journal of Memory and Language*, *48*, 304–319. **DOI:** https://doi.org/10.1016/S0749-596X(02)00507-7

Estes, Z., & Jones, L. L. (2006). Priming via relational similarity: A copper horse is faster when seen through a glass eye. *Journal of Memory and Language*, *55*, 89–101. **DOI:** https://doi.org/10.1016/j.jml.2006.01.004

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170. **DOI:** https://doi.org/10.1207/s15516709cog0702_3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. **DOI:** https://doi.org/10.1016/0010-0285(83)90002-6

Green, A. E., Fugelsang, J. A., & Dunbar, K. N. (2006). Automatic activation of categorical and abstract analogical relations in analogical reasoning. *Memory & Cognition*, *34*, 1414–1421. **DOI:** https://doi.org/10.3758/BF03195906, **PMID:** 17263066

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

Hobeika, L., Diard-Detoeuf, C., Garcin, B., Levy, R., & Volle, E. (2016). General and specialized brain correlates for analogical reasoning: A meta-analysis of functional imaging studies. *Human Brain Mapping*, *37*, 1953–1969. **DOI:** https://doi.org/10.1002/hbm.23149, **PMID:** 27012301, **PMCID:** PMC6867453

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

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., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. **DOI:** https://doi.org/10.3389/neuro.06.004.2008, **PMID:** 19104670, **PMCID:** PMC2605405

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*, 195–208. **DOI:** https://doi.org/10.1207/s15516709cog0402_4

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42–55. **DOI:** https://doi.org/10.1038/nrn.2016.150, **PMID:** 27881854

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766. **DOI:** https://doi.org/10.1016/j.jesp.2013.03.013

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy*

of Sciences, U.S.A., 116, 4176–4181. **DOI:** https://doi.org/10.1073/pnas.1814779116, **PMID:** 30770443, **PMCID:** PMC6410800

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 373, 20170137. **DOI:** https://doi.org/10.1098/rstb.2017.0137, **PMID:** 29915005, **PMCID:** PMC6015821

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., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems* (Vol. 2, pp. 3111–3119). Lake Tahoe, NV: Curran Associates, Inc.

Moore, K. S., Yi, D.-J., & Chun, M. (2013). The effect of attention on repetition suppression and multivoxel pattern similarity. *Journal of Cognitive Neuroscience*, 25, 1305–1314. **DOI:** https://doi.org/10.1162/jocn_a_00387, **PMID:** 23489143

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, 59, 2636–2643. **DOI:** https://doi.org/10.1016/j.neuroimage.2011.08.076, **PMID:** 21924359, **PMCID:** PMC3251697

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 217–248). Burlington, Canada: Academic Press. **DOI:** https://doi.org/10.1016/B978-0-12-385527-5.00007-3

Pernet, C. R., Wilcox, R. R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, 3, 606. **DOI:** https://doi.org/10.3389/fpsyg.2012.00606, **PMID:** 23335907, **PMCID:** PMC3541537

Popov, V., & Hristova, P. (2015). Unintentional and efficient relational priming. *Memory & Cognition*, 43, 866–878. **DOI:** https://doi.org/10.3758/s13421-015-0514-6, **PMID:** 25754399

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Repetition suppression and semantic enhancement: An investigation of the neural correlates of priming. *Neuropsychologia*, 44, 2284–2295. **DOI:** https://doi.org/10.1016/j.neuropsychologia.2006.05.017, **PMID:** 16806317

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 177–192. **DOI:** https://doi.org/10.1002/wcs.1336, **PMID:** 26263071

Slotnick, S. D. (2017). Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. *Cognitive Neuroscience*, 8, 150–155. **DOI:** https://doi.org/10.1080/17588928.2017.1319350, **PMID:** 28403749

Spellman, B. A., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory & Cognition*, 29, 383–393. **DOI:** https://doi.org/10.3758/BF03196389, **PMID:** 11407415

van Maanen, L., Forstmann, B. U., Keuken, M. C., Wagenmakers, E.-J., & Heathcote, A. (2016). The impact of MRI scanner environment on perceptual decision-making. *Behavior Research Methods*, 48, 184–200. **DOI:** https://doi.org/10.3758/s13428-015-0563-6, **PMID:** 25701105, **PMCID:** PMC4819926

Vargas, R., & Just, M. A. (2020). Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. *Cerebral Cortex*, 30, 2157–2166. **DOI:** https://doi.org/10.1093/cercor/bhz229, **PMID:** 31665238

Vendetti, M. S., & Bunge, S. A. (2014). Evolutionary and developmental changes in the lateral frontoparietal network: A little goes a long way for higher-level cognition. *Neuron*, 84, 906–917. **DOI:** https://doi.org/10.1016/j.neuron.2014.09.0.35, **PMID:** 25475185, **PMCID:** PMC4527542

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, 9, 100–106. **DOI:** https://doi.org/10.1111/mbe.12080

Wang, W.-C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2018). Neural basis of goal-driven changes in knowledge activation. *European Journal of Neuroscience*, 48, 3389–3396. **DOI:** https://doi.org/10.1111/ejn.14196, **PMID:** 30290029, **PMCID:** PMC6350035

Wang, W.-C., Wing, E. A., Murphy, D. L. K., Luber, B. M., Lisanby, S. H., Cabeza, R., et al. (2018). Excitatory TMS modulates memory representations. *Cognitive Neuroscience*, 9, 151–166. **DOI:** https://doi.org/10.1080/17588928.2018.1512482, **PMID:** 30124357, **PMCID:** PMC6340811

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., et al. (2018). Organizational principles of abstract words in the human brain. *Cerebral Cortex*, 28, 4305–4318. **DOI:** https://doi.org/10.1093/cercor/bhx283, **PMID:** 29186345

Wendelken, C., & Bunge, S. A. (2010). Transitive inference: Distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *Journal of Cognitive Neuroscience*, 22, 837–847. **DOI:** https://doi.org/10.1162/jocn.2009.21226, **PMID:** 19320546, **PMCID:** PMC2858584

Wendelken, C., Chung, D., & Bunge, S. A. (2012). Rostrolateral prefrontal cortex: Domain-general or domain-sensitive? *Human Brain Mapping*, 33, 1952–1963. **DOI:** https://doi.org/10.1002/hbm.21336, **PMID:** 21834102, **PMCID:** PMC3984972

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

Whitaker, K. J., Vendetti, M. S., Wendelken, C., & Bunge, S. A. (2018). Neuroscientific insights into the development of analogical reasoning. *Developmental Science*, 21, e12531. **DOI:** https://doi.org/10.1111/desc.12531, **PMID:** 28295877, **PMCID:** PMC5887920

Wilcox, R. (2004). Inferences based on a skipped correlation coefficient. *Journal of Applied Statistics*, 31, 131–143. **DOI:** https://doi.org/10.1080/0266476032000148821

Wodka, E. L., Simmonds, D. J., Mahone, E. M., & Mostofsky, S. H. (2009). Moderate variability in stimulus presentation improves motor response control. *Journal of Clinical and Experimental Neuropsychology*, 31, 483–488. **DOI:** https://doi.org/10.1080/13803390802272036, **PMID:** 18686112, **PMCID:** PMC2892264

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