Barsalou, L. W. (2017) What does semantic tiling of the cortex tell us about semantics? *Neuropsychologia*, 105, pp. 18-38. (doi:10.1016/j.neuropsychologia.2017.04.011)

What Does Semantic Tiling of the Cortex Tell us about Semantics?

Lawrence W. Barsalou

Institute of Neuroscience and Psychology
School of Psychology
University of Glasgow

4 April 2017

Abstract

Recent use of voxel-wise modeling in cognitive neuroscience suggests that semantic maps tile the cortex. Although this impressive research establishes distributed cortical areas active during the conceptual processing that underlies semantics, it tells us little about the nature of this processing. While mapping concepts between Marr's computational and implementation levels to support neural encoding and decoding, this approach ignores Marr's algorithmic level, central for understanding the mechanisms that implement cognition, in general, and conceptual processing, in particular. Following decades of research in cognitive science and neuroscience, what do we know so far about the representation and processing mechanisms that implement conceptual abilities? Most basically, much is known about the mechanisms associated with: (1) features and frame representations, (2) grounded, abstract, and linguistic representations, (3) knowledge-based inference, (4) concept composition, and (5) conceptual flexibility. Rather than explaining these fundamental representation and processing mechanisms, semantic tiles simply provide a trace of their activity over a relatively short time period within a specific learning context. Establishing the mechanisms that implement conceptual processing in the brain will require more than mapping it to cortical (and sub-cortical) activity, with process models from cognitive science likely to play central roles in specifying the intervening mechanisms. More generally, neuroscience will not achieve its basic goals until it establishes algorithmic-level mechanisms that contribute essential explanations to how the brain works, going beyond simply establishing the brain areas that respond to various task conditions.

Keywords: semantics; conceptual processing; neural encoding and decoding; multi-voxel pattern analysis; explanatory levels; cognitive mechanisms

1. Introduction

In a recent article, Huth, de Heer, Griffiths, Theunissen, and Gallant (2016) established a semantic atlas on the cortical surface of the human brain. My aim here is to examine this atlas and explore its implications. In the process, I more generally examine the contributions of research in voxel-wise modeling and multi-voxel pattern analysis (MVPA) associated with neural encoding and decoding. Whereas neural encoding focuses on inferring the likely neural activity that results from presented stimuli (e.g., inferring the neural activity associated with hearing "junk food"), neural decoding focuses on inferring likely presented stimuli based on observed patterns of neural activity (e.g., using an observed neural state to predict that "junk food" was just heard). For excellent reviews of neural encoding and decoding, see Naselaris, Kay, Nishimoto, and Gallant (2011), Haxby, Connolly, and Guntupalli (2014), and Weichwald et al. (2015). Although I will be critical of this research in various ways, I admire and respect it in many others. Besides being state-of-the-art in technical sophistication, it is ambitious, has made significant contributions, and has considerable promise. As will also become clear, these methods can be used in ways consistent with the arguments to follow (and have been).

Although I begin with work that addresses semantics in the brain, my arguments more generally address conceptual processing. Whereas "semantics" technically refers to the meanings of linguistic forms, other important forms of conceptual processing occur throughout human cognition (e.g., Barsalou, 2012; McRae & Jones, 2013; Murphy, 2002). In vision and the other senses, for example, conceptual knowledge plays central roles in recognizing objects, scenes, and events, and in subsequently drawing inferences about them (e.g., Henderson & Hollingworth, 1999; Kersten, Mamassian, & Yuille, 2004; Schyns & Oliva, 1999). Similarly, conceptual knowledge plays central roles in the multimodal simulation and imagery that operates across perceptual, cognitive, and motor processes (e.g., Barsalou, 1999, 2008). Although these conceptual processes are undoubtedly shaped by language, they go significantly beyond semantics (Barsalou et al., 1993). Additionally, conceptual processes in non-human species play central roles throughout their perception, cognition, and action in the absence of language (e.g., Barsalou, 2005; Hernnstein, 1984, 1990; Roitblat, Terrace, & Bever, 1984). For these reasons, my focus here is on conceptual processing, while assuming that semantics constitutes a significant subset in humans.

I further assume that a concept is a dynamical distributed network in the brain that represents a category in the environment or experience, and that controls interactions with the category's instances (Barsalou, 2003b, 2009, 2012, 2016a, 2016d, in press). The concept of *bicycle*, for example, represents the category of bicycles in the world and controls interactions with them. Across interactions with bicycles, aggregate information becomes established in relevant neural systems as an agent perceives, evaluates, and interacts with them. On encountering future bicycles, the distributed network acquired becomes active to dynamically produce extensive context-specific inferences that support effective goal-directed action. Within an individual's conceptual system, thousands of concepts represent diverse categories of settings, agents, objects, actions, events, bodily states, mental states, features, relations, and so forth.

1.1. Semantic selectivity and semantic tiling of the cortex

1.1.1. Methods and preliminary analyses.

To assess semantic processing in the brain, Huth et al. (2016) first established semantic selectivity on the cortical surface, and then used it to establish semantic tiling. Previous work from their group has similarly established semantic selectivity across the cortex for a diverse collection of semantic, conceptual, and perceptual materials (e.g., Huth, Nishimoto, Vu, & Gallant, 2012; Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; Nishimoto et al., 2011).

In the work of interest here on neural encoding, Huth et al. (2016) used fMRI to image the BOLD activity of 7 participants as they listened to 10 autobiographical narratives recorded before a live studio audience, each about 10-15 min in duration, over the course of 2 imaging sessions. After transcribing the narratives, Huth et al. constructed a 10,470 word lexicon of the unique words occurring across them. Using methods from computational linguistics for corpus analysis, Huth et al. represented each story word with respect to 985 basis functions. Specifically, they took 985 diverse topic words from Wikipedia's *List of 1000 basic words* (e.g., above, worry, month, tall, mother),[1] and looked up how often each story word occurred within 15 words of the topic word in a large text corpus.

As Figure 1A illustrates, the result of the co-occurrence analysis was a 985-value context vector for each of the 10,470 story words, reflecting how frequently each story word co-occurred with each topic (with these vectors log-transformed and standardized). As much previous work has shown, such context vectors provide indices of lexical semantics (e.g., Baroni & Lenci, 2010; Erk, 2012; Landauer, McNamara, Dennis, & Kintsch, 2013). As the context vectors for two words become increasingly similar, their semantics tend to become increasingly similar as well. By representing the 10,470 story words for their co-occurrence with the same 985 topics, Huth et al. essentially represented story word semantics with a common set of basis functions.

As Figure 1B illustrates, Huth et al. then regressed the BOLD signal in each cortical voxel onto the time series of context vectors across individual story words as they were heard sequentially in the scanner. Because each of the 985 basis functions had a frequency-based value for every story word, its values varied over time across the stories. Of primary interest was establishing voxels whose neural activity tracked the fluctuating values of a given basis function. Whereas some voxels had a positive correlation with a basis function's standardized co-occurrence frequencies, others had a negative relation or no relation. Using machine learning methods, Huth et al. established a regression coefficient for how well each of the 985 basis functions predicted the BOLD activity in each of 10,000+ cortical voxels (on the order of 100,000 coefficients; see Figure 2A). Although four time points along the BOLD function were modeled for each voxel, a combined measure was reported in the central results. Auditory processing of the story words was also modeled, with this variance removed before semantic modeling was performed. The semantic model was subsequently used to predict neural activations for the story words in an eleventh story.[2]

To better understand the information in the coefficient matrix (Figure 2A), Huth et al. submitted the 985 vectors of regression coefficients across the 10,000 best-predicted voxels to principal component analysis. As Figure 2B illustrates, four components explained significant variance at the group level. To the extent that different basis functions behaved similarly across voxels, they loaded on the same component. Notably, these four components only explained about 20% of the variance in the vectors for the 985 basis functions, indicating that much unexplained variance remained. Because additional components were not significant, the remaining variance was relatively unsystematic. Of further interest, the amount of explained variance nearly doubled to about 35% when principle components were extracted for individual participants, indicating that neural activity for the basis functions was individual specific. Rather than being perfectly stable across individuals, semantic processing in the brain often appears to exhibit strong individual differences (e.g., Haxby et al., 2011, 2014; Renoult et al., 2012, 2016).

To interpret the four group principle components, Huth et. al projected the 10,470 story words into the four-dimensional space that the components defined (via the words' context vectors and the basis functions' loadings on components). As a result, a four-valued vector of component scores represented each story word's semantics. After finding the 458 best words in the original space of 10,470 words, this small subset was submitted to cluster analysis, producing the 12 clusters shown on the left of Table 1 (e.g., clusters of visual, tactile, locational, and mental words).

To the right of each cluster label in Table 1 are examples of words that I sampled from each cluster.[3] On the left side of each row are words that strike me as reasonable cluster members. In the visual cluster, for example, it seems reasonable to assume that colour, yellow, stripes, wide, and shaped all have clear visual senses. It's nevertheless worth noting that even these reasonable cluster members form quite a heterogeneous category, including words associated with color, patterning, size, and shape. Typically, accounts of vision and visual semantics are likely to distinguish these sub-categories. More problematically, the words that I have designated as questionable in Table 1 seem to have little relevance for their respective clusters. In the visual cluster, for example, it seems difficult to justify that fur, steel, skull, fielder, cloth, and seal, are visual in the same sense as the reasonable cluster members, given that many other features seem equally, if not more, salient for these words. As can be seen across clusters, each cluster is extremely heterogeneous in the words included, with none appearing to constitute a clearly coherent semantic category. Of further concern, the words shown here are the 458 best words in the four-dimensional component space. It is difficult to imagine how the remaining 10,012 words could be worse fits with the clusters than those classified in Table 1 as questionable, but it seems likely, probably making these clusters even more heterogeneous than illustrated here.

Huth et al. (2016) then used these relatively amorphous clusters to interpret the four group components by assessing where the clusters fell in the four-dimensional space, two components at a time. Figure 2A in Huth et al. illustrates this process (shown here at the top of Figure 3). As can be seen, the social, emotional, violent, and communal clusters were associated with one end of the first component, whereas the tactile, locational, numeric, and visual clusters were associated with the other, suggesting to Huth et al. that this component varied with social vs. non-social semantics (or, alternatively, with animacy). As can similarly be seen, the visual and tactile clusters were associated with one end of the second component, whereas the mental, professional, and temporal clusters were associated with the other, suggesting to Huth et al. that this component varied with perceptual vs. non-perceptual semantics (or, alternatively, with concreteness). Interestingly, Huth et al. (p, 454) concluded that the third and fourth components were uninterpretable. For the third component, the professional, location, and visual clusters were associated with one end, whereas the mental, abstract, and emotional clusters were associated with the other.[4] For the fourth component, the communal, emotional, and social clusters were associated with one end, whereas the violent, tactile, and temporal clusters were associated with the other.

Putting these results in perspective, Huth et al. have essentially shown that word meanings vary in animacy and concreteness, hardly a novel finding. Furthermore, semanticists are likely to find the heterogeneity and amorphousness of both the clusters and the dimensions in these analyses to be not only uninformative but disconcerting. It's not readily apparent that a coherent semantic analysis has been achieved, or that the four component scores representing the story words offer any insight into their semantics. The inability to interpret the third and fourth components underlines these concerns.

1.1.2. Semantic selectivity. Huth et al. (2016) next project the group components onto the cortical surface. As shown in Figure 3 here, Huth et al.'s Figure 2b presents projections of the first three components for one participant in detail, and their Figure 2c presents partial projections for three additional participants. In these figures, as a voxel becomes redder, it becomes more social; as it becomes greener, it becomes more concrete; as it becomes bluer, it has a higher value on the uninterpretable third component (all four components are projected individually for one participant in Huth et al.'s Extended Data Figure 3a). As is apparent, semantic selectivity on the cortical surface varies systematically with respect to these components. It is interesting and impressive that techniques exist for establishing semantic selectivity on the cortical surface in this extensive and precise manner, voxel by voxel.

1.1.3. Semantic tiling and a semantic atlas. Finally, Huth et al. (2016) use the semantic selectivity just described to establish "a dense, tiled map of functionally homogenous brain areas" (p. 455). In this analysis, a semantic tile satisfies two conditions: (1) It is a region of contiguous voxels that share relatively similar values on the four group components, (2) the values in a homogenous region contrast clearly with values in neighboring regions. Huth et al.'s Figure 3c illustrates the semantic tiles established in this analysis (Figure 4 here), with color again exhibiting the most important components underlying a tile's semantics. Gray regions illustrate regions whose semantics were relatively graded and overlapping, such that semantic tiles could not be established.

Huth et al. (2016, p. 457) propose that their semantic tiles constitute "a comprehensive atlas of semantically sensitive areas" that will be "useful for many researchers investigating the neurobiological basis of language." In an online tool, interested readers can examine the contents of this atlas in detail.[5] By clicking on a semantic tile, useful pieces of information become available about the tile and associated analyses.

Problematically, however, interacting with the online tool suggests that the semantic atlas may actually be relatively uninformative for researchers who aim to understand conceptual and semantic processing. In my opinion at least, this atlas doesn't actually constitute a coherent semantic account in any conventional sense of what an atlas is supposed to provide. Consider the red tile in the right temporal-parietal junction labeled LPC R5. The story words most associated with this tile include:

> cousin, murdered, pregnant, pleaded, arrested, refused, son, wife, sister, husband, mother, aunt, asked, daughter, confessed

Rather than constituting a homogenous semantic group, these words come from diverse semantic categories that include relatives, crimes, and communicative acts. Given the dramatic autobiographical narratives from which these story words were drawn, it might appear that this tile processes relatives discussing criminal activity (probably not a general semantic category).

