23 The Neural Representation of Concrete and Abstract Concepts

Robert Vargas and Marcel Adam Just

Introduction

Although the study of concept knowledge has long been of interest in psychology and philosophy, it is only in the past two decades that it has been possible to characterize the neural implementation of concept knowledge. With the use of neuroimaging technology, it has become possible to ask previously unanswerable questions about the representation of concepts, such as the semantic composition of a concept in its brain representation. In particular, it has become possible to uncover some of the fundamental dimensions of representation that characterize several important domains of concepts.

Much of the recent research has been done with fMRI to predict and localize various concept representations and discover the semantic properties that underlie them. Commonly used experimental designs in this research area present single words or pictures of objects, measure the resulting activation pattern in multiple brain locations, and develop a mapping between the topographically distributed activation pattern and the semantic representation of the concept. The primary research topics concerning concept representations pertain to three issues: The composition of concept representations; the neurally defined underlying semantic dimensions; and the relation between neuroimaging findings and cognitive and psycholinguistic findings. It is these types of relationships between cortical function and meaning representation that allow us to understand more about both the way knowledge is organized in the human brain and the functional role that various brain systems play in representing the knowledge.

Concepts are often qualitatively different from one another with regard to their perceptual grounding. As a result, one area of research has largely focused on the neural representations of concrete object concepts. However, as imaging technology and analytic techniques continue to improve, the neural representations of seemingly ethereal, abstract concepts such as *ethics* and *truth* have recently become a topic of increasing interest. In addition to the

Due to a production issue, this chapter appears as the final chapter; it was intended to be in Part II.

interest in such highly abstract concepts, recent research has also investigated hybrids between concrete and abstract concepts such as emotions, physics concepts, and social concepts. These hybrid concepts are not directly perceptually grounded but they can nevertheless be experienced. This chapter provides an overview of contemporary neuroimaging research examining the neural instantiation of concrete concepts, abstract concepts, and concepts that fall somewhere in between, which we call hybrid concepts.

Contemporary Approaches to Analyzing Concept Representations

Univariate-Based Analyses

The initial approach of task-related fMRI imaging was to measure the difference in activation for a class of stimuli (such as a semantic category, like houses) relative to a "rest" condition. At each 3-dimensional volume element in the brain (a voxel), a general linear regression model (GLM) is fit to relate the occurrence of the stimuli to the increase in activation relative to the rest condition. The result is a beta weight whose magnitude reflects the degree of condition-relevant activation in each voxel. This approach proves useful for investigating the involvement of cortical regions whose activation systematically increases or decreases relative to rest for a specific mental activity. However, with this voxel-wise univariate approach, complex relations between the activation in different brain regions within a network are often not apparent (Kriegeskorte, Goebel, & Bandettini, 2006; Mur, Bandettini, & Kriegeskorte, 2009). Moreover, treating each voxel independently of the others misses the fact that the activation pattern corresponding to a concept consists of a set of co-activating voxels that may or may not be proximal to each other. Nevertheless, the univariate approach was successful in identifying which brain regions were activated in response to a given class of concepts.

Multivariate Pattern Analysis (MVPA)

The advent of higher-resolution imaging analyses aided in shifting the research focus from identifying the cortical regions involved in the representation of concepts to focusing on the coordinated activation across a network of brain regions or subregions (Haxby et al., 2001; Haynes & Rees, 2006). Instead of assessing the activation evoked by a class of concepts in terms of individual voxels in various brain regions considered independently of each other, multivariate analyses treated the activating voxels in conjunction with each other, as multiple dependent variables. Multivariate pattern analysis (MVPA) is graphically illustrated in Figure 23.1. MVPA refers to a family of analyses designed to take into account the multivariate relationships among the voxels

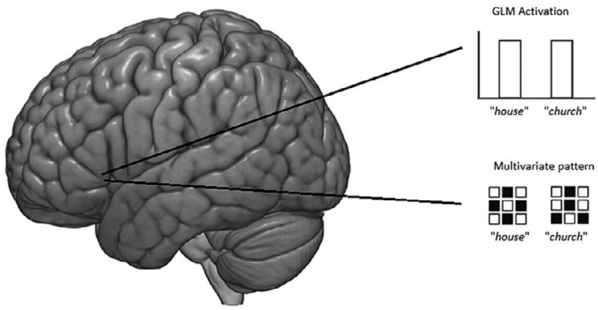

Figure 23.1 *Conceptual schematic showing differences between GLM activation-based approaches and pattern-oriented MVPA, where the same number of voxels activate (shown as dark voxels) for two concepts but the spatial pattern of the activated voxels differs.*

that represent various concepts. Some of the most common analyses for investigating concept representations include: 1) Representational Similarity Analysis (RSA), which enables *comparison* of the multivariate activation patterns of different concepts; 2) Factor Analysis or Principle Components Analysis (PCA), which enables discovery of the lower-dimensional structure of distributed patterns of activation; 3) Predictive Modeling, which enables assessment of various postulated interpretations of underlying semantic structures by predicting activation patterns of concepts; and 4) Encoding Models, which enable quantitative assessment of various organizational structures hypothesized to drive the activation. These techniques tend to answer somewhat different questions.