Even more problematically, if one examines individual voxels within a specific tile, they are often associated with even more diverse words, suggesting that the tile doesn't actually establish a homogenous semantic region. Consider three voxels within the tile just discussed, LPC R5. In voxel [18, 75, 33], the words most associated with it include:

> spend, staying, date, last, place, weeks, days, year, month, visit, rent, trip, till, and vacation

In voxel [15, 77, 29], the most associated words include:

> insisted, himself, asked, waited, home, sent, leave, decided, told, arrives, him, promptly, friend's

In voxel [15, 77, 25], the most associated words include:

> parents, murdered, arrives, mother, wife, refused, sister, home, husband, sent, arrive, visit, aunt, leave, house, father, lived, apartment, whereabouts, relatives

Although remote associations can be established between the semantics of these three voxels, it seems questionable that they all belong to a homogenous semantic tile. More generally, it seems questionable that a meaningful and useful semantic atlas has been established on the cortex.

1.2. What have we learned about semantics?

What have we learned about the meaning of a word from Huth et al.'s results? From this perspective, a word's meaning is a distributed pattern of activation across the entire cortex, reflecting how the word's co-occurrence frequencies activate cortical voxels via regression coefficients for basis functions. Although the interaction between co-occurrence frequencies and regression coefficients causes some cortical areas to be more important in representing a word's meaning than others, in principle, the entire cortical surface carries information about the word's semantics (given that every voxel receives predictive information about the word via the shared basis functions projecting to it).

What have we learned about how the brain processes meaning? As we have seen, the semantic selectivity of a voxel reflects the regression coefficients that project to it from the basis functions. From this perspective, semantic processing in the brain results from integrating activations for weighted basis functions within each voxel, which in turn depends on patterns of co-occurrence frequencies across words. As a consequence, a voxel can be characterized in terms of the basis functions and words that activate it most highly.

Once basis functions are reduced to principle components, similar conclusions follow: A word's meaning is represented as a set of component scores on significant components, and a voxel's semantic selectivity is also represented as a set of component scores.

Finally, this approach assumes that particular words can be linked to specific voxels and cortical tiles to produce a semantic atlas. It seems rather peculiar and unusual, however, to assume that specific words are linked to specific voxels in this phrenological manner. Although this assumption is perhaps useful if one's goals are associated with neural encoding and decoding, it is perhaps misguided and misleading if one's goal is to understand how the brain implements semantic processing. Rather than simply being implemented as a pattern of voxels, a word's meaning is more likely to be implemented by mechanisms that represent and process conceptual information.

Although the ability to establish semantic selectivity across the cortical surface for a large text corpus is impressive, it's not clear that researchers who work on semantics and conceptual processing will find these results to offer significant new insights. Other than demonstrating that basis functions and principle components can characterize semantic selectivity across the cortical surface, not much further appears to follow from this work. Furthermore, the considerable heterogeneity and ambiguity associated with interpreting the principle components diminishes the ability to draw useful conclusions about semantics per se, or about how the brain produces semantic processing.[6]

2. Levels of explanation

It is informative to view research on neural encoding and decoding from the perspective of Marr's (1982) classic levels of analysis. According to Marr, an intelligent system can be characterized at three levels of explanation: computational, algorithmic, and implementation. Specifically, the computational level describes the tasks that the system performs, and why it performs them. In the process, the computational level may formally describe relevant stimuli and behavioral responses, along with systematic relations between them (e.g., physical analyses of stimuli and response, behaviorist laws of conditioning, Bayesian models). In turn, the algorithmic level describes the information processing mechanisms within the system that perform the task. From the perspective of the Cognitive Revolution, these are cognitive mechanisms inferred as latent variables from stimulus-response relations (e.g., Lachman, Lachman, & Butterfield, 1979). Although these mechanisms often take the form of representations and processes in classic cognitive models, the distinction between representations and processes becomes blurred in some approaches, such as neural nets and dynamical systems. Finally, the implementation level describes the physical medium that implements information processing mechanisms. A particular set of algorithmic mechanisms, for example, could be implemented in biological tissue (e.g., a human) or in silicon (e.g., a robot). As Marr (1982) further argued, an intelligent system is only understood fully once an *integrated* account across levels exists.

Since Marr's original proposal, his framework has been applied widely in diverse research domains, while continuing to receive considerable attention and development (see Peebles & Cooper, 2015, and the special issue of *Topics in Cognitive Science* that follows). As a result, our understanding of Marr's levels and their interaction continues to evolve. Nevertheless, the general form of Marr's framework remains widely accepted, continuing to play important roles in modern research.

2.1. Stressing the importance of mechanistic accounts at the algorithmic level

Impressive methods associated with big data, crowd sourcing, voxel-wise modeling, and so forth increasingly demonstrate their power and value. Often simultaneously (but not necessarily), attention diminishes to relevant mechanisms at the algorithmic level. As a consequence, researchers continue to remind the community about the importance of algorithmic mechanisms, arguing that the scientific goals of explanation and control cannot be achieved without them.

Schyns, Gosselin, and Smith (2009), for example, noted that cognitive neuroscientists often focus on relations between tasks and neural activity, ignoring the algorithmic mechanisms that produce this activity. Schyns et al. further illustrated how reverse correlation methods combined with automata theory can be used effectively to develop algorithmic accounts of visual perception. Ince et al. (2015) et al. developed these methods further (also see Schyns, van Rijsbergen, & Ince, 2016). An important conclusion from this work is that we do not understand how the brain implements an intelligent activity if we don't understand the mechanisms that produce it. Love (2015), too, argued that mechanistic accounts at the algorithmic level are essential for understanding the brain's operation (for related arguments, see other articles from the same 2015 issue of *Topics in Cognitive Science*).

In an analysis of Bayesian modeling, Jones and Love (2011) similarly argued that Bayesian models typically focus on principles that underlie rational and optimal responses to stimuli, while ignoring process models that implement these principles mechanistically. Jones and Love echoed Marr's (1982) concern that accounts of intelligent behavior are incomplete when they lack an algorithmic account of the underlying mechanisms, and that such accounts play critical roles in understanding cognition. Indeed, mechanistic accounts are typically viewed as essential for explaining phenomena across the sciences (e.g., Bechtel, 2008, 2009; Bechtel & Abrahamsen, 2005; Bechtel & Shagrir, 2015).

2.2. Levels of explanation in neural encoding and decoding

What can we learn about Huth et al.'s (2016) findings, specifically, and about neural encoding and decoding, more generally, from framing them with Marr's explanatory levels? Figure 5a illustrates one insight that results from such framing. From Marr's perspective, neural encoding and decoding focus on mappings between the computational and implemental levels while ignoring the algorithmic level. By presenting people with conceptual stimuli, such as words, and measuring the distributed patterns of voxel activity that follow, it becomes possible to later predict a person's neural activity in response to particular stimuli, and to predict the stimuli they're observing from their neural activity. If the goal is to simply perform neural encoding and decoding in this manner, establishing mappings between the computational and implementation levels can clearly be useful.

Interestingly, neural encoding and decoding do not require the postulation and assessment of algorithmic mechanisms. Because establishing the mapping between conceptual stimuli and neural activity is sufficient for successful performance (to some degree), no need exists for establishing mediating mechanisms at the algorithmic level. As argued later, however, excluding algorithmic mechanisms may significantly limit the success of neural encoding and decoding, whereas including them may accomplish significantly more.

Figure 5B illustrates examples of algorithmic mechanisms that might be useful for increasing the success of neural encoding and decoding. More importantly, however, these mechanisms are likely to play central roles in explaining conceptual and semantic processing in the brain. If we want to understand how the brain produces conceptual and semantic processing, we may well need to include such mechanisms in our accounts. Distributed patterns of neural activity will not be sufficient.

Notably, the methods used in neural encoding and decoding are frequently used to address hypotheses at the algorithmic level (e.g., Peelen & Downing, 2007; Wang et al., 2016; and many others). Nothing intrinsic about the methods associated with voxel-wise modeling precludes using them to understand algorithmic mechanisms. A focus on neural encoding and decoding, however, often draws attention to mappings between the computational and implementation levels, with the algorithmic level dropping out.

2.3. Neurobehaviorism

Interestingly, the Cognitive Revolution in the mid-twentieth century originated in a similar set of issues (Lachman et al., 1979; also see Jones & Love, 2011). At the time, Behaviorists focused on stimuli and responses, largely ignoring the cognitive processes that mediated them. In response, cognitive scientists increasingly articulated theoretical arguments for the importance of mediating processes, with empirical support accumulating rapidly. Since then, the central roles of mechanisms and process models at the algorithmic level have become fundamental constructs across most areas of psychology and cognitive science (e.g., Barsalou, 2016a).

One could argue that an analogous state-of-affairs has developed in neuroscience, with neuroscientists sometimes implicitly adopting what might be called *Neurobehaviorism*. As we have just seen, research on neural encoding and decoding often focuses on the relations between tasks and neural activity, without addressing mediating algorithmic mechanisms. Neural encoding and decoding are hardly isolated examples. Considerable amounts of other neuroscience research often only establish the neural activity associated with a task, without specifying the algorithmic mechanisms responsible. Thus, a potential lack of algorithmic mechanisms not only exists horizontally across the computational level in Behaviorism, but also exists vertically from the computational to the implementation level in Neurobehaviorism.

2.4. Reifying brain states at the implementation level as algorithmic mechanisms

Figure 5C illustrates one potential position that neural encoding and decoding researchers could adopt in responding to concerns about the algorithmic level. Rather than proposing that algorithmic mechanisms aren't important, these researchers could argue that the distributed neural patterns their methods identify at the implementation level constitute mechanisms at the algorithmic level. Regarding Huth et al. (2016), for example, one could argue that the distributed cortical pattern of activation established for a word constitutes its semantic representation. Indeed, this appears to be a basic assumption of their approach and results, at least implicitly.

If so, then several significant issues follow. First, is the distributed cortical representation of a word uniform with no structure, or does it contain parts that enter into larger organizations, with specific parts implementing particular functions? If so, then such accounts add additional constructs at the algorithmic level, going significantly beyond homogenous neural patterns at the implementation level (e.g., Bechtel & Abrahamsen, 2005). Second, what kinds of processes operate on these distributed representations to perform basic conceptual tasks, such as categorization, inference, and concept composition? An atlas of distributed semantic patterns does not offer an account of the processing performed on these patterns to implement conceptual functions. To the extent that processes are added, they further implicate additional algorithmic constructs beyond homogenous neural patterns.

3. Mechanisms of conceptual processing at the algorithmic level

Decades of research have addressed algorithmic mechanisms in conceptual and semantic processing across multiple areas of psychology, across the disciplines of cognitive science, and across relevant areas of cognitive and social neuroscience. As summarized in Figure 5B, the remainder of this article focuses on mechanisms central to explaining conceptual and semantic processing at the algorithmic level. In the next two sections, I address basic representational mechanisms that enjoy widespread empirical support: (1) feature and frame representations, (2) multiple representations of conceptual knowledge. In three subsequent sections, I address important classes of processing mechanisms associated with: (3) knowledge-based inference, (4) conceptual composition, (5) conceptual flexibility. I hasten to add that many other mechanisms are potentially relevant, with different researchers likely to focus on different ones. For each class of mechanisms addressed here, space only allows a brief summary of relevant research. Rather than providing an exhaustive review of each class, I simply try to motivate its importance with classic and representative examples.

3.1. Feature and frame representations

As we saw earlier, one approach to understanding a word's meaning in the brain is that an undifferentiated state of activation across the cortex represents it. Alternatively, researchers have argued for decades that word meanings (and other conceptual structures) are not undifferentiated holistic states, but instead contain a variety of representational elements, including features and frame structure. Furthermore, researchers assume widely that these representational elements operate as mechanisms at the algorithmic level, causally affecting conceptual processing and related behaviors.

3.1.1. Features. Perhaps the first, most natural place to look for the importance of features is in linguistics (e.g., Fromkin, Rodman, & Hyams, 2013). For decades, linguists have employed features at every level of linguistic analysis, from phonetics to semantics. In morphology and semantics, linguists have proposed that fundamental features are essential for distinguishing different classes of word meanings, including animacy (*human* vs. *statue*), gender (*boy* vs. *girl*), number (*bird* vs. *birds*), and so forth. Not only do features like these distinguish significant word classes, they are central for combining words syntactically. When combining a subject with a verb, for example, the subject and verb must often agree in animacy, number, and many other features (e.g., *the woman eats* vs. *the rock *eats*; *it walks* vs. *they *walks*). Indeed, it would be difficult, if not impossible, to characterize basic language structure at any level without features.