Representational Similarity Analysis (RSA)

RSA is often used to measure the similarity (or dissimilarity) of representational structures of various individual concepts or categories of concepts. The representation of a concept or a category of concepts can be defined as the evoked activation levels of some set of voxels. These activation patterns can be computed with respect to all of the voxels in the whole cortex but are often restricted to the voxels in semantically relevant regions. The most common technique is to redefine the representation of a concept from being an activation pattern to a similarity pattern with respect to the other concepts in the set (Kriegeskorte, Mur, & Bandettini, 2008a). For example, the neural representation of a concept like *robin* can be thought of in terms of its similarities to a set of other birds. This approach makes it possible to compare various brain subsystems in terms of the types of information they represent, and thus to characterize the processing characteristics of each subsystem. For example, RSA has been used to demonstrate the similarities in the visuospatial subsystems of humans and monkeys in the representations of visually depicted objects (Kriegeskorte et al., 2008b). The strength of this approach is its higher level of abstraction of the neural representation of concepts, representing them in terms of their relations (similarities) to other concepts. The cost of this approach is its limited focus on the representation of the properties of individual concepts.

Extracting Dimensions of Semantics (Factor Analysis / PCA)

Factor analysis and PCA are used to extract neurally meaningful dimensions from high-dimensional activation patterns. "Neurally meaningful" refers to a subset of concepts systematically evoking activation from a subset of relevant voxels. For example, concrete objects that entail interaction with parts of the human body (such as hand tools) evoke activation in motor and pre-motor areas, such that a neural dimension of body–object interaction emerges (Just, Cherkassky, Aryal, & Mitchell, 2010). This approach focuses on dimensions that are shared by some concepts and de-emphasizes the differences among the concepts that share the dimension. The regions corresponding to the dimension can be localized to particular brain areas (by noting the factor loadings of various clusters of voxels).

After the dimension reduction procedure finds a dimension and the items and voxels associated with it, the dimension requires interpretation. The source of the interpretation often comes from past knowledge of the functional roles of the regions involved and the nature of the items strongly associated with the dimension. For example, if hand tools obtain the highest factor scores on some factor, then that factor might plausibly be interpreted as a body–object interaction factor. (The items' factor scores indicate the strength of the association between the items and the factor). One approach to assessing an interpretation of a dimension (such as a body–object interaction dimension in this example) is to first obtain ratings of the salience of the postulated dimension, say body–object interaction, to each of the items from an independent group of participants. For example, the raters may be asked to rate the degree to which a concept, such as *pliers*, is related to the hypothesized dimension *body–object interaction* (Just et al., 2010). Then the correlation between the behavioral ratings and the activation-derived factor scores of the items provides a measure of how well the interpretation of the dimension fits the activation data. This technique has been used to extract and interpret semantically meaningful dimensions underlying the representations of both concrete nouns and abstract concepts (Just et al., 2010; Vargas & Just, 2019).

Predictive Modeling

The goal of a predictive modeling procedure is to assess whether the activation pattern of a concept that was left out of the modeling can be predicted with reasonable accuracy, given some theoretical basis. The prediction process starts by generating a hypothesis about the underlying factor or dimension (which is based on how the items are ordered by their factor scores and on the locations of the voxels with high factor loadings). Then a linear regression model is used to define the mapping between the salience ratings of all but one item and the activation levels evoked by those items in factor-related locations (voxel clusters with high factor loadings in factor analyses that excluded the participant in question). The mapping is defined for all of the underlying factors. Then the activation prediction for the left-out item is generated by applying the mappings for all of the factors to the ratings of the left-out item. This process is repeated, each time leaving out a different item, generating an activation prediction for all of the items. Activation predictions for each concept can be made within each participant and then averaged over participants. The accuracy of the predictions provides converging evidence for the interpretation of the neurosemantic factors. Unlike correlations between behavioral ratings and factor scores for items, this approach develops a mapping that is generative or predictive, applying to items uninvolved in the modeling.

Hypothesis-Driven Encoding Modeling

Encoding models provide another more general way to test whether a hypothesized semantic organization structure is capable of explaining the activation data for some set of concepts. A first step in the modeling is the specification of a theoretically plausible feature set that is hypothesized to account for the relationship between a stimulus set and the corresponding evoked activation patterns (Naselaris, Kay, Nishimoto, & Gallant, 2011). For example, the co-occurrence of noun concepts with verbs in a large text corpus may account for the relationship between individual concepts and activation patterns for those concepts, say in a regression model. The resulting beta-weights from the regression model quantify the degree to which each feature determines the relationship between the stimuli and neural activity (Mitchell et al., 2008). The ability of this mapping to generalize to novel concepts, either in activation space or in feature space, provides a quantitative assessment of the plausibility of the hypothesized relation. This approach is especially useful for representations that are less clearly mapped in the brain, such as abstract concepts, enabling an evaluation of the neural plausibility of theories of abstract concept representation (Wang et al., 2018).

More recently, encoding models have been used with semantic vectors, a feature structure constructed by extracting information from the co-occurrence of words in a large text corpus, to serve as a basis for predictions of large-scale sets of concept representations (Pereira et al., 2018). Encoding models have also been used to measure the ability for theoretically-derived semantic feature structures to explain neural activation data for sentences (Yang, Wang, Bailer, Cherkassky, & Just, 2017). Encoding models are a flexible tool that allow for the quantitative evaluation of the ability of theoretically motivated feature structures to account for brain activation patterns.

Neurosemantic Structure of Concrete Object Representations

Object concepts are the most perceptually driven of concept representations. Consequently, the neural representation of object concepts is fairly well understood because the neural organization of low-level perceptual information is well understood (Grill-Spector & Malach, 2004; Martin, 2007). Haxby et al. (2001) showed that pictures of different objects could be related to each other based on their pattern of activation in the visuospatial pathway, specifically in the fusiform face area (FFA) and parahippocampal place area (PPA). Patterns of activation in these regions were distinguishable in terms of the object categories being represented (i.e., *faces* vs. *houses*). It seems clear that a substantial part of concrete object representations consists of the representation of their perceptual properties.

Moreover, it has been possible to determine the sequence in which various types of perceptual information becomes activated as the thought of a concrete object emerges. Recent MEG research has shown that the temporal trajectory of the neural activation for object representations starts with low-level visual properties such as image complexity, which begins to be activated about 75 ms after stimulus onset in the early bilateral occipital cortex. Later, at 80–120 ms, information concerning more complex categorically defined shapes (e.g., *has eyes*, *has four legs*) begins to be activated along the left ventral temporal cortex and anterior temporal regions (Clarke, Taylor, Devereux, Randall, & Tyler, 2013). The early onset object representation suggests that coarse categorical distinctions between objects are rapidly represented along a left-hemispheric feed-forward neural pipeline. After this initial representation is generated, more complex semantic features take form through recurrent activation and the integration of more distributed cortical systems at 200–300 ms. This temporal trajectory from simple to complex information suggests a cumulating pipeline designed to construct meaning from distributed semantic features.

Beyond the understanding that concrete object representations are based in large part on the objects' perceptual properties, several interesting questions remain, such as how the differing perceptual properties of an object are integrated in the object representation and what semantic properties underlie the organization of the representations of differing objects.

Hub-and-Spoke Model of Feature Integration in Concept Representations

Any individual concept representation is thought to be composed of a network of semantic features (Collins & Loftus, 1975). Connections to more similar (closer) semantic representations are more likely and easier to come to mind than more distal ones. The anterior temporal lobe (ATL), sometimes referred to as the *convergence zone* or *hub*, has been credited with incorporating individual semantic features of concepts (the spokes, in this analogy) into an integrated representation of that concept (Meyer & Damasio, 2009). Recent fMRI research suggests that this integration of semantic features in the brain is localized to the ATL. One study showed that combining color-related activation coded in the right V4 region of the occipital cortex and shape-related activation coded in the lateral occipital cortex (LOC) allowed visual objects to be distinguished in the ATL (Coutanche & Thompson-Schill, 2015). Although the ATL has also been shown to activate for abstract concepts (Hoffman 2016), a study similar to Coutanche and Thompson-Schill (2015) has yet to be conducted showing that individual abstract concepts can be decoded from ATL based on their composite semantic features. In sum, the ATL is thought to act as a cognitive mechanism that integrates perceptual and verbal (i.e., concrete and abstract) information comprising the representation of a concept (Lambon Ralph, 2014).

Semantic Dimensions of Concrete Concepts

Contemporary research into concrete object concepts has progressed beyond the focus on perceptual aspects of concept representations and begun to examine higher-level semantic properties of concrete concept representations. This approach generally utilizes dimension reduction techniques such as factor analysis, first on an individual participant level then at the group level, to investigate semantic dimensions that are present in the neural representations across individuals (Just, Cherkassky, Buchweitz, Keller, & Mitchell, 2014). This dimension reduction approach applied to a set of activation patterns has the advantage of discovering neurally driven dimensions of meaning rather than imposing a previously hypothesized semantic organization.

Just et al. (2010) utilized this approach to uncover three semantic dimensions underlying the representation of 60 words referring to concrete nouns (e.g., *hammer*, *apple*). Specifically, they found that these 60 concrete concepts could be characterized by the way they relate to *eating*, *manipulation* (or *body–object interaction*), and *shelter* (or *enclosure*). Moreover, each of these dimensions was associated with a small set of cortical regions. The *shelter* dimension was associated with activation in regions of bilateral parahippocampal place area, bilateral precuneus, and left inferior frontal gyrus. The *manipulation* dimension was associated with activation in regions of left supramarginal gyrus and left pre- and post-central gyrus (the participants were right-handed). The *eating* dimension was associated with activation in regions of the left inferior and middle frontal gyrus and left inferior temporal gyrus. These results indicate the beginnings of a biologically plausible basis set for concrete nouns and highlight semantic properties beyond a visuospatial domain.

Other research has sought to discover semantic dimensions of non-word or picture concept representations using a different approach. Principal Components Analysis (PCA) was applied to the activation evoked by 1800 object and action concepts shown in short movie clips (Nishimoto et al., 2011). This approach sub-divided the brain based on the similarities of the activation patterns among the concepts to their co-occurrence with a large text corpus. This technique was also applied to the activation patterns evoked by natural continuous speech (Huth, De Heer, Griffiths, Theunissen, & Gallant, 2016). Both the video clip and the natural-speech studies related neural activation similarities to corpus co-occurrence information to locate semantically consistent regions within the cerebral cortex based on domain-specific information. This parcellation approach associated the activation of various regions and semantic categories with individual concepts. The 12 interpretable semantic categories from the PCA were: *mental* (e.g., *asleep*); *emotional* (e.g., *despised*); *social* (e.g., *child*); *communal* (e.g., *schools*); *professional* (e.g., *meetings*); *violent* (e.g., *lethal*); *temporal* (e.g., *minute*); *abstract* (e.g., *natural*); *locational* (e.g., *stadium*); *numeric* (e.g., *four*); *tactile* (e.g., *fingers*); and *visual* (e.g., *yellow*). Aside from the format of stimulus presentation, the notable distinction between the dimension reduction approaches in Just et al. (2010) and Huth et al. (2016) was that Huth et al. generated semantic dimensions based on the mapping between activation and co-occurrence, while Just et al. generated dimensions from the activation patterns. The exploration of the underlying dimensions of concrete concepts helps provide a basis for the semantic organization of perceptible concepts beyond basic visuospatial properties.