Features have been no less central to the study of concepts and lexical meaning in cognitive psychology and psycholinguistics (e.g., Cree & McRae, 2003; Hampton, 1979; McRae, Cree, Seidenberg, & McNorgan, 2005; Rosch & Mervis, 1975; Wu & Barsalou, 2009). When researchers ask participants to generate the features of a concept, they do so rapidly, producing diverse features in the process (e.g., *feathers, wings, flies, and nests* for *BIRDS*). Clearly, a wide variety of features are associated with any reasonably familiar concept in memory. At a minimum, accounts of concepts at the algorithmic level need to explain knowledge of these features and their associations with concepts.[7]

More critically, however, overwhelming evidence indicates that these features play functional roles in conceptual and semantic processing, implicating them as causal mechanisms at the algorithmic level. Consider Garner's (1976) classic filtering task. In the baseline condition, two stimuli that differ on only one feature (e.g., a red square and a blue square) are presented one at a time, repeating in a random order, with the task being to indicate whether the current stimulus is red or blue (via binary responses with the left vs. right hand, respectively). In the critical filtering condition, the task remains the same (discriminate red vs. blue), but now the stimuli can be both squares and circles, such that participants must filter out shape while focusing on color. In general, participants are excellent at filtering features from one another, sometimes exhibiting modest cross-talk between features, while still generally exhibiting excellent filtering performance (e.g., Melara & Marks, 1990). Most importantly, the ability to perform the filtering task implicates features as causal mechanisms at the algorithmic level. By focusing attention on some feature representations in memory and inhibiting others, participants perform with high accuracy. Without functional feature representations in memory, it's difficult to explain how people perform this task.

A wide variety of additional tasks that combine filtering with learning further implicate features as causal mechanisms at the algorithmic level (e.g., Colzato, Van Wouwe, Lavender, & Hommel, 2006; Kirkham, Cruess, & Diamond, 2003; Zelazo, Frye, & Rapus, 1996). In the intra-dimensional shift task, participants first learn which feature from a dimension is currently being rewarded and respond when it's present (e.g., round stimuli), but then must shift to a new feature on the same dimension when it becomes rewarded instead (e.g., triangular stimuli). Similarly, in the inter-dimensional shift task, participants must shift from a feature on one dimension to a feature on another dimension to receive reward (e.g., shifting from round shape to green color). For decades, performance on these tasks has been used to measure developmental stage, intellectual ability, brain damage, and aging, indicating that the representation and processing of features plays central roles in human intelligence.

In traditional conditioning literatures, attending to features during learning is often central to task performance (e.g., Bower & Hilgard, 1981; Domjan, 2014; Mackintosh, 1975; Rescorla & Wagner, 1972). In the classic blocking paradigm, for example, organisms initially learn that a feature predicts reward or punishment, with a second predictive feature being added gradually, as learning the first feature stabilizes. Interestingly, participants typically don't learn that the second feature also predicts the outcome, because the first feature blocks its use. From a cognitive perspective, attention focuses on a feature that produces good performance, while minimizing distraction from other features. Many other classic learning phenomena across species similarly illustrate the central functional roles of features in perception, attention, and learning.

In learning literatures associated specifically with human cognition, features are equally important, often controlled by language. In explicit rule learning paradigms, for example, learners often specify rules as combinations of features (e.g., Trabasso & Bower, 1968). In most category learning paradigms, learning is strongly associated with predictive features that control categorization via attention and reward (e.g., Ashby & Maddox, 2005; Kruschke, 2003; Nosofsky, 1984). Again, it seems difficult to explain conceptual processing without invoking features as fundamentally important causal mechanisms at the algorithmic level.

Some of the most compelling research in this area demonstrates that learning new features has considerable impact on behavior, further implicating their causal roles in cognition. In Schyns and Rodet (1997), for example, participants learned complex visual features that distinguished different categories of (fictional) Martian rocks from one another. In a key comparison, participants learned two features, X and XY, one at a time, with two different groups learning them in opposite orders (i.e., X then XY vs. XY then X). During the critical test phase, participants in the first group successfully used feature Y to categorize Martian rocks, whereas participants in the second group didn't. Whereas the transition from X to XY isolated Y as a functional feature, the transition for XY to X did not. Schyns, Goldstone, and Thibaut (1998) discuss how these and many other findings implicate features as causal mechanisms at the algorithmic level.

In the domain of chick sexing, Biederman and Shiffrar (1987) provided another classic demonstration of how learning new features can have substantial effects on behavior. To segregate male and female chicks at birth, a difficult visual discrimination of their genitalia must be made, with expertise typically requiring much practice. In a one-minute intervention, however, Biederman and Shiffrar presented undergraduates with schematic depictions of male and female genitalia (motivated by geon theory; Biederman, 1987) that produced a 50% improvement in classification accuracy, relative to the pretest baseline (comparable to expert performance on the task). Again, it seems difficult to explain such a large change in performance without postulating the representation of learned features at the algorithmic level that causally affect classification performance.

Finally, the BUBBLES technique offers a powerful method for establishing algorithmic features that control conceptual behavior (e.g., Schyns et al., 2009). As random apertures ("bubbles") expose regions of a visual stimulus (e.g., a face), participants must categorize it in some way (e.g., gender, emotion). Considerable work demonstrates that exposing particular features of faces increases the accuracy of certain categorizations, demonstrating the causal role of features in performing these categorizations effectively (Schyns, Bonnar, & Gosselin, 2002; Smith, Cottrell, Gosselin, & Schyns, 2005). Whereas the eyes are often used to make gender discriminations, the mouth is often used to make emotion discriminations, with the specific patterns of relevant features varying cross-culturally (e.g., Jack, Garrod, Yu, Caldara, & Schyns, 2012; Jack, Caldara, & Schyns, 2012). As much work in this area illustrates, features function as causal mechanisms at the algorithmic level to produce conceptual processing and behavior.

3.1.2. The status of features in neural encoding and decoding research. It might appear that features play central roles in neural encoding and decoding. Perhaps the basis functions often used to represent conceptual stimuli in this research can be viewed as features of the relevant stimuli. Perhaps establishing voxel selectivity to these basis functions can be viewed as establishing these basis functions as features in the brain.

In Huth et al. (2016), for example, every story word was coded with a context vector that represented the word's co-occurrence frequencies across a common set of basis functions (i.e., topic words). Furthermore, the selectivity of every cortical voxel to each basis function was captured in the regression cooefficients that predicted the voxel's BOLD time course. At least since Mitchell et al.'s (2008) pioneering work on neural encoding and decoding, researchers have been using basis functions in this manner to code stimulus meaning and to establish voxel selectivity in the brain (for further examples, see Haxby et al., 2014).

Behavioral research also increasingly uses basis functions to represent concepts and establish relations between them. Crutch, Troche, Reilly, and Ridgway (2013), for example, had participants evaluate 400 words on 12 diverse features (sensation, action, emotion, thought, social interaction, morality, time, space, quantity, and polarity), such that each word was defined by a profile of values across basis functions, thereby establishing it as a point in a high-dimensional space. In the critical analyses, Euclidean distances between words in the high-dimensional space predicted the comprehension deficits of a stroke patient. Troche, Crutch, and Reilly (2014) took a similar approach, using a common set of 12 basis functions to distinguish concrete vs. abstract concepts in a multiple-dimensional space. Finally, Binder et al. (2016) evaluated the meanings of 535 words on 65 basis functions that captured neurally-inspired features (e.g., colour, bright, texture, taste, path, number, time, consequential, self, pleasant, attention). In multiple analyses, they found that representing word meanings as profiles across basis functions (points in a multidimensional space) recovered similarity relations and taxonomic structure within the word set.

Although one might be tempted to conclude that this approach represents the features of concepts as basis functions, it is again instructive to consider Marr's levels. Rather than being functional mechanisms at the algorithmic level, basis functions appear to be technical descriptions of stimuli at the computational (task) level. When co-occurrence vectors are used to represent a word's meaning, these are external descriptions of the stimuli, because they describe how a critical stimulus—a word—correlates with other words in written language (e.g., Huth et al., 2016; Mitchell et al., 2008). Similarly, when rating vectors are used to represent a word's meaning, they describe how a critical stimulus is normed by one group of participants, external to other participant(s) tested in the critical analyses (e.g., Binder et al., 2016; Crutch et al., 2013; Troche et al., 2014). In neither case, has a basis function been shown to be an internal mechanism that operates as a causal feature to control conceptual processing and behavior within an individual's cognitive system.

What about the semantic selectivity of a voxel? Could all the voxels that respond positively to a basis function be viewed as a feature? From Marr's perspective, this selectivity simply shows a correlation between a stimulus descriptor at the computational level and the neural activity it produces at the implementation level. It doesn't necessarily specify a functional feature mechanism that controls conceptual processing and behavior.

To see this, it is informative to explore why basis functions distinguish stimuli so effectively and recover their underlying structure. If basis functions are sufficiently numerous and diverse, they serve as a measurement tool for capturing similarities and differences between concepts. Similar concepts will tend to behave similarly when assessed by a given basis function, whereas different concepts will tend to behave differently. To the extent that basis functions are included that measure important differences between important groups of concepts, they distinguish these groups, enabling the recovery of taxonomic structure. Nevertheless, a basis function such as *colour* in Binder et al. (2016) probably doesn't map directly onto a feature for a concept in the brain at the algrorithmic level, such as *yellow* for *BANANA* (i.e., *yellow* probably plays a more important role than *colour* when *BANANA* is processed conceptually). Instead, the basis function, *colour*, only measures the causal feature, *yellow,* indirectly.

Thus, basis functions don't appear to function as causal mechanistic constructs in the brain. To be effective, they only need to differentiate the representations and processes that operate at the algorithmic level. As a result, an effective set of basis functions could have no direct overlap with the feature mechanisms that represent concepts. Mitchell et al. (2008, Figure 5) offered an informative demonstration of this point. When Mitchel et al. iteratively sampling 25 basis functions randomly from 5,000 possible basis functions, they typically observed neural decoding that was well above chance. Although decoding wasn't as good as when a set of basis functions was hand picked, decoding was nevertheless effective. This important result suggests that basis functions work simply because differences between them allow capturing semantic similarity indirectly, without directly assessing the underlying feature mechanisms that represent concepts at the algorithmic level.

3.1.3. Frame structure. Although features play central roles in conceptual processing, they are not the only representational mechanisms necessary for explaining concepts and meaning. Instead, additional representational mechanisms are necessary, in particular, those associated with frame structure (e.g., Barsalou, 1992, 1999; Fillmore, 1985; Gentner, 1983, 2010; Löbner, 2014). It is not sufficient to characterize a concept as a conjunctive list of binary features (e.g., representing the concept of *FACE* as *eyes* & *mouth* & *nose*). Instead, considerable research across the cognitive sciences illustrates that concepts contain additional structure associated with arguments and values, conceptual relations, and recursion—concepts are not "flat" structures (e.g., Barsalou & Hale, 1993; Fodor & Pylyshyn, 1988; Smolensky, 1990).

In fully representing the concept of *FACE*, for example, recursion is necessary to establish that an *eye* contains an *eyeball*, which contains a *pupil*, which contains an *iris*, etc. Similarly, attribute-value relations are necessary for specifying that *EYE COLOUR* is an attribute that can take values, such as *blue*, *green*, and *brown*. Finally, conceptual relations are required to specify that the eyes align at the same level above the nose, which is above the mouth. Every concept appears to exhibit the basic properties associated with frame structure (Barsalou, 1992; Löbner, 2014), with this structure not being rigid but varying dynamically across contexts (Barsalou, 2003a). Furthermore, when knowledge engineers and researchers have attempted to articulate detailed conceptual and semantic content, they have typically gone beyond feature conjunctions to frame structure (e.g., Lenat, 1995). Although features can be a useful heuristic for representing aspects of a concept in various tasks, features are simply fragments of more complex underlying frame structure (Barsalou & Hale, 1993).

Besides being necessary for representing the detailed structure of knowledge, frame structure has strong empirical support in cognitive science. Consider a classic demonstration from Markman and Gentner (1993). On each baseline trial, the experimenter showed participants two pictures, such as: (1) a delivery man handing a bag to a woman, and (2) a woman feeding a squirrel. The experimenter then pointed to an entity in the first picture (e.g., the woman), and asked what in the second picture went with it. On these trials, participants typically made identity matches, mapping the entity identified in the first picture to the same entity in the second picture (e.g., the woman).

On critical trials, participants were asked to judge the similarity of the two pictures before performing the mapping task. Because similarity judgments often produce deep conceptual processing, Markman and Gentner predicted that participants would activate frame structure to interpret each picture, thereby conceptualizing picture elements as values of frame attributes. On interpreting the first picture described above, for example, participants should activate the *DELIVER* frame, bind the delivery man to the *AGENT* attribute, and bind the woman to the *RECIPIENT* attribute. Similarly, on interpreting the second picture, participants should activate the *FEED* frame, bind the woman to the *AGENT*, and bind the squirrel to the *RECIPIENT*. Most importantly, Markman and Gentner further predicted that using frame structure to interpret the pictures in this manner should change the mapping from the first picture to the second. Rather than performing an identity mapping (i.e., mapping the woman in the first picture to the woman in the second), participants should perform an attribute mapping (i.e., mapping the woman in the first picture to the squirrel in the second, because using frame structure to interpret the pictures caused the woman and the squirrel to both be conceptualized as *RECIPIENTS*). As predicted, argument mappings increased by about 50% following similarity judgments, indicating that frame structure had become active to interpret the pictures.

Many other findings similarly implicate frame structure in conceptual processing and similarity judgments (e.g., Gentner, 1983, 2010; Goldstone, Medin, & Gentner, 1991; Medin, Goldstone, & Gentner, 1990). Most importantly, such findings implicate frame structure as a causal mechanism at the algorithmic level, given its effects on conceptual processing and behavior. Without the presence of frame mechanisms, it is a challenge to explain many basic findings associated with conceptual processing.