Neurosemantic Signatures of Abstract Concepts

The representations of abstract concepts, such as *ethics* and *law*, are neurally and qualitatively distinct from those of concrete concepts. Abstract concepts, by definition, have no direct link to perception, with the exception of some form of symbolic representation (e.g., lady justice holding a scale to represent the concept of *law* or *justice*). The conventional view of abstractness portrays it as an absence of a perceptual basis, that is, the opposite of concreteness (Barsalou, 1999; 2003; Brysbaert, Warriner, & Kuperman, 2014; Wang, Conder, Blitzer, & Shinkareva, 2010). Although it is easy to define abstract concepts such as those lacking concreteness, this definition does not describe the psychological or neurocognitive properties and mechanisms of abstract concepts.

Concrete and abstract concepts generally evoke different activation patterns, as a meta-analysis showed (Wang et al., 2010). This meta-analysis indicated that the two types of concepts differ in their activation in areas related to verbal processing, particularly the left inferior frontal gyrus (LIFG).

Abstract concepts elicited greater activation than concrete concepts in such verbal processing areas. By contrast, concrete concepts elicited greater activation than abstract concepts in visuospatial processing (precuneus, posterior cingulate, and fusiform gyrus). This meta-analysis was limited to univariate comparisons of categories of concepts and did not have access to the activation patterns evoked by individual concepts. This limitation potentially overlooks nuanced distinctions in the representational structure. Univariate contrasts potentially overlook critical relationships across neural states and neural regions (Mur et al., 2009). Through the use of MVPA techniques, more recent studies have begun to examine the underlying semantic structure of sets of abstract concepts. The next section focuses on various imaging studies examining the neural activation patterns associated with abstract concepts and explores the possible semantic structures that are specific to abstract concepts.

Neurosemantic Dimensions of Abstract Meaning

As in the case of concrete concepts, the semantic dimensions underlying abstract concept categories can be identified from their activation patterns. One of the first attempts to decode the semantic content of abstract semantic information was conducted by Anderson, Kiela, Clark, and Poesio (2017). A set of individual concepts that belonged to various taxonomic categories (*tools*, *locations*, *social roles*, *events*, *communications*, and *attributes*) were decoded from their activation patterns. Whether a concept belonged to one of two abstract semantic categories (i.e., *Law* or *Music*) was also decoded from the activation patterns of individual concepts. Although these abstract semantic categories could be decoded based on their activation patterns, the localization of this dissociation is unclear.

Neurally-based semantic dimensions underlying abstract concepts differ from the dimensions underlying concrete concepts. Vargas and Just (2019) investigated the fMRI activation patterns of 28 abstract concepts (e.g., *ethics*, *truth*, *spirituality*) focusing on individual concept representation and the relationship between the activation profiles of these concept representations.

Factor analyses of the activation patterns evoked by the stimulus set revealed three underlying semantic dimensions. These dimensions corresponded to 1) the degree to which a concept was *Verbally Represented*, 2) whether a concept was *External* (or *Internal*) to the individual, and 3) whether the concept contained *Social Content*. The Verbal Representation dimension was present across all participants and was the most salient of the semantic dimensions. Concepts with large positive factor scores for this factor included *compliment*, *faith*, and *ethics*, while concepts with large negative scores for this factor included *gravity*, *force*, and *acceleration.* The former three concepts seem far less perceptual than the latter three. For the Externality factor, a concept that is external is one that requires the representation of the world outside oneself and the relative non-involvement of one's own state. An internal concept is one that involves the representation of the self. At one extreme of the dimension lie concepts that are external to the self (e.g., *causality*, *sacrilege*, and *deity*). At the other extreme lie concepts that are internal to the participant (e.g. *spirituality* and *sadness*). The last semantic dimension was interpreted to correspond to Social Content. The concepts at one extreme of the dimension included *pride*, *gossip*, and *equality*, while the concepts at the other extreme included *heat*, *necessity*, and *multiplication*. Together these semantic dimensions underlie the neural representations of the 28 abstract concepts.

One surprising finding was that the regions associated with the Verbal Representation dimension were the same as those found in the meta-analysis conducted by Wang et al. (2010) that contrasted the activation between concrete and abstract concepts. Activation in the LIFG (a region clearly involved in verbal processing) was evoked by concepts such as *faith* and *truth*, while the left supramarginal gyrus (LSMG) and left lateral occipital complex (LOC), both of which are involved in different aspects of visuospatial processing, were associated with concepts such as *gravity* and *heat*.

Moreover, the output of the factor analysis (i.e., factor scores) for the Verbal Representation factor also suggested that the abstractness of the neural patterning in these regions for an individual concept is represented as a point on a continuum between language systems and perceptual processing systems. This interpretation corresponds to the intuition that abstractness is not a binary construct but rather a gradient-like translation of a concept into a more verbal encoding. This conclusion is somewhat surprising given that the set of 28 concepts are all qualitatively abstract, in that they have no direct perceptual referent. The amount of activation in LIFG evoked by a given abstract concept corresponds to its Verbal Representation factor score.

These results raise an interesting theoretical and psychological question regarding the role of neural language systems, particularly LIFG, in the verbal representation of abstract concepts. That is, what does it mean, neurally and psychologically, for an abstract concept to be verbally represented?

Abstract Concepts as Verbal Representations

What does it mean for an abstract concept to be represented in regions involved in verbal processing and to evoke activation in the LIFG? When the LIFG is artificially lesioned through the repeated use of transcranial magnetic stimulation (TMS), healthy participants show a 150 ms slower response time for comprehending abstract concepts (e.g., *chance*) (Hoffman, Jefferies, & Lambon Ralph, 2010). This same TMS-based lesioning procedure showed no influence in the amount of time needed to respond to concrete concepts (e.g., *apple*). However, these differences in the impact of TMS were nullified when the abstract concepts were presented within a context (e.g., "*You don't stand a chance*"). These results suggest that the abstractness of a concept is dependent on whether it requires integration of meaning across multiple contexts (Crutch & Warrington 2005; 2010; Hoffman 2016; Hayes & Kraemer, 2017). Moreover, the LIFG seems to be involved in the context-dependent integration of meaning.

Given that LIFG appears to be involved in the contextualization of the meaning of abstract concepts (Hoffman et al., 2010) and that the magnitude of activation in LIFG is directly proportional to the degree that it is verbally represented (Vargas & Just, 2019), taken together these results suggest that the activation in LIFG reflects the magnitude of mental activity required to contextualize the meaning of a lexical concept. LIFG has been shown to elicit greater activation for sentence-level representations as compared to word-level concepts (Xu, Kemeny, Park, Frattali, & Braun, 2005). It may be the case that the central cognitive mechanism underlying the neural activation in LIFG represents the integration of meaning across multiple representations in order to form a new representation that is a product of its components. That is, the components of meaning of *apple* require less computation (in LIFG) to generate a composite representation than the concept of *chance.* Also, providing a context for *chance*, as in "*You don't stand a chance*", reduced the cognitive workload by providing a more explicit version of its meaning. A similar mechanism can account for the greater activation in LIFG for sentences than for individual words, because constructing a sentence-level representation requires combining the meanings of individual concept representations in a mutually context-constraining way.

As previously discussed, another region involved in the integrating of meaning for concepts is the anterior temporal lobe (ATL). ATL has been implicated in the integration of semantic features to form a composite representation of object concepts (Coutanche & Thompson-Schill, 2015). However, unlike LIFG, ATL does not appear to differentiate between abstract concepts that vary based on the degree that they are verbally represented (as defined by their factor scores in Vargas & Just (2019)).

In sum, the integration of abstract concept representations with other concepts in a sentence seems to require additional computation. However, it is unclear whether these integrating computations are processing some episodic contexts (as suggested by the results of Hoffman et al., 2010), or some specific concept representations, or use some more general amodal representational format.

Hybrid Concepts: Neither Completely Concrete nor Completely Abstract

Hybrid concepts are concepts that can be experienced directly but require additional processing beyond the five basic perceptual faculties to be evoked. These concepts do not neatly fit within the dichotomy of concrete vs. abstract. For example, the concept *envy* cannot be tasted, seen, heard, smelled, or touched, but it inarguably can be experienced as an internal event which could have perceptual repercussions (e.g., feeling lethargic, crying). We propose that *envy* and other concepts referring to psychological states are hybrid.

The view of embodied cognition (Barsalou, 1999) expands upon the definition of perception beyond our five basic perceptual faculties to also include the experiences of proprioception and emotions. Hybrid concepts fall outside the strict realm of the sensory-perceptual but within the realm of psychological experience as described by embodiment theory (e.g., proprioception and emotion). Emotions, physics, and social concepts are not usually defined exclusively with respect to their concreteness/abstractness but serve as excellent exemplars of hybrid concepts in that they can be perceptually experienced without evoking any of our five senses directly. Additionally, the neural understanding of the semantic underpinning of hybrid concepts is not well understood. The following three sections describe research investigating the neurosemantic organization of hybrid concepts, specifically the neurosemantic organization of emotions, physics concepts, and social concepts.

Neurosemantic Dimensions of Meaning Underlying Emotions Concepts

Meta-analyses of activation contrasts investigating emotion concepts reveal six functional networks (Kober et al., 2008) including limbic regions (i.e., amygdala, hypothalamus, and thalamus), areas related to top-down executive control function (i.e., dorsal lateral prefrontal cortex), the processing of autobiographical information (i.e., posterior cingulate cortex) (Klasen, Kenworthy, Mathiak, Kircher, & Mathiak, 2011), visual association regions, and subregions within the motor cortex (Phan, Wager, Taylor, & Liberzon, 2002). These networks suggest that emotion representations partially involve cognitive functions related to more complex perceptual functioning (i.e., motion and visual association). Furthermore, the involvement of regions related to top-down executive functioning and regions related to the processing of autobiographical information suggest that emotion concepts recruit cognitive faculties for not only basic perceptual representations but also for higher-ordered cognitive functions. Although these findings identify regions involved in emotion representation and processing, they do not provide insight into how different emotions are neurally distinguished.

Recent MVPA analyses examining the neural representations of emotion concepts have provided insight into the way the representations of different emotions are neurally organized. Kassam, Markey, Cherkassky, Loewenstein, and Just (2013) examined the evoked neural activation patterns of 18 emotion concepts such as *happiness*, *pride*, *envy*, and *sadness*. The participants in this study didn't just think about the meaning of a presented emotion word, they tried to evoke the emotion in themselves at that moment. Factor analyses of the activation profiles for the 18 emotion concepts followed by a predictive model to validate the interpretations revealed three underlying dimensions of meaning. The underlying semantic dimensions organize these emotions concepts according to the *valence* of the emotion (positive or negative), its degree of *arousal* (*fury* vs. *annoyance*) and degree of *social involvement* (i.e., whether another person is included in the representation, as is the case for *envy* but not necessarily so for *sadness*). Each of these dimensions of representation were found to correspond to activation distributed across several cortical regions. The brain locations associated with the *valence* of an emotion concept included the right medial prefrontal cortex, left hippocampus, right putamen, and the cerebellum. The brain locations associated with the *arousal* dimension included the right caudate and left anterior cingulum. Finally, brain locations associated with *sociality* included the bilateral cingulum and somatosensory regions. Both univariate and multivariate approaches provided neural evidence for the involvement of perceptual and higher-cognitive faculties. The multivariate analyses provided additional insight into the dimensions along which the individual emotion concepts are differentiated from each other. So even though emotions are very different from object concepts, the principles underlying their neural representations are rather similar to those of other types of concepts.