3.1.4. Summary. The evidence just reviewed for features and frame structure only begins to cover relevant findings. As we have seen, however, conceptual knowledge appears to contain mechanisms for features and frame structure that play causal roles in conceptual processing. Specifically, these mechanisms often mediate relations between conceptual stimuli (e.g., words, pictures) and diverse conceptual behaviors (e.g., filtering, categorization, mapping). It is highly likely that such mechanisms also mediate between conceptual stimuli and distributed patterns of neural activity. When voxel-wise modeling only establishes mappings between the computational and implementation levels, however, we have no understanding of the mechanistic representations that produced these mappings.

3.2. Multiple representations of conceptual information

Different types of representations appear to implement concepts at the algorithmic level, each having different implications for conceptual processing. Currently, three kinds of representation are receiving significant attention across diverse research communities in neuroscience and cognitive science: (1) grounded representations in the modalities, (2) abstractions in association areas, and (3) distributed linguistic representations in the language system (Barsalou, 2016b).[8] Problematically, undifferentiated holistic patterns of neural activity don't inform which kinds of representation operate in a particular setting, nor do they establish the roles that these representations play in conceptual processing, individually or together.

3.2.1. Grounded representations in the modalities. Much behavioral and neuroscience research demonstrates that the modalities play central roles in grounding conceptual processes (e.g., Barsalou, 2008; Coello & Fischer, 2016a, 2016b; De Vega, Glenberg, & Graesser, 2008; Kemmerer, 2015; Martin, 2007, 2016; Pecher & Zwaan, 2005). When representing information about a concept, the brain reuses modality-specific resources to represent relevant information, including the representation of perceptual features, actions, and internal states (e.g., Anderson, 2010; Barsalou, 2016b). When representing conceptual information about instances in their absence, the brain simulates likely states that the brain would be in if instances were present, thereby providing useful inferences about them. Such simulations may often be unconscious, exhibit bias, vary widely in detail, be context-dependent, and reflect strong biological constraints (e.g., Barsalou, 1999, 2008, 2016b).

In cognitive neuroscience, pioneering research by Alex Martin and his colleagues demonstrated the activation of modality-specific regions during the conceptual processing of tools and animals (e.g., Martin, 2007, 2016). When people represent a tool, for example, the fusiform gyrus becomes active to represent its shape; premotor cortex becomes active to represent action with it; parietal cortex becomes active to represent the trajectory of its manipulation; temporal cortex becomes active to represent its perceived motion. From Martin's theoretical perspective, the conceptual representation of a category emerges from neural activity across the distributed brain regions that process its features. Related to the research reviewed earlier on features, Martin's work establishes the brain areas that implement features likely to play critical roles in conceptual processing at the algorithmic level.

Subsequent research has produced many related findings, demonstrating that the conceptual representation of a feature often reuses relevant modality-specific resources. When conceptual features for color are processed (e.g., *yellow* for *BANANA*), their representations often recruit color processing resources in the visual system (e.g., Hsu, Frankland, & Thompson-Schill, 2012; Martin, 2016; Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Simmons et al., 2007; Wang et al., 2013). When conceptual features for sound are processed (e.g., *loud* for *EXPLOSION*), their representations often recruit processing resources in the auditory system (e.g., Bonner & Grossman, 2012; Hoenig et al., 2011; Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008; Trumpp, Kliese, Hoenig, Haarmeier, & Kiefer, 2013). When conceptual features for the taste and pleasure of foods are processed (e.g., *sweet* for *CHOCOLATE*), their representations often recruit processing resources in the gustatory and reward systems (e.g., Chen, Papies, & Barsalou, 2016; Martin, 2016; Simmons, Martin, & Barsalou, 2005; van der Laan, de Ridder, Viergever, & Smeets, 2011). When conceptual features of actions are processed (e.g., *using the foot* for *KICK*), their representations often recruit processing resources in the motor and spatial systems (e.g., Hauk, Johnsrude, & Pulvermüller, 2004; Kemmerer, 2015; Martin, 2016; Pulvermüller, 2013). When conceptual features for visual motion are processed (e.g., *swaying* for *TREES*), their representations often recruit motion processing resources in the visual system (e.g., Kemmerer, 2015; Martin, 2016; Watson, Cardillo, Ianni, & Chatterjee, 2013).

Given the accumulation of evidence for grounded representations of these features and many others, it appears likely that most, if not all, kinds of feature representations draw on such resources. Indeed, even representations of abstract features such as *mental state*, *magnitude*, and *self* appear grounded (e.g., Barsalou, 2016b; Leshinskaya & Caramazza, 2016; Wilson-Mendenhall, Simmons, Martin, & Barsalou, 2013). As described later in the section on conceptual flexibility, grounded representations do not always become active during conceptual processing, and thus are not obligatory. Importantly, however, modality-specific features are not unique in this regard. No conceptual features appear obligatory during conceptual processing, with all features instead being subject to contextual influence (e.g., Lebois, Wilson-Mendenhall, & Barsalou, 2015; also see section 3.5 here).

3.2.2. Abstractions in association areas.

For some time, evidence has been accumulating that the brain's association areas also contribute to conceptual processing. Evidence for this conclusion comes from establishing the brain areas that represent word meanings after controlling for word orthography and phonology (e.g., subtracting activations for carefully matched pseudo-words from activations for words).

Introducing this methodological approach in classic work, Binder et al. (1999) found that semantic processing activated areas of the default mode network similar to those active during the resting state (including dorsomedial prefrontal cortex, posterior cingulate cortex, and angular gyrus). Much subsequent research has similarly shown that association areas are typically active during conceptual processing. In a meta-analysis across 120 experiments, Binder, Desai, Graves, and Conant (2009) found that semantic processing consistently activated association areas in the parietal lobes (angular gyrus, supramarginal gyrus, posterior cingulate cortex, ventral precuneus), in the frontal lobes (dorsomedial prefrontal cortex, ventromedial orbitofrontal cortex, lateral orbitofrontal cortex, left inferior frontal gyrus), and in the temporal lobes (middle temporal gyrus, left fusiform cortex, left parahippocampal cortex).

Recently, Binder (2016) reviewed additional evidence showing that association areas become more active as the amount of conceptual processing increases. Specifically, the above areas tend to become more active for familiar proper names and high-frequency words than for unfamiliar proper names and low-frequency words. These areas also become more active when highly related and associated concepts are processed together than when less related and associated concepts are processed together. Finally, processing concepts deeply activates these areas more than processing them shallowly. Because association areas become increasingly active as conceptual processing increases, they again appear to be associated with conceptual processing.

Across the evolution of monkeys and primates, association areas have expanded in size, while modality-specific systems have remained relatively constant (e.g., Buckner & Krienen, 2013). This pattern suggests that association areas contribute to the impressive cognitive and social abilities of humans. The further finding that association areas are consistently active during conceptual processing suggests that sophisticated conceptual processing is central to important human abilities.

What specific roles might association areas play in conceptual processing? According to Binder (2016), association areas contain conjunctive neurons that integrate features for a concept across the modalities, thereby implementing abstraction and synthesis of grounded information (also see Damasio, 1989). Simmons and Barsalou (2003) similarly proposed that conjunctive neurons in association areas pattern topographically according to their similarity (as defined by overlapping feature information). More generally, Barsalou (2016b) suggests that association areas implement data compression of multimodal information, where compressions could take a variety of forms, including conjunctive neurons, prototypes, and/or principal components.

The previously-mentioned findings that familiarity, frequency, relatedness, and depth of processing all increase activations in association areas implicates data compression (Binder, 2016). If association areas implement data compression, then as conceptual processing increases, association areas should become more active (i.e., because the amount of compressed data being processed increases).

Fernandino et al. (2016) offered additional evidence for this conclusion. In their experiment, participants processed 900 words for concreteness during fMRI, whose meanings had been rated by other participants for the salience of color, shape, motion, sound, and manipulation. When Fernandino et al. regressed BOLD activity for the words onto each of these five sensory-motor features, they found that the association areas reviewed earlier tracked their salience. As the salience of sensory-motor information increased, activation in association areas grew stronger. Again, if association areas implement data compression, then as a concept contains more data relevant to a sensory-motor feature, association areas should become more active. In a final analysis, Fernandino et al. assessed whether any brain areas tracked all five sensory-motor features together, and found that central areas of the default mode network did so (anterior and posterior regions along the cortical mid-line, bilateral angular gyrus). These results again suggest that association areas represent abstractions of grounded features via some form of data compression.

A related possibility is that the anterior temporal lobes have a special status in representing abstractions associated with concepts. According to hub-and-spoke theories, conjunctive neurons in the anterior temporal lobes integrate grounded features across the modalities (e.g., Lambon Ralph, Sage, Jones, & Mayberry, 2010; Patterson, Nestor, & Rogers, 2007; Reilly, Peelle, Garcia, A., & Crutch, 2016; Rogers & McClelland, 2004). In contrast to other theories that focus on a wider set of association areas, hub-and-spoke theories propose that abstractions for all concepts reside in the anterior temporal lobes. Other researchers propose instead that anterior temporal lobes only contain conjunctive neurons for certain kinds of concepts, such as those for individuals and social cognition (e.g., Binder, 2016; Drane et al., 2008; Martin, 2016; Martin, Simmons, Beauchamp, & Gotts, 2014; cf. Wong & Gallate, 2012).

Still another possibility is that association areas represent abstract features and relations associated with concepts (Barsalou, 2016b; Binder, 2016; Jamrozik, McQuire, Cardillo, & Chatterjee, 2016; Leshinskaya & Caramazza, 2016; Wilson-Mendenhall et al., 2013). From this perspective, abstract features associated with self, mental states, thematic roles, integrated event structure, and so forth are represented in association areas. As these features become active for relevant concepts, association areas become active to represent them.

Much remains to be learned about contributions of association areas to conceptual processing. Clearly, however, they are robustly active during conceptual processing, suggesting that they play central roles.

3.2.3. Distributed linguistic representations. Finally, considerable evidence implicates distributed linguistic representations in conceptual processing (e.g., Andrews, Frank, & Vigliocco, 2014; Andrews, Vigliocco, & Vinson, 2009; Barsalou, 2016b; Connell & Lynott, 2013; Louwerse, 2011; Louwerse & Connell, 2011; Zwaan, 2016). In the spirit of Latent Semantic Analysis and related approaches from computational linguistics (e.g., Baroni & Lenci, 2010; Erk, 2012; Landauer et al., 2013), words associated with a concept provide information about its meaning (e.g., word associates of *BICYCLE*, such as *wheel*, *ride*, *helmet*, and *exercise*). When processing a concept, associated word forms become active, providing useful information about it. Although meanings for these word forms may not become active, the word forms alone (phonological, orthographic) may nevertheless provide useful information for performing the current task.

Consider an empirical finding that illustrates how distributed linguistic representations play causal roles at the algorithmic level. In Solomon and Barsalou (2004), participants were asked to verify parts of objects. On each trial, they received an object word followed by a property word and had to indicate whether the property was a part of the object (e.g., *CAT-claw*, *BATHTUB-drain*). Although all participants received the same 100 true pairs, two different groups received unassociated vs. associated false pairs, respectively, where unassociated false pairs included *PLIERS-river* and *BRIEFCASE-wick*, and associated false pairs included *CAT-litter* and *TABLE-furniture*.

When participants received unassociated false trials, the best predictor of reaction times and errors on *true trials* was the associative strength between object and part words. When the other group of participants received associated trials, however, the best predictor of reaction times and errors on true trials was the size of the part (as the part became larger, performance declined). This pattern of results suggests that distributed linguistic representations controlled performance when false trials were unassociated, whereas grounded visual representations controlled performance when false trials were associated. Specifically, when false trials were unassociated, participants simply needed to determine whether the object and property words were associated or unassociated—information that is available in distributed linguistic representations. When the object and property words were associated, participants responded true; when they were unassociated, they responded false. Because associative strength between the two words perfectly predicted correct performance, participants could use this linguistic information to make decisions. Conversely, when false trials were associated, associative strength between the object and property words no longer predicted correct performance, given that the words for both true and false properties were associated with their respective object words. Instead, participants had to consult another source of knowledge—grounded visual representations—to assess whether the object physically contained the property as a part (explaining why property size became important).

Kan, Barsalou, Solomon, Minor, and Thompson-Schill (2003) replicated this experiment using fMRI, and found that a fusiform area associated with imagining objects only became active with associated false trials, not with unassociated false trials. Again, participants appeared to be using grounded visual representations when associated false trials made using this kind of representation necessary, but used distributed linguistic representations when unassociated false trials made using this kind of representation possible instead. Increasingly, results like these demonstrate that distributed linguistic representations contribute to conceptual processing under relevant conditions.

3.2.4. Summary. As we have seen, different types of representations contribute to conceptual processing. Undoubtedly, they work together rather than independently. Nevertheless, we know relatively little at this point about how they work together. One finding is that distributed linguistic representations appear to operate initially as heuristics, with "deeper" simulations being used when accurate and detailed conceptual knowledge is required (e.g., Barsalou, Santos, Simmons, & Wilson, 2008; Connell & Lynott, 2013; Glaser, 1992; Kan et al., 2003; Paivio, 1986; Solomon & Barsalou, 2004; Zwaan, 2016). A related possibility is that abstractions in association areas are also used as preliminary heuristic representations when deeper representations are not necessary (e.g., Binder, 2016; Patterson et al., 2007). Alternatively, abstractions may primarily serve to index and activate grounded representations (e.g., Damasio, 1989; Simmons & Barsalou, 2003).