Neurosemantic Dimensions of Meaning Underlying Physics Concepts

Research investigating the neural representation of physics concepts suggests their neural organization somewhat reflects the physical world they refer to, such as the movements or interactions of objects. Mason and Just (2016) investigated the neural activation patterns of 30 elementary physics concepts (e.g., *acceleration*, *centripetal force*, *diffraction*, *light*, *refraction*). Factor analyses of the activation patterns evoked by the 30 concepts revealed four underlying semantic dimensions. These dimensions were *periodicity* (typified by words such as *wavelength*, *radio waves*, *frequency*), *causal-motion/visualization* (e.g., *centripetal force*, *torque*, *displacement*), *energy flow* (*electric field*, *light*, *direct*, *current*, *sound waves*, and *heat transfer*), and *algebraic/equation representation* (*velocity*, *acceleration*, and *heat transfer*) which are associated with familiar equations.

The regions associated with each semantic dimension provide insight into the underlying cognitive role of the region. The *periodicity* dimension was associated with dorsal premotor cortex, somatosensory cortex, bilateral parietal regions, and the left intraparietal sulcus. These regions have been shown to activate for rhythmic finger tapping (Chen, Zatorre, & Penhune, 2006). The *causal-motion/visualization* dimension was associated with the left intraparietal sulcus, left middle frontal gyrus, parahippocampus, and occipital-temporal-parietal junction. These regions have been shown to be involved in attributing causality to the interactions between objects and data (Fugelsang & Dunbar, 2005; Fugelsang, Roser, Corballis, Gazzaniga, & Dunbar, 2005). The *algebraic/equations* dimension includes the precuneus, left intraparietal sulcus, left inferior frontal gyrus, and occipital lobe. These regions have been implicated in the executive processing and integration of visuospatial and linguistic information in calculation (Benn, Zheng, Wilkinson, Siegal, & Varley, 2012) and more general arithmetic processing. The regions associated with *energy flow* were middle temporal and inferior frontal regions. In the context of physics concepts, these regions are attributed with representing the visual information associated with abstract concepts (Mason & Just, 2016). Together, these results suggest that the neural representations of physics concepts, many of them developed only a few hundred years ago, draw on the human brain's ancient ability to perceive and represent physical objects and events.

Neurosemantic Dimensions of Meaning Underlying Social Concepts

Research comparing the neural representation of social concepts between healthy controls and individuals with high-functioning autism has revealed three semantic dimensions involved in the neural representations of social interactions (Just et al., 2014). Participants in this study thought about the representations of eight verbs describing social interactions (*compliment*, *insult*, *adore*, *hate*, *hug*, *kick*, *encourage*, and *humiliate*) considered from the perspective of either the agent or recipient of the action. Factor analyses of neural activation profiled for these 16 concept–role combinations revealed semantic dimensions associated with *self-related* cognition (*hate* in the agent role and *humiliate* in the recipient role), *social valence* (*adore* and *compliment*), and *accessibility/familiarity* relating to the ease or difficulty of semantic access.

The *self* dimension was associated with activation in the posterior cingulate: An area commonly implicated in the processing of autobiographical information. The *social valence* factor included the caudate and putamen for both controls and individuals with autism. The *accessibility/familiarity* factor included regions that are part of the default mode network, particularly middle cingulate, right angular gyrus, and right superior medial frontal.

Because this study involved a comparison between young adult healthy controls and participants with high-functioning ASD, it provided an important glimpse into how a psychiatric or neurological condition can systematically alter the way a certain class of concepts is thought about. The use of fMRI neuroimaging allows the precise measurement of how a given concept is neurally represented, and specify precisely how a condition like ASD can alter the representation. The interesting finding was that the members of the two participant groups could be very accurately distinguished by their neural representations of these social interaction concepts. More specifically, the ASD group showed a lack of a *self* dimension, showing little activation in the regions associated with the *self* dimension in the healthy control group. The findings suggest that when the ASD participants thought about a concept like *hug*, it involved very little thought of themselves. By contrast, the control group thought about themselves when thinking about what *hug* means. Thus, the assessment of neural representations of various classes of concepts has the potential to identify the presence and the nature of concept alterations in psychiatric or neurological conditions. The neurosemantic architecture of hybrid concepts (as exemplified by emotions, physics, and social concepts) suggests these concepts relate us with the external world (e.g., causal-motion visualization dimension for physics concepts or self/other for social concepts). Moreover, the neural activation associated with magnitudes of perceptual experience are also captured by the neural representations (e.g., degree of arousal with emotion concepts). Taken together, these results suggest that hybrid concepts are composed, in part, of perceptual states that translate our perceptual world into various mental states.

Commonality of Individual Concrete and Abstract Concepts across People

The commonality across participants of the neurally-defined dimensions underlying various semantic domains foreshadows one of the most interesting findings concerning the neural representations of individual concepts. The surprising finding is that the neural representations of all the concepts studied so far are rather similar across people. This section focuses on the commonality of individual concept representations across individuals. The general approach to quantitatively evaluating the commonality of individual concept representations is to train a machine learning classifier on the labeled activation data of all but one participant for a given set of concepts, and then to classify or make predictions concerning the concept representations of the left-out individual. In a cross-validation protocol, this process is repeated with a different person left out on each iteration. The accuracies of the predictions are then averaged across iterations. This averaged accuracy measures the commonality of a set of concept representations.