Clearly, we have *much* to learn about how these different representational systems interact to produce conceptual processing at the algorithmic level. Successfully addressing this issue is likely to become one of the most important goals in establishing the neural bases of conceptual processing. When, however, voxel-wise modeling only address mappings between the computational and implementation levels, we gain no insight into the specific representational mechanisms that produced the mapping, nor how they did so.

3.3. Knowledge-based inference

Researchers who study conceptual processing often propose that a fundamental purpose of having a conceptual system is to produce useful inferences (e.g., Barsalou, 2012; Murphy, 2002). Consider the conceptual process of categorization. Rather than being an end in itself, categorization is the gateway to knowledge that produces diverse inferences essential for perception, cognition, and action. Once one knows that a perceived entity is an *apple* instead of a *tennis ball*, for example, different understandings of its origins and morphology follow, as well as different actions for using it effectively. "Going beyond the information given" is what makes an active agent intelligent and effective (Bruner, 1973), and this is indeed what conceptual processing is typically all about. Not surprisingly, one of the most central themes in modern cognitive science and neuroscience is the fundamental importance of prediction across intelligent activities (e.g., Bar, 2007, 2009;

Barsalou, 2009; Clark, 2013; Friston, 2010). Conceptual knowledge is typically the source of these predictions.

Problematically, however, mapping concepts at the computational level to neural activity at the implementation level doesn't inform us about inference. Not only do we learn nothing about the mechanisms that implement inference at the algorithmic level, we have no idea whether the neural activations we observe at the implementation level reflect the concepts presented on a task vs. inferences that go beyond them (e.g., in maps of semantic selectivity, such as those in Huth et al., 2016).

3.3.1. Knowledge-based inference during situated action. Effectively pursuing goals during situated action requires extensive supporting inferences. Figure 6 summarizes this important class of inference processes. In memory, thousands of event frames represent the typical content of a particular type of event, including its typical settings, objects, people, agents, goals, bodily states, emotions, actions, and outcomes (Panel A). In the current situation, entities and/or events activate the most relevant event frame in a Bayesian manner (in Panel B, a familiar setting and an object activate the best-fitting event frame, via priors and likelihoods). Once the event frame becomes active, it generates inferences about other likely elements of the situation present or likely to follow (Panel C).

On encountering an apple at a seminar, for example, an event frame for eating becomes active. Besides generating the general inference that an eating event might follow, the event frame's activation generates many specific inferences about the event not yet observable (going beyond the information given). For example, the active event frame might suggest: (1) the goal of consuming the apple, (2) actions useful for consuming it, (3), the bodily state of hunger as a condition for consumption, (4) a pleasant taste and positive emotion during consumption, and (5) hunger reduction and health benefits as outcomes.

Considerable research in cognitive science and neuroscience supports this general account of knowledge-based inference during situated action. As we saw earlier (3.1.3), frames are often viewed as a uniform type of representation that underlies conceptual knowledge, with frames for events playing central roles in integrating conceptual processing (e.g., Barsalou, 1992, 1999; Fillmore, 1985; Gentner, 1983, 2010; Löbner, 2014). From this perspective, event frames in memory are abstracted from relevant populations of events experienced in the world (e.g., Barsalou, 2003b, 2016c, 2016d; Barsalou, Niedenthal, Barbey, & Ruppert, 2003). On later encountering an element of a familiar event, the best-fitting frame becomes active in a Bayesian manner to generate multimodal predictions about what is likely to unfold in the current situation (e.g., Barsalou, 2009, 2011).

Much evidence demonstrates the wide variety of knowledge-based inferences generated from event frames to support situated action. Consider three such inferences. On encountering a familiar setting, a relevant event frame becomes active to generate inferences about likely objects present (e.g., Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman, Rabinowitz, Glass, & Webb, 1974; Palmer, 1975; Yeh & Barsalou, 2006). On encountering a kitchen, for example, one expects to find food and utensils; on encountering a park, one expects to find walking paths and dogs. Conversely, one expects to find certain kinds of objects in specific settings, such as encountering bicycles on roads. Computational models that explain the interaction between bottom-up and top-down processing offer elegant accounts of setting-object inferences (e.g., McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982; Van Dantzig, Raffone, & Hommel, 2011).

Evaluative inferences constitute a second fundamental form of situated inference. On encountering a familiar object or event, the brain immediately produces an evaluation of it. Most basically, evaluative inferences provide information about affective valence: Is the perceived object or event something that is liked or disliked (e.g., Barrett & Bliss-Moreau, 2009; Russell, 2003)? Is the entity or event something that should be approached or avoided (e.g., Carver, 2006)? In what ways is the entity or event self-relevant (e.g., Baumeister, 1998; Northoff et al., 2006)? Diverse behavioral paradigms demonstrate the central roles of evaluative inferences that occur habitually and immediately on encountering familiar entities and events (e.g., De Houwer, Thomas, & Baeyens, 2001; Herring et al., 2013; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). As this work further illustrates, the valence of a stimulus can be conditioned rapidly, later producing evaluative inferences that control behavior. In neuroscience, much is known about the pathways that produce evaluative inferences, with orbitofrontal cortex and the ventral striatum playing central roles (e.g., Berridge, Robinson, & Aldridge, 2009; Rudebeck & Murray, 2014; Wilson, Takahashi, Schoenbaum, & Niv, 2014). On perceiving a familiar object or food, for example, visual processing rapidly projects to orbitofrontal cortex, producing an immediate evaluation of the object (e.g., Chaumon, Kveraga, Barrett, & Bar, 2014; Chen et al., 2016; Lebrecht, Bar, Barrett, & Tarr, 2012).

Affordances constitute a third fundamental form of situated inference. On encountering a familiar object, the brain immediately produces inferences about actions that can be used with it to achieve desirable goals. On encountering a hammer, for example, one's dominant hand may prepare to grasp the handle with a power grip, and then to subsequently swing the hammer, thereby achieving its typical function of pounding a nail. Although the action is often not executed, an implicit simulation of the object occurs that represents inferences about its typical use, which could be potentially useful to undertaking a goal in the current setting, or to understanding another agent's intentions. Since classic behavioral research by Tucker and Ellis (1998), numerous behavioral researchers have demonstrated affordances in diverse ways (for a review, see Caligiore, Borghi, Parisi, & Baldassarre, 2010). Since classic neuroimaging research by Chao and Martin (2000), much neuroscience evidence has similarly accumulated for the generation of motor affordances to objects (e.g., Lewis, 2006).

Many additional kinds of knowledge-based inferences underlie situated action beyond the three just reviewed here (for further examples, see Barsalou, 2016d). Notably, such knowledge-based inferences must have occurred continuously and frequently as Huth et al.'s (2016) participants listened to stories. The voxel-wise modeling performed, however, did not identify the neural activity associated with these inferences, nor establish the mechanisms responsible for producing them.

3.3.2. Event inferences during language comprehension. Psycholinguistics has also produced extensive evidence for knowledge-based inference during event processing, at least since the classic work of Bransford and Johnson (1972, 1973). As people comprehend language, they produce extensive inferences from words, phrases, and sentences that go beyond the linguistic forms presented. On reading that *Mary pounded a nail into the wall*, for example, readers infer that a hammer was probably used and that the nail's orientation was probably horizontal (e.g., Graesser, Singer, & Trabasso, 1994; Stanfield & Zwaan, 2001; Zwaan & Radvansky, 1998).

Research by McRae, Elman, Hare, and their colleagues offers a particularly thorough account of event inferences during language comprehension (e.g., Hare, Jones, Thomson, Kelly, & McRae, 2009; Metusalem et al., 2012). In a typical experiment, participants received a word prime that described some feature of an event. Following a short SOA (e.g., 250 ms), participants received a target word that described another event feature, and then performed a semantic judgment on the word (e.g., evaluating its animacy or concreteness). In some experiments, for example, participants received the name of an event, followed by a typical value of an event attribute (e.g., *TRIAL-judge*). Relative to a matched baseline, the time to evaluate the target was typically facilitated by about 30 to 60 msec. According to these researchers, activating the concept for the word prime activated the associated event frame, together with its attributes and typical values, which then facilitated reading words for these values. Reading *TRIAL*, for example, activated the *TRIAL* event frame, which in turn activated the attribute for *AGENT*, which in turn activated the value for *judge*.

Additionally, these experiments demonstrated that reading the word for one attribute value typical of an event frame propagated activation to correlated attribute values of the same event. Priming *classroom* as the *SETTING* attribute of the *SCHOOL* frame, for example, activated *student* as a value of the *AGENT* attribute. Across numerous experiments in multiple articles, these researchers demonstrated knowledge-based inferences for many of the attributes illustrated in Figure 6.

3.3.3. Summary. As we have seen, knowledge-based inferences make situated action possible. Once something in the current situation activates relevant knowledge, inferences follow that support perceptual anticipations, relevant goals, appropriate actions, and desired outcomes. The constant interplay between perception, knowledge, action, and outcomes produces a continual stream of inferences that make intelligent action effective and efficient.

It is difficult to imagine how this constant interplay works without postulating mechanisms at the algorithmic level. Not only are mechanisms for specific knowledge structures relevant (e.g., event frames), so are numerous mechanisms for activating these structures, generating inferences from them, tracking inference accuracy, making corrections when necessary, and learning from experience. Without accounts of such mechanisms, we know nothing about perhaps the most central role of conceptual processing in intelligence, namely, to support situated action.

3.4. Concept composition

Although concepts and meanings are often studied in isolation for the purpose of experimental control, they rarely occur in isolation when encountered during everyday activity. When processing a concept that represents an element of a perceived scene, it is typically processed together with concepts that represent other aspects of the scene and their integration (3.3.1). When processing the meaning of a word encountered in spoken or written language, it is typically processed together with the meanings of other words in the same phrase, sentence, and text (3.3.2).

The composition of concepts to create larger conceptual representations implicates another central set of mechanisms at the algorithmic level. When perceiving a scene, compositional mechanisms integrate the concepts describing its elements and integration. When understanding language, compositional mechanisms integrate the meanings of relevant linguistic units. Importantly, composition doesn't simply result from storing and activating patterns of co-occurring concepts, but instead reflects productive mechanisms capable of producing infinite compositions, including many novel ones never experienced (e.g., Barsalou, 1999; Fodor & Pylyshyn, 1988). Typically, frames play central relational roles in integrating concepts into larger more complex structures that enable productive composition. Much recent research across the cognitive sciences addresses compositional mechanisms (e.g., Werning, Hinzen, & Machery, 2012; Winter & Hampton, in press).

3.4.1. Integrative priming. Recent findings from Estes and Jones (2009) demonstrate how natural and pervasive composition is in conceptual processing. Using the phenomena of integrative priming, Estes and Jones demonstrated how the activation of a concept immediately generates inferences and processing machinery relevant to combining itself with other concepts present. Because concepts are almost always processed together in compositions, these inclinations to combine should not be surprising.

On the critical trials in Estes and Jones' experiments, participants first received a word prime for 500-2000 msec, and then received a letter string in a lexical decision task. Of primary interest were trials where the critical letter string contained a word that could be integrated with the prime, but that shared no overlapping features with it, nor any measurable association (e.g., when the prime was *FARM*, and the target was *mouse*). Relative to various baselines, Estes and Jones observed significant facilitation while making lexical decisions on these trials, in the range of about 15-50 msec. Notably, these facilitation effects were comparable to those in other conditions when the prime and target shared overlapping semantic features or exhibited strong associative strength.

Estes and Jones explain integrative priming effects as follows. When participants read a prime word, they not only activate features of its meaning, but also thematic relations typically associated with it (e.g., relations from relevant frames). When reading the prime *FARM*, for example, a locative relation becomes active, specifying that *FARM* is a location attribute that takes other objects located there as values (e.g., *crops, tractors*). Although *mouse* may not have co-occurred sufficiently often with *FARM* to establish a strong association with it, and although *mouse* and *FARM* do not share semantic features, a *mouse* is nevertheless something that constitutes a potential value of the locative relation associated with *FARM's* meaning. Thus, reading *FARM* followed by *mouse* implicitly forms a natural compositional structure of a *mouse found on a farm*.

Other research on concept composition supports this account. When researchers have assessed the thematic relations associated with nouns, they have found that some are more central to a word's meaning than others (e.g., Gagné & Shoben, 1997; Gagné & Spalding, 2014; Wisniewski, 1997). Thus, when two nouns are combined into a noun-noun phrase, the better they fit default thematic relations that become active initially, the faster they are combined (with the modifier's thematic relations typically being more dominant than those of the head noun). Consider the relative ease of comprehending *KITCHEN floor* vs. *KITCHEN plan*. Because *KITCHEN* tends to prime a locative thematic relation faster than an instrument thematic relation, participants tend to process *KITCHEN floor* faster. As for integrative priming, thematic relations appear to become active rapidly, anticipating composition with relevant concepts.

3.4.2. Non-linear composition. When concepts combine, their combined meaning could be a linear combination of their individual meanings. Interestingly, however, linear combination rarely seems to occur, ruling out certain classes of compositional mechanisms at the algorithmic level and implicating others (e.g., Costello & Keane, 2000; Hampton, 1997, 2007; Medin & Shoben, 1988; Murphy, 1988; Wu & Barsalou, 2009).