This approach has shown that there is considerable commonality of the neural representations of concepts across healthy participants. The commonality was present for concrete, abstract, and hybrid concepts. Decoding accuracies across participants were high and approximately equivalent for concrete, abstract, and hybrid concepts (i.e., mean rank accuracy =.72 for concrete concepts (Just et al., 2010); .74 for abstract concepts (Vargas & Just, 2019); .71 for physics concepts (Mason & Just, 2016); .7 for emotion concepts (Kassam et al., 2013); and .77 for social concepts (Just et al., 2014).

Although a large proportion of the concepts in a brain reading study are accurately predicted across participants, there are always a few items at the negative tail of the accuracy distribution, and it would be interesting to know if the items with lower across-participant commonalities had some distinguishing properties. In a study of sentence decoding across three languages (Portuguese, Mandarin, and English), Yang et al. (2017) found lower across-language, across participant decoding accuracies for concepts that are more abstract and related to social and mental activities (e.g., *happy*, *negotiation*, *artist*). They attributed this lower degree of commonality across languages of such items to some abstract and socially-related concept domains being more culturally-determined.

For the set of 28 abstract concepts presented in the Vargas and Just (2019) study, the concepts which were more prototypically abstract (e.g., *sacrilege* and *contract*) were somewhat less accurately predicted across participants than concepts that tend to be more hybrid (e.g., *force* and *pride*). However, there were a number of exceptions to this trend. For example, concepts such as *anger* and *gossip* were less well predicted than others across participants (although still with an above-chance accuracy), and these concepts tended to be highly instantiable. By contrast, concepts such as *necessity* and *causality*, which are highly verbally represented, were more accurately predicted across participants.

Relating Neuroimaging Findings and Corpus Co-occurrence Measures

One particular class of encoding models, as was previously discussed, attempts to relate neural representations to some well-defined feature set. Defining the meaning of a concept in some computationally tractable way has long been a challenge, and it is relevant here because it has the potential to be systematically related to the neural representation of the concept. One of the early answers to this challenge suggested that concepts can be characterized in terms of the concepts with which they co-occur in some large text corpus (Landauer & Dumais, 1997). The lower dimensions (about 300) of a large co-occurrence matrix produce a *semantic vector* representation of the words in the corpus (Pennington, Socher, & Manning, 2014; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). The method of deriving this lower dimensional feature space can vary, depending on the specific approach. The utility of semantic vector representations comes from their convenience in natural language processing applications. But can the semantic vector representation of a concept like *apple* be informative about the neural representation of *apple*?

The semantic vector representations can be used as the predictive basis of an encoding model. Predicted images can be generated from the learned mapping relating brain activation data from a matrix containing semantic vectors. This learned mapping can then be used to generate predicted brain images for concepts with no previously collected data (Mitchell et al., 2008). This approach provides the basis for generating a set of concept representations which can then be explored for its semantic properties (Pereira et al., 2018). Moreover, it enables the study of many more concept representations than can easily be acquired in time- and cost-limited fMRI studies. However, it is unclear whether encoding models based on semantic vector representations illuminate the difference between concrete and abstract concepts representations.

Co-occurrence structures have also been used to evaluate the neural instantiation of the associative theories of abstract concept representations. Wang et al. (2018) utilized RSA to compare the organizational structures of 360 abstract concept representations by examining the representational structure of fMRI activation patterns across the whole brain and concept co-occurrence properties in a large corpus. The goal was to show that each of these viable organization principles is instantiated uniquely within the brain.

Co-occurrence properties represent the theoretical view that abstract concepts are represented in terms of their association with other concepts. Their results showed that the relationship between co-occurrence representations and brain activity for 360 abstract concepts was largely left lateralized and seemed to uniquely activate areas traditionally associated with language processing such as left lateral temporal, inferior parietal, and inferior frontal regions.

Conclusion

The understanding of how concepts are represented in the human brain has advanced significantly based on innovations in imaging technology and multivariate machine learning techniques. One new insight concerns how human and self-centric concept representations are neurally structured. No dictionary definition has specified how a hammer is to be wielded, and yet that is an important part of how it is neurally represented. Thus, part of the neural representation of a physical object specifies how our bodies interact with the object (Hauk & Pulvermüller, 2004; Just et al., 2010). Part of the neural representation of *gossip* specifies a social interaction. The concept of *spirituality* evokes self-reflection. Thus, this insight is that many neural representations of concepts contain human-centric information in addition to semantic information.

A second insight concerns the dependence of abstract concepts on the verbal representations of other concepts. Representing the meaning of abstract concepts may require a greater integration of meaning across multiple other concept representations than is the case for concrete concepts. Abstract concepts evoke activation in cortical regions associated with language processing, particularly LIFG, which may reflect the neurocomputational demand for this increased integration of meaning.

A third insight is that the semantic components of a neural representation of a concept consist of the representations within various neural subsystems, such as the motor system, the social processing system, and the visual system. These neural subsystems constitute the neural indexing or organizational system.

A fourth insight concerns the remarkable degree of commonality of neural representations across people and languages. Although concept representations phenomenologically seem very individualized, the neural representations indicate very substantial commonality, while still leaving room for some individuality. The commonality probably arises from the commonality of human brain structures and their capabilities, and from commonalities in our environment. We all have a motor system for controlling our hands, and all apples have a similar shape, so our neural representations of holding an apple are similar.

A fifth insight is that the principles regarding the neural representations of physical objects extend without much modification to more concrete and hybrid concepts. Although it is easy to see why the concept of *apple* is similarly neurally represented in all of us, it is more surprising that an emotion like *anger* evokes a very similar activation pattern in all of us. Moreover, even abstract concepts like *ethics* have a systematic neural representation that is similar across people.

Although there is much more to human thought than the representation of concepts, these representations constitute an important set of building blocks from which thoughts are constructed. The neuroimaging of these concept representations reveals several of their important properties as well as hints as to how they might combine to form more complex thoughts.

Acknowledgments

This research was supported by the Office of Naval Research Grant N00014–16-1-2694.

References

Anderson, A. J., Kiela, D., Clark, S., & Poesio, M. (2017). Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns. *Transactions of the Association for Computational Linguistics*, *5*, 17–30.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22* (4), 577–660.

Barsalou, L. W. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1435), 1177–1187. doi.org/10.1098/rstb.2003.1319

Benn, Y., Zheng, Y., Wilkinson, I. D., Siegal, M., & Varley, R. (2012). Language in calculation: A core mechanism? *Neuropsychologia*, *50*(1), 1–10. https://doi.org/10.1016/j.neuropsychologia.2011.09.045

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *NeuroImage*, *32*(4), 1771–1781. https://doi.org/10.1016/j.neuroimage.2006.04.207

Clarke, A., Taylor, K. I., Devereux, B., Randall, B., & Tyler, L. K. (2013). From perception to conception: How meaningful objects are processed over time. *Cerebral Cortex*, *23*(1), 187–197. https://doi.org/10.1093/cercor/bhs002

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Coutanche, M. N., & Thompson-Schill, S. L. (2015). Creating concepts from converging features in human cortex. *Cerebral Cortex*, *25*(9), 2584–2593. https://doi.org/10.1093/cercor/bhu057

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(3), 615–627. https://doi.org/10.1093/brain/awh349

Crutch, S. J., & Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: Complementary semantic interference and facilitation effects. *Cognitive Neuropsychology*, *27*(1), 46–71. https://doi.org/10.1080/02643294.2010.491359

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the Association for Information Science and Technology*, *41*(6), 391–407.

Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43*(8), 1204–1213. https://doi.org/10.1016/j.neuropsychologia.2004.10.012

Fugelsang, J. A., Roser, M. E., Corballis, P. M., Gazzaniga, M. S., & Dunbar, K. N. (2005). Brain mechanisms underlying perceptual causality. *Cognitive Brain Research*, *24*(1), 41–47. https://doi.org/10.1016/j.cogbrainres.2004.12.001

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*(1), 649–677. https://doi.org/10.1146/annurev.neuro.27.070203.144220

Hauk, O., & Pulvermüller, F. (2004) Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping*, *21*(3), 191–201. DOI: 10.1002/hbm.10157

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex, *Science*, *293*(5539), 2425–2430. doi: 10.1126/science.1063736.

Hayes, J. C., & Kraemer, D. J. M. (2017). Grounded understanding of abstract concepts: The case of STEM learning. *Cognitive Research: Principles and Implications*, *2*(1), 7. https://doi.org/10.1186/s41235-016-0046-z

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hoffman, P. (2016). The meaning of "life" and other abstract words: Insights from neuropsychology. *Journal of Neuropsychology*, *10*(2), 317–343. https://doi.org/10.1111/jnp.12065

Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2010). Ventrolateral prefrontal cortex plays an executive regulation role in comprehension of abstract words: Convergent neuropsychological and repetitive TMS evidence. *Journal of Neuroscience*, *30*(46), 15450–15456. https://doi.org/10.1523/JNEUROSCI.3783-10.2010

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PLoS ONE*, *5*(1), e8622. https://doi.org/10.1371/journal.pone.0008622

Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A., & Mitchell, T. M. (2014). Identifying autism from neural representations of social interactions: Neurocognitive markers of autism. *PLoS ONE*, *9*(12), e113879. https://doi.org/10.1371/journal.pone.0113879

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying emotions on the basis of neural activation. *PLoS ONE*, *8*(6), e66032. https://doi.org/10.1371/journal.pone.0066032

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal representation of emotions. *Journal of Neuroscience*, *31*(38), 13635–13643. https://doi.org/10.1523/JNEUROSCI.2833-11.2011

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion. *NeuroImage*, *42*(2), 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008a). Representational similarity analysis – Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(11), 1–28. https://doi.org/10.3389/neuro.06.004.2008

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

Lambon Ralph, M. A. (2014). Neurocognitive insights on conceptual knowledge and its breakdown. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120392. https://doi.org/10.1098/rstb.2012.0392

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

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

Mason, R. A., & Just, M. A. (2016). Neural representations of physics concepts. *Psychological Science*, *27*(6), 904–913. https://doi.org/10.1177/0956797616641941

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, *32*(7), 376–382. https://doi.org/10.1016/j.tins.2009.04.002

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191–1195. https://doi.org/10.1126/science.1152876

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI – An introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109. https://doi.org/10.1093/scan/nsn044

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, *21*(19), 1641–1646. https://doi.org/10.1016/j.cub.2011.08.031

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1532–1543.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(1), 963. https://doi.org/10.1038/s41467-018-03068-4

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Vargas, R., & Just, M. A. (2019). Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. *Cerebral Cortex*, *30*(4), 2157–2166. https://doi.org/10.1093/cercor/bhz229

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

Wang, X., Wu, W., Ling, Z., Xu, Y., Fang, Y., Wang, X., ... Bi, Y. (2018). Organizational principles of abstract words in the human brain. *Cerebral Cortex*, *28*(12), 4305–4318. https://doi.org/10.1093/cercor/bhx283

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *NeuroImage*, *25*(3), 1002–1015. https://doi.org/10.1016/j.neuroimage.2004.12.013

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *NeuroImage*, *146*, 658–666. https://doi.org/10.1016/j.neuroimage.2016.10.029