Consider an example of the non-linearity that often occurs, along with implications for potential compositional mechanisms at the algorithmic level. When Wu and Barsalou (2009) gave participants object words and asked them to generate features of their meanings, participants overwhelmingly produced features from the objects' exteriors, relative to their interiors. When producing features of *LAWN*, for example, participants produced more external features such as *green* and *blades* than internal features such as *roots* and *dirt*. Interestingly, however, when another group of participants received *LAWN* combined with the modifier *rolled-up* (i.e., *rolled-up LAWN*), they produced more internal features than external features of *LAWN*. Notably, this shift from external to internal features did not occur for all concept combinations, such as *rolled-up SNAKE*, indicating that the shift is not a simple linear function of the modifier.

This pattern of results demonstrates, first, that the meaning of a concept composition is not a simple linear function of its individual word meanings. If it were, then the relative proportion of external to internal features for a concept should remain constant across processing the head noun in isolation vs. processing it in a noun phrase. Instead, concept composition is often highly non-linear, such that major shifts in feature salience occur.

Second, these results suggest that occlusion mechanisms in the visual system contribute to concept composition via simulation and imagery. Wu and Barsalou predicted that when people produce features of an object, they typically construct a multimodal simulation of it, and then report the features that they perceive in the image. As a consequence, people tend to produce unoccluded features more easily than occluded features, which are less visible. When people produce features for *LAWN*, they produce more external than internal features, given that external features are unoccluded (e.g., *blades, green*). Conversely, when objects are combined with modifiers such as *rolled-up*, a simulation of removing the external surface may be performed, such that internal features become more salient, and are thus produced more often (e.g., *roots, dirt*). Because simulating a *rolled-up SNAKE* doesn't expose its internal features, the shift from external to internal features does not occur. As this pattern of results suggests, simulation mechanisms can play important roles in determining whether the meanings of combined comments exhibit various patterns of feature salience.

3.4.3. Summary. As we have seen, concepts combine with other concepts naturally and ubiquitously, with frames, thematic roles, and simulation playing important roles. Again, it is difficult to imagine how concept composition could occur without postulating mechanisms at the algorithmic level. Indeed, theoretical accounts of concept composition have universally included such mechanisms. In the absence of such mechanisms, we understand little about the powerful human ability to construct infinite complex concepts, many of which have never been experienced.

Finally, taking concept composition into account seems essential for successful neural encoding and decoding. To see this, consider the design matrix for Huth et al.'s (2016) study in Figure 1B. As can be seen, only context vectors for individual words were used as regressors to predict BOLD activity. No regressors for the combinations of these words were included, even though each word was undoubtedly combined with others. As a consequence, word lists, rather than coherently integrated phrases and sentences, established the basis for neural encoding. Although word lists enable effective neural encoding to some extent, as Huth et al.'s (2016) Figure 1 illustrates, significantly better encoding would probably result if information about concept composition were included as well.

3.5. Conceptual flexibility

For decades, flexibility has been a hallmark of conceptual processing. Although researchers often adopt the idealization that a concept's content remains constant, as in Huth et al. (2016), considerable evidence illustrates that it does not. As we just saw, for example, the information active for a concept like *LAWN* varies non-linearly across concept compositions that contain it, exhibiting conceptual flexibility as it adapts to accompanying words.

3.5.1. Further examples of conceptual flexibility. As Yeh and Barsalou (2006) review, conceptual flexibility has been demonstrated since the early 1970s across behavioral literatures associated with perception, memory, concepts, and language. Several classic examples associated with semantic processing are described next.

In Barclay, Bransford, Franks, McCarrell, and Nitsch (1974), participants were asked to comprehend sentences so that they could answer questions about them later. For each critical word in the experiment, a given participant received it in one of two sentences that made different features of its meaning relevant. For *PIANO*, the sentence, "The man tuned the piano," made features about the *sound* of *PIANO* relevant, whereas the sentence "The man lifted the piano," made features about its *weight* relevant. Of interest was whether these context sentences altered the features that became active for *PIANO*, or whether a stable set of features was active in each context. To assess this issue, Barclay et al. gave their participants a surprise recall test, asking them to remember nouns from the sentences. To facilitate recall, participants received cues relevant to both primed meanings of each target word (even though only one meaning had been primed for a given participant). Cues relevant to *PIANO*, for example, were *something with a nice sound* and *something heavy*. Barclay et al. found that words were best recalled on trials when participants were cued with phrases consistent with the meaning that had been primed for them during learning (e.g., when participants had studied the sentence, "The man lifted the piano," *something heavy* produced higher recall of *PIANO* than *something with a nice sound*).

In a verification paradigm, Barsalou (1982) had participants read sentences and then verify whether a property read after the sentence was true or false of the sentence's subject noun. If, for example, the sentence's subject noun was *BASKETBALL*, participants might have to verify that *floats* was a property. To assess conceptual flexibility, half the participants received a subject noun in a neutral sentence (e.g., "The *BASKETBALL* was well-worn from much use"), whereas the other half received it in a priming sentence (e.g., "The *BASKETBALL* was used as a life preserver when the boat sank"). Of interest was whether priming the critical property with the sentence sped its subsequent verification. Consistent with conceptual flexibility, a 145 ms priming effect occurred. Drawing attention to a property of the subject noun with a sentence increased the property's salience considerably, demonstrating that the noun's representation varied dynamically with context. In subsequent priming experiments with greater control and improved designs, Greenspan (1986) demonstrated similar findings (for review of additional priming experiments, see Yeh & Barsalou, 2006).

Finally, Hampton (1988) found that category membership varies during conceptual composition, further demonstrating conceptual flexibility. When, for example, participants were asked whether *chess* belongs to the category of *SPORTS*, they tended to say no. Interestingly, however, when participants were asked whether *chess* belongs to the category of *SPORTS WHICH ARE ALSO GAMES* (a more restricted category than *SPORTS*), the likelihood that *chess* was included surprisingly became larger (rather than becoming smaller). Because *chess* is typically viewed as a *GAME*, the meaning of *SPORTS* in *SPORTS WHICH ARE ALSO GAMES* expanded toward *GAMES*. Using the feature listing task, Hampton (1987) further demonstrated that the features active during these concept compositions changed to enable this expansion, again demonstrating conceptual flexibility (also see Hampton, 1997).

3.5.2. Recent proposals of conceptual flexibility. Although demonstrations of conceptual flexibility have been prevalent since the early 1970s, current researchers continue to echo similar themes strongly. Lebois et al. (2015), for example, argued that concepts do not have conceptual cores activated automatically across contexts. In an initial literature review, they illustrated how classic phenomena associated with automaticity actually exhibit considerable flexibility and context-dependence (e.g., in Stroop, Simon, SNARC, attentional cuing, and grounded congruency tasks). When processing a colour word such as "red" in the Stroop task, for example, the semantic feature *red* varies considerably in its accessibility across task contexts. Rather than becoming active in an automatic ballistic manner, its accessibility varies with its relevance to the current task. Lebois et al. further report experiments from the grounded congruency paradigm illustrating how the accessibility of salient spatial features in concepts varies across task contexts (e.g., *high* for *SKY*, *low* for *DIRT*; also see Santiago, Román, & Ouellet, 2011; van Dam, Brazil, Bekkering, & Rueschemeyer, 2014). Lebois et al. concluded that concepts are constructed in a Bayesian manner, such that the information most likely to be incorporated into a concept on a given occasion reflects: (1) information that has been frequently and habitually active for the concept previously (priors), (2) information that is relevant in the current context (likelihoods). Because context can exert powerful effects on the information incorporated into a concept's construction, it can override habitually associated information, such that conceptual cores aren't observed across contexts.

Connell and Lynott (2014) similarly proposed that every time a concept is represented, it is represented in a different manner. Rather than a fixed stable representation always being retrieved across contexts, a novel representation is constructed, shaped by three constraints: (1) perception of the current situation and task demands, (2) distributed linguistic representations that shape conceptual representations and provide place holders (3.2.3), and (3) continuous change in the conceptual system over time in long-term memory. As a result of these factors, no concept is ever represented the same way twice. Casasanto and Lupyan, (2015) similarly proposed that all concepts are ad hoc. Rather than a concept being associated with static representation in memory, a unique representation is constructed on each occasion it is processed, reflecting a wide variety of current and long-term contributions from cognitive and social mechanisms. For similar accounts, see Barsalou (1987, 1989, 1993), Barsalou et al. (2008), and Barsalou, Wilson, and Hasenkamp (2010).

Finally, Yee and Thompson-Schill (2016) reviewed evidence that a concept's content varies as a function of its long-term context, recent context, immediate context, and ongoing context. They further argued that a concept's content cannot be separated from the contexts in which it occurs, with background contexts often being absorbed into a concept's current form (Schwanenflugel, 1991; also see Barsalou, in press). Finally, Yee and Thompson-Schill proposed algorithmic mechanisms capable of explaining these effects, drawing on the architecture of recurrent neural networks.

3.5.3. Summary. As we have seen, concepts exhibit considerable flexibility. Rather than a static concept representing a category, temporary conceptualizations of the category are constructed dynamically across situations, adapting to current constraints. As we have also seen, researchers propose that many mechanisms underlie flexibility, including grounded representations, distributed linguistic representations, features, frames, recurrence, and various memory processes. Again, it seems difficult, if no impossible, to explain conceptual flexibility without postulating mechanisms at the algorithmic level. Without such accounts, we understand little about how concepts are constructed flexibly across the situations in which they are processed.

Finally, taking conceptual flexibility into account seems essential for successful neural encoding and decoding. To see this, consider Figures 3 and 4, showing Huth et al.'s (2016) results for semantic selectivity and semantic tiling, respectively. As can be seen, these figures depict static semantic maps that do not change across the contexts in which a word's meaning is processed. Thus, one problem for this account is that it doesn't capture a basic fact about semantic processing: The semantic representation of a word changes with context. Additionally, this approach sanctions the mistaken assumption that concepts take static forms.

Another more practical problem is that failing to incorporate conceptual flexibility limits the success of neural encoding and decoding. If one's goal is to decode a neural state, for example, then anticipating and utilizing conceptual flexibility should be useful if not essential. Because a concept takes different forms in different situations, optimal decoding should occur when the concept's predicted neural pattern matches the contextualized pattern that occurs.

One possibility would be to map every composition containing the same concept into a different neural state over the course of voxel-wise modeling, essentially constructing a large library of states for a concept across all its compositions. A problem with this approach, however, is that it doesn't productively predict the neural states for new combinations. An alternative would be to develop theories of compositional mechanisms at the algorithmic level that can be used to productively represent and predict neural states at the implementation level. Rather than simply using a look-up table, neural encoding and decoding would utilize predictions based on the algorithmic mechanisms likely to be operating.

4. Conclusions

As we have seen, decades of research across diverse research communities have established representation and processing mechanisms at the algorithmic level that produce conceptual processing. Specifically, we have seen evidence for mechanisms associated with features and frame structure, multiple representations, knowledge-based inference, concept composition, and conceptual flexibility. Clearly, much remains to be learned about each class of mechanisms, and we are far from having anything close to an adequate account of the algorithmic mechanisms that produce conceptual and semantic processing. Many other mechanisms not covered here are certainly relevant as well. Furthermore, some of the mechanisms proposed here could be incorrect, or at least in significant need of revision. Nevertheless, it would be quite surprising if conceptual processing could be explained successfully without incorporating mechanisms like those reviewed here.

Again, one might argue that we don't need algorithmic accounts of conceptual processing. Perhaps mappings between the computational (task) level and the implementation (brain) level will be sufficient (Figure 5A). Perhaps distributed neural states constitute the critical algorithmic mechanisms (Figure 5C). Arguments like these didn't succeed for Behaviorism prior to the Cognitive Revolution, however, and are unlikely to succeed for what again might be called

Neurobehaviorism. Mechanisms have traditionally played central roles across the sciences, and undoubtedly play central roles in neuroscience as well (e.g., Bechtel, 2008, 2009; Bechtel & Abrahamsen, 2005; Bechtel & Shagrir, 2015). From atomic particles in physics, to genetics in biology, to representations and processes in cognitive science, postulating mechanisms and testing hypotheses about them has played central roles throughout the history of science.

4.1. Voxel-wise modeling provides traces—not explanations—of semantic processing

Assuming that mechanisms at the algorithmic level produce semantic processing, what does the voxel-wise modeling associated with neural encoding and decoding accomplish? Most obviously, it develops effective tools for predicting neural states from stimuli, and for inferring presented stimuli from neural states (e.g., Haxby et al., 2014; Huth et al., 2016; Naselaris et al., 2011; Weichwald et al., 2015).

Less obviously, voxel-wise modeling provides traces at the implementation level of mechanistic processing at the algorithmic level. If one assumes that algorithmic mechanisms produce semantic processing, then voxel-wise modeling establishes how various brain areas participate in mechanistic activity. In Huth et al., for example, we saw that voxel-wise modeling establishes how the cortical surface responds to computational-level descriptions of animacy and concreteness (although "animacy" and "concreteness" might not actually be the most accurate ways of describing these complex dimensions). Critically, however, the actual mechanisms that represent and implement animacy and concreteness in conceptual processing remain unclear. Voxel-wise modeling only tells us what parts of the brain become active when concepts related to these external descriptors are processed. It doesn't tell us what functions these brain areas perform mechanistically. Because no proposals about mechanisms have been made or tested, neural voxel-wise modeling tells us nothing about them.

Furthermore, the traces left behind by mechanistic processing most likely reflect the specific content active for concepts during the specific time period when voxel-wise modeling was performed, rather than establishing "deep" invariants of conceptual content. Because conceptual processing is flexible, adapting to current task conditions, it produces context-specific traces that reflect current task conditions. In Huth et al. (2016), for example, we saw that the semantic selectivity of tile LPC R5 appeared to reflect the specific content processed in the autobiographical narratives (i.e., relatives discussing criminal activity). When training with a different set of texts (e.g., philosophers vacationing on a cruise), the observed semantic selectivity of an area might change considerably.

As mentioned earlier, much other research using the techniques of voxel-wise modeling *does* address specific mechanisms, attempting to establish them in the brain. In the absence of a mechanistic orientation, however, voxel-wise modeling only establishes traces of the mechanistic processing engaged by stimulus and task structures operative during voxel-wise modeling.

4.2. Understanding conceptual and semantic processing in the brain

Presumably, it is important to establish the neural mechanisms that produce semantic processing, specifically, and conceptual processing, more generally. If we want to understand how the brain implements semantic and conceptual processing, it seems essential to understand the mechanisms responsible for these remarkable abilities. Otherwise, how do we explain the brain's natural capacity to produce them?

Cognitive science offers one obvious source of potentially relevant mechanisms. As we have seen, researchers across the cognitive sciences have been proposing and testing mechanisms of conceptual processing for decades. There is no shortage of cognitive mechanisms to explore in the brain.

It is an intriguing question whether these kinds of mechanisms will ultimately turn out to be useful in explaining how the brain implements conceptual processing. One possibility is that basic forms of these mechanisms will remain but evolve as they become better understood anatomically and physiologically. Another possibility is that these mechanisms will be largely replaced with new information processing mechanisms that conform more closely to the principles of neural computation. As we increasingly understand how specific forms of cytoarchitecture implement information processing activities, new constructs of representation and processing may enter the mechanistic vocabulary (Amunts & Zilles, 2015). It wouldn't be surprising if both of these possibilities were realized to some extent. What *would* be surprising, however, is discovering that information processing mechanisms of some kind aren't necessary.

Assuming that mechanisms remain important, it becomes essential to understand how the brain implements them, and how they work together to produce conceptual processing. Rather than simply trying to localize these mechanisms in the brain, it will be necessary to understand how they participate in coordinated systems to implement specific processes. Not only will it be important to characterize the input, output, and functional role of a given mechanism, it will be important to specify the organized operation of interacting mechanisms, and the temporal dynamics of their joint activity. Although fMRI has some ability to establish mechanisms and processing circuits in this manner, other neuroimaging methods such as EEG, MEG, and TMS will probably be necessary for establishing temporal dynamics successfully. Indeed, we may be into completely new technologies by the time we are ready to establish algorithmic processing effectively in the brain.

Nevertheless, various areas of neuroscience research have begun establishing algorithmic networks successfully in this manner, including perception (e.g., Ince et al., 2015; Schyns et al., 2009, 2016), conceptual processing (e.g., Mack, Love, & Preston, 2016; Mack, Preston, & Love, 2013), and decision making (O'Doherty, Hampton, & Kim, 2007). Additionally, methodological tools are increasingly being designed to establish algorithmic networks, including dynamic causal modeling (Friston, Harrison, & Penny, 2003), structural equation modeling (Gates, Molenaar, Hillary, & Slobounov, 2011; Molenaar, Beltz, Gates, & Wilson, 2016), and a variety of approaches adapted from mathematical psychology (Turner, Forstmann, Love, Palmeri, & Van Maanen, 2016).

Finally, simply establishing intrinsic functional connectivity falls short of establishing algorithmic networks (e.g., Hansen, Battaglia, Spiegler, Deco, & Jirsa, 2015; Sporns, 2010; Yeo et al., 2011). Rather than simply establishing correlated neural activity between related brain areas, further establishing each brain area's mechanistic role in coordinated activity to implement a specific task is essential.

4.3. Integrating levels of explanation

Following Marr (1982), understanding an intelligent activity, such as conceptual processing, is likely to require *integrated* explanations across the computational, algorithmic, and implementation levels. Precise descriptions of stimuli, responses, and their relations at the computational level are certainly essential, as is establishing the corresponding neural regions that process them at the implementation level. When, however, neuroscientists believe that mapping task descriptions to anatomical structures and physiological activity is sufficient for good neuroscience, they fail to appreciate that they probably haven't achieved what they really care about. They haven't really established how the brain works. Without specifying the mechanisms that make performance possible, the brain's operation remains unexplained. Furthermore, projects such as neural encoding and decoding are likely to become much more successful when they incorporate algorithmic mechanisms than when they rely solely on lengthy stimulus-response training and crowd sourcing.

There is likely to be significant disagreement about this. I predict, however, that neuroscience will never achieve its most basic goals until it focuses on mechanisms at the algorithmic level, and integrates them effectively with task descriptions, neural structure, and neural activity at the computational and implementation levels.

References

Amunts, K., & Zilles, K. (2015). Architectonic mapping of the human brain beyond Brodmann. *Neuron*, *88*, 1086–1107.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*, 245–266.

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*, 359–370.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–289.

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1235–1243.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, *13*, 471–481.

Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, *36*, 673–721.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. In Mark P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. Volume 41, pp. 167–218). Academic Press.

Barsalou, L. W. (in press). Cognitively plausible theories of concept composition. In Y. Winter & J. A. Hampton, *Compositionality and concepts in linguistics and psychology*. London: Springer Publishing.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). Cambridge: Cambridge University Press.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp. 21–74). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Erlbaum.

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

Barsalou, L. W. (2003a). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 1177–1187.

Barsalou, L. W. (2003b). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, *18*, 513–562.

Barsalou, L. W. (2005). Continuity of the conceptual system across species. *Trends in Cognitive Sciences*, *9*, 309–311.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1281–1289.

Barsalou, L. W. (2011). Integrating Bayesian analysis and mechanistic theories in grounded cognition. *Behavioral and Brain Sciences*, *34*, 191–192. h

Barsalou, L. W. (2012). The human conceptual system. In M. Spivey, K. McRae, & M. F. Joanisse, *The Cambridge handbook of psycholinguistics* (pp. 239–258). New York: Cambridge University Press.

Barsalou, L. W. (2016a). Can cognition be reduced to action? Processes that mediate stimuli and responses make human action possible. In A. K. Engel, K. J. Friston, & D. kragic, *Where's the action? The pragmatic turn in cognitive science (Strüngmann Forum Reports, Vol. 18. J. Lupp, Series Ed.)* (pp. 81–96). Cambridge, MA: MIT Press.

Barsalou, L. W. (2016b). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142.

Barsalou, L. W. (2016c). Situated conceptualization offers a theoretical account of social priming. *Current Opinion in Psychology*, *12*, 6–11.

Barsalou, L. W. (2016d). Situated conceptualization: Theory and applications. In Y. Coello & M. H. Fischer, *Foundations of embodied cognition: Volume 1. Perceptual and emotional embodiment* (pp. 11–37). East Sussex: Psychology Press.

Barsalou, L. W., & Hale, C. (1993). Components of conceptual representation. From feature lists to recursive frames. In I. Van Mechelen, J. A. Hampton, R. Michalski, & P. Theuns, *Categories and concepts: Theoretical views and inductive data analysis* (pp. 97–144). San Diego: Academic Press.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross, *Psychology of Learning and Motivation* (Vol. 43, pp. 43–92). New York: Academic Press.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M.

Glenberg, & A. C. Graesser, *Symbols, embodiment, and meaning* (pp. 245–283). Oxford: Oxford University Press.

Barsalou, L. W., Wilson, C. D., & Hasenkamp, W. (2010). On the vices of nominalization and the virtues of contextualizing. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 334–360). New York: Guilford Press.

Barsalou, L. W., Yeh, W., Luka, B. J., Olseth, K. L., Mix, K. S., & Wu, L.-L. (1993). Concepts and meaning. In K. Beals, G. Cooke, D. Kathman, K. E. McCulloch, S. Kita, & D. Teste, *Chicago Linguistics Society 29: Papers from the parasession on conceptual representations* (pp. 23–61). University of Chicago: Chicago Linguistics Society.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology, Vols. 1 and 2 (4th ed.)* (pp. 680–740). New York: McGraw-Hill.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, *22*, 543–564.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441.

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*, 312–322.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology*, *9*, 65–73.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Webb, E. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, *103*, 597–600.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 640–645.

Binder, J. R. (2016). In defense of abstract conceptual representations. *Psychonomic Bulletin & Review*, *23*, 1096–1108.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, 1–45.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796.

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

Bonner, M. F., & Grossman, M. (2012). Gray Matter Density of Auditory Association Cortex Relates to Knowledge of Sound Concepts in Primary Progressive Aphasia. *Journal of Neuroscience*, *32*(23), 7986–7991.

Bower, G. H., & Hilgard, E. R. (1981). *Theories of learning*. New York: Prentice-Hall.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase, *Visual information processing* (pp. 383–438). Oxford: Academic Press.

Bruner, J. S. (1973). *Beyond the information given: Studies in the psychology of knowing*. Oxford, England: W. W. Norton.

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*, *17*, 648–665.

Caligiore, D., Borghi, A. M., Parisi, D., & Baldassarre, G. (2010). TRoPICALS: A computational embodied neuroscience model of compatibility effects. *Psychological Review*, *117*, 1188–1228.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, *30*, 105–110.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence, *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge, MA: MIT Press.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*, 478–484.

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2014). Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex*, *24*, 2899–2907.

Chen, J., Papies, E. K., & Barsalou, L. W. (2016). A core eating network and its modulations underlie diverse eating phenomena. *Brain and Cognition*, *110*, 20–42.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 1–73.

Coello, Y., & Fischer, M. H. (Eds.). (2016a). *Foundations of embodied cognition: Volume 1. Perceptual and emotional embodiment*. Oxford, UK: Routledge.

Coello, Y., & Fischer, M. H. (Eds.). (2016b). *Foundations of embodied cognition: Volume 2. Conceptual and interactive embodiment*. Oxford, UK: Routledge.

Colzato, L. S., Van Wouwe, N. C., Lavender, T. J., & Hommel, B. (2006). Intelligence and cognitive flexibility: fluid intelligence correlates with feature 'unbinding' across perception and action. *Psychonomic Bulletin & Review*, *13*, 1043–1048.

Connell, L., & Lynott, D. (2013). Flexible and fast: Linguistic shortcut affects both shallow and deep conceptual processing. *Psychonomic Bulletin & Review*, *20*, 542–550.

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*, 390–406.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, *24*, 299–349.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*, 163–201.

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: the role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience*, *7*, Article 186.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*, 25–62.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.

De Vega, M., Glenberg, A. M., & Graesser, A. C. (Eds.). (2008). *Symbols, embodiment, and meaning*. Oxford: Oxford University Press.

Domjan, M. (2014). *The principles of learning and behavior*. Independence, KY: Cengage Learning.

Drane, D. L., Ojemann, G. A., Aylward, E., Ojemann, J. G., Johnson, L. C., Silbergeld, D. L., … Tranel, D. (2008). Category-specific naming and recognition deficits in temporal lobe epilepsy surgical patients. *Neuropsychologia*, *46*, 1242–1255.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, *6*, 635–653.

Estes, Z., & Jones, L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology: General*, *138*, 112.

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., … Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, *26*(5), 2018–2034.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni Di Semantica*, *6*, 222–254.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*, 1273–1302.

Fromkin, V., Rodman, R., & Hyams, N. (2013). *An introduction to language*. Cengage Learning.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 71.

Gagné, C. L., & Spalding, T. L. (2014). Conceptual composition: The role of relational competition in the comprehension of modifier-noun phrases and noun-noun compounds. *The Psychology of Learning and Motivation*, *59*, 97–130.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*, 98–123.

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, *54*, 1151–1158.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*, 222–262.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Greenspan, S. L. (1986). Semantic flexibility and referential specificity of concrete nouns. *Journal of Memory and Language*, *25*, 539–557.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, *15*, 55–71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 12.

Hampton, J. A. (1997). Conceptual combination. In K. Lamberts & D. R. Shanks, *Knowledge, concepts, and categories* (pp. 133–159). East Sussex: Psychology Press.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science*, *31*, 355–384.

Hansen, E. C. A., Battaglia, D., Spiegler, A., Deco, G., & Jirsa, V. K. (2015). Functional connectivity dynamics: Modeling the switching behavior of the resting state. *NeuroImage*, *105*, 525–535.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*, 151–167.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Henderson, J. M., & Hollingworth, and A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Hernnstein, R. J. (1984). Objects, categories, and discriminative stimuli. In H. L. Roitblat, H. S. Terrace, & T. G. Bever, *Animal cognition* (pp. 233–262). Hillsdale, NJ: Erlbaum.

Hernnstein, R. J. (1990). Levels of stimulus control: A functional approach. *Cognition*, *37*, 133–166.

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., … Crites, S. L. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin*, *139*, 1062–1089.

Hoenig, K., Müller, C., Herrnberger, B., Sim, E.-J., Spitzer, M., Ehret, G., & Kiefer, M. (2011). Neuroplasticity of semantic representations for musical instruments in professional musicians. *NeuroImage*, *56*, 1714–1725.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390–421.

Hsu, N. S., Frankland, S. M., & Thompson-Schill, S. L. (2012). Chromaticity of color perception and object color knowledge. *Neuropsychologia*, *50*, 327–333.

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*, 453–458.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*, 1210–1224.

Ince, R. A. A., van Rijsbergen, N. J., Thut, G., Rousselet, G. A., Gross, J., Panzeri, S., & Schyns, P. G. (2015). Tracing the flow of perceptual features in an algorithmic brain network. *Scientific Reports*, *5*, 17681.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*, 19–25.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*, 7241–7244.

Jamrozik, A., McQuire, M., Cardillo, E.R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review*, *23*, 1080–1089.

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 169–188.

Kan, I. P., Barsalou, L. W., Solomon, K. O., Minor, J. K., & Thompson-Schill, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, *20*, 525–540.

Kemmerer, D. (2015). Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. *Psychonomic Bulletin & Review*, *22*, 1068–1075.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, *28*, 12224–12230.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*, 449–467.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*, 171–175.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.

Lambon Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, *107*, 2717–2722.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2013). *Handbook of Latent Semantic Analysis*. East Sussex: Psychology Press.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*, 1764–1801.

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology*, *3*, 107.

Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Commun. ACM*, *38*, 33–38.

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin & Review*, *23*, 991–1001.

Lewis, J. W. (2006). Cortical networks related to human use of tools. *The Neuroscientist*, *12*, 211–231.

Löbner, S. (2014). Evidence for frames from human language. In T. Gamerschlag, D. Gerland, R. Osswald, & W. Petersen, *Frames and concept types* (pp. 23–67). New York: Springer.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273–302.

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–398.

Love, B. C. (2015). The algorithmic level Is the bridge between computation and brain. *Topics in Cognitive Science*, *7*, 230–242.

Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proceedings of the National Academy of Sciences*, *113*, 13203–13208.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from Its neural implementation. *Current Biology*, *23*, 2023–2027.

Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt.

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

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, *23*, 979–990.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, *270*, 102–105.

Martin, A., Simmons, W. K., Beauchamp, M. S., & Gotts, S. J. (2014). Is a single 'hub', with lots of spokes, an accurate description of the neural architecture of action semantics? *Physics of Life Reviews*, *11*, 261–262.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*, 375–407.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*, 547–559.

McRae, K., & Jones, M. N. (2013). Semantic memory. In D. Reisberg, *The Oxford handbook of cognitive psychology* (pp. 206–219). Oxford, UK: Oxford University Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, *1*, 64–69.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*, 158–190.

Melara, R. D., & Marks, L. E. (1990). Dimensional interactions in language processing: investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 539.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*, 545–567.

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*, 1191–1195.

Molenaar, P. C. M., Beltz, A. M., Gates, K. M., & Wilson, S. J. (2016). State space modeling of time-varying contemporaneous and lagged relations in connectivity maps. *NeuroImage*, *125*, 791–802.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, *12*, 529–562.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*, 902–915.

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*, 1641–1646.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage*, *31*, 440–457.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114.

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford: Oxford University Press.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.

Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. New York: Cambridge University Press.

Peebles, D., & Cooper, R. P. (2015). Thirty Years After Marr's Vision: Levels of Analysis in Cognitive Science. *Topics in Cognitive Science*, *7*(2), 187–190.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–648.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*, 458–470.

Reilly, J., Peelle, J. A., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: The dynamic multilevel reactivation framework. *Psychonomic Bulletin & Review*, *23*, 1002–1014.

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*, 550–558.

Renoult, L., Tanguay, A., Beaudry, M., Tavakoli, P., Rabipour, S., Campbell, K., … Davidson, P. S. R. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia*, *83*, 242–256.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory, 2*, 64–99.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA, US: MIT Press.

Roitblat, H. L., Terrace, H. S., & Bever, T. G. (1984). *Animal cognition*. Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rudebeck, P. H., & Murray, E. A. (2014). The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron*, *84*, 1143–1156.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Santiago, J., Román, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. In A. Maass & T. W. Schubert, *Spatial dimensions of social thought* (pp. 41–110). Berlin: Mouton de Gruyter.

Schwanenflugel, P. J. (1991). Why are abstract concepts so hard to understand? In P. J. Schwanenflugel, *The psychology of word meanings* (pp. 223–250). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402–409.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–17.

Schyns, P. G., Gosselin, F., & Smith, M. L. (2009). Information processing algorithms in the brain. *Trends in Cognitive Sciences*, *13*, 20–26.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*, 243–265.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 681–696.

Schyns, P. G., van Rijsbergen, N. J., & Ince, R. A. A. (2016). Realizing the promise of brain imaging within an information processing framework. *Manuscript in Preparation*.

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, *20*, 451–486.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*, 1602–1608.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*, 2802–2810.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*, 184–189.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*, 159–216.

Solomon, K. O., & Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, *32*, 244–259.

Sporns, O. (Ed.). (2010). *Analysis and function of large-scale brain networks*. Washington, DC: Society for Neuroscience.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153–156.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research*. New York: Wiley.

Troche, J., Crutch, S., & Reilly, J. (2014). Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Frontiers in Psychology*, *5*.

Trumpp, N. M., Kliese, D., Hoenig, K., Haarmeier, T., & Kiefer, M. (2013). Losing the sound of concepts: Damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex*, *49*, 474–486.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 830–846.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2016). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*.

van Dam, W. O., Brazil, I. A., Bekkering, H., & Rueschemeyer, S.-A. (2014). Flexibility in embodied language processing: Context effects in lexical access. *Topics in Cognitive Science*, *6*, 407–424.

Van Dantzig, S., Raffone, A., & Hommel, B. (2011). Acquiring contextualized concepts: A connectionist approach. *Cognitive Science*, *35*, 1162–1189.

van der Laan, L. N., de Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage*, *55*, 296–303.

Wang, J., Cherkassky, V. L., Yang, Y., Chang, K. K., Vargas, R., Diana, N., & Just, M. A. (2016). Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. *Cognitive Neuropsychology*, *33*, 257–264.

Wang, X., Han, Z., He, Y., Caramazza, A., Song, L., & Bi, Y. (2013). Where color rests: Spontaneous brain activity of bilateral fusiform and lingual regions predicts object color knowledge performance. *NeuroImage*, *76*, 252–263.

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action concepts in the brain: An activation likelihood estimation meta-analysis. *Journal of Cognitive Neuroscience*, *25*, 1191–1205.

Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., & Grosse-Wentrup, M. (2015). Causal interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, *110*, 48–59.

Werning, M., Hinzen, W., & Machery, E. (2012). *The Oxford handbook of compositionality*. Oxford: Oxford University Press.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*, 267–279.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. *Journal of Cognitive Neuroscience*, *25*, 920–935.

Winter, Y., & Hampton, J. A. (Eds.). (in press). *Compositionality and concepts in linguistics and psychology*. London: Springer Publishing.

Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, *4*, 167–183.

Wong, C., & Gallate, J. (2012). The function of the anterior temporal lobe: A review of the empirical evidence. *Brain Research*, *1449*, 94–116.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*, 1015–1027.

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *The American Journal of Psychology*, *119*, 349–384.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–1165.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*, 1028–1034.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

Author Notes

I am grateful to Philippe Schyns for ongoing discussion about the issues addressed in this article, and to Tyler Morgan for technical assistance and discussion. I am also grateful to Alex Martin for sharing his insightful analysis of these issues many years ago. Finally, I'm grateful to an anonymous reviewer, David Kemmerer, Brad Love, and Alex Martin for helpful comments on an earlier draft. Please address correspondence to Lawrence W. Barsalou, Institute of Neuroscience and Psychology, 58 Hillhead Street, University of Glasgow, Glasgow G12 8QB UK, lawrence.barsalou@glasgow.ac.uk, http://barsaloulab.org

Figure Captions

Figure 1. Panel A illustrates the context vectors established for each of the 10,470 story words with respect to the 985 topic words (basis functions) in Huth et al. (2016). Panel B illustrates the design matrix used to predict the BOLD signal during fMRI as participants listened to the stories. To model neural selectivity for the 985 basis functions, the BOLD time course in each voxel was regressed onto the fluctuating values of each basis function across context vectors for the individual words. See the text for further details.

Figure 2. Panel A illustrates the matrix of regression coefficients that Huth et al. (2016) computed during the regression analysis illustrated in Figure 1. One regression coefficient was computed for the predictive relationship between each basis function and each cortical voxel. Panel B illustrates the principle component analysis that Huth et al. performed to establish a low-dimensional semantic space. At the group level, four components each explained significant variance, together explaining about 20% of the variance across basis functions. Principle component analyses were also computed for individual participants (as illustrated in Figure 3), tending to explain about 35% of the variance. See the text for further details.

Figure 3. Reproduction of Huth et al.'s (2016) Figure 2 (permission pending). Panel a illustrates semantic interpretations of the four group components significant at the group level, two components at the time. To interpret each pair of components illustrated, positions of clusters from the cluster analysis are embedded in the respective two-dimensional space. Panel b illustrates maps of semantic selectivity across the cortex for one participant, with the color map indicating selectivity with respect to the three dimensions rendered. Panel c illustrates semantic selectivity maps for three additional participants. See the text for further details.

Figure 4. Reproduction of Huth et al.'s (2016) Figure 3, Panel c (permission pending), depicting a map of semantic tiles across the cortex (estimated from the group data). Each colored tile represents a relatively homogenous region of voxels as measured by principle component scores, contrasting discretely with tiles containing a different pattern of relatively homogenous values. This analysis establishes cortical regions of semantic prediction shared across participants. See the text for further details.

Figure 5. Illustrations of explanatory levels adapted from Marr (1982). Panel A illustrates omission of the algorithmic level in neural encoding and decoding, where the focus is on establishing relations between the computational and implementation levels for predictive purposes. Panel B illustrates representation and processing mechanisms at the algorithmic level central for conceptual and semantic processing. Panel C illustrates reification of the implementation level at the algorithmic level, assuming that distributed patterns of voxel activity constitute representational mechanisms. See the text for details.

Figure 6. Illustration of knowledge-based inference during situated action. Panel A illustrates a general frame for an event and its attributes. Panel B illustrates a setting and objects in the setting that activate a matching event frame in memory. Panel C illustrates subsequent projection of inferences from the event frame back into the situation, establishing self-relevance, producing relevant bodily and motor states, and predicting likely outcomes. See the text for further details.

Footnotes

[1] Only 985 words were available at the time the work was done, not 1000. For the current list, see: https://en.wiktionary.org/wiki/Appendix:1000_basic_English_words

[2] The ability of each voxel's 985 regression coefficients to explain its BOLD time course across the stories can be found at http://gallantlab.org/huth2016/ by clicking on the voxel and the displaying its value for model performance.

[3] For the full clusters, see page 33 in Huth et al.'s (2016) Supplementary Information.

[4] Although Huth et al. concluded that this component was uninterpretable, it might be construed as associated with external vs. internal information.

[5] To use the interactive tool, visit http://gallantlab.org/huth2016/

[6] The construct of *distributed* is used in several different ways throughout this article. In the above section and at many later points, a *distributed pattern of activation* refers to a distributed informational state across voxels established by voxel-wise modeling (during neural encoding and/or neural decoding). Elsewhere, a *distributed network* refers to a set of local brain areas distributed throughout the brain that are central to implementing a task or process, with these areas perhaps being established through functional connectivity methods. Note that these two senses of *distributed* can overlap, as when distributed patterns of activation occur within local areas of distributed networks. Finally, a *distributed linguistic representation* (as described in 3.2.3) is a collection of linguistic forms (typically words) that operates as a construct at the computational and/or algorithmic levels, described shortly. The specific form of *distributed* being used at a given point in the text should be clear from the context.

[7] A common criticism of psychological features is that they often appear to have the same status as the concepts they describe (e.g., the feature *seat* for the concept *chair* also appears to be a concept; cf. Binder et al., 2016). One approach to this issue is to assume that concepts for objects and features similarly represent their respective referents in perception and/or simulation (Barsalou, 2003a; Wu & Barsalou, 2009). Whereas *chair* is a concept that describes an object, *seat* is a concept that describes a feature of a chair (and a similar feature of other related objects). From this perspective, features are not constitutive of concepts but are simply related to them via various thematic relations, such as *part-of*, *made-of*, etc.

[8] A fourth kind of representation, amodal symbols, is also of interest to some researchers, but is not addressed here for reasons provided in Barsalou (2016b).

Table 1. Examples of words from the clusters in Huth et al.'s (2016) semantic analysis used to interpret the group principle components.

CLUSTER	REASONABLE	QUESTIONABLE
VISUAL	colour yellow stripes wide shaped	fur steel skull fielder cloth seal
TACTILE	fingers pinch soft smooth reach	clouds meters screens barrel sheets
LOCATIONAL	building stadium shops landscape	visitors golf evenings art company
MENTAL	knew memories asleep experience	senses talked replies moments hadn't
ABSTRACT	qualities artificial intricate natural	roots flesh environment hip folk
NUMERIC	four pair half drop cent per quarter	deck floors top purse sold
EMOTIONAL	fear anger hatred peaceful troubled	alive truth nature religion illness speak
TEMPORAL	minute clock schedule arrive weekend	rumbling heading travel parking trip
SOCIAL	married relatives pregnant widow son	arrest suicide calls informed whom
COMMUNAL	community public culture society	male sons banker wealthy interests
PROFESSIONAL	office business bank meeting owner	year school staying visit estate
VIOLENT	lethal painful die poison	pause tongue instantly reaction

Note. Clusters were originally reported in the Supplementary Information for Huth et al. (2016). The assignment of words to the Reasonable and Questionable groups has been added here for illustrative purposes, and is not from the original report.

Figure 1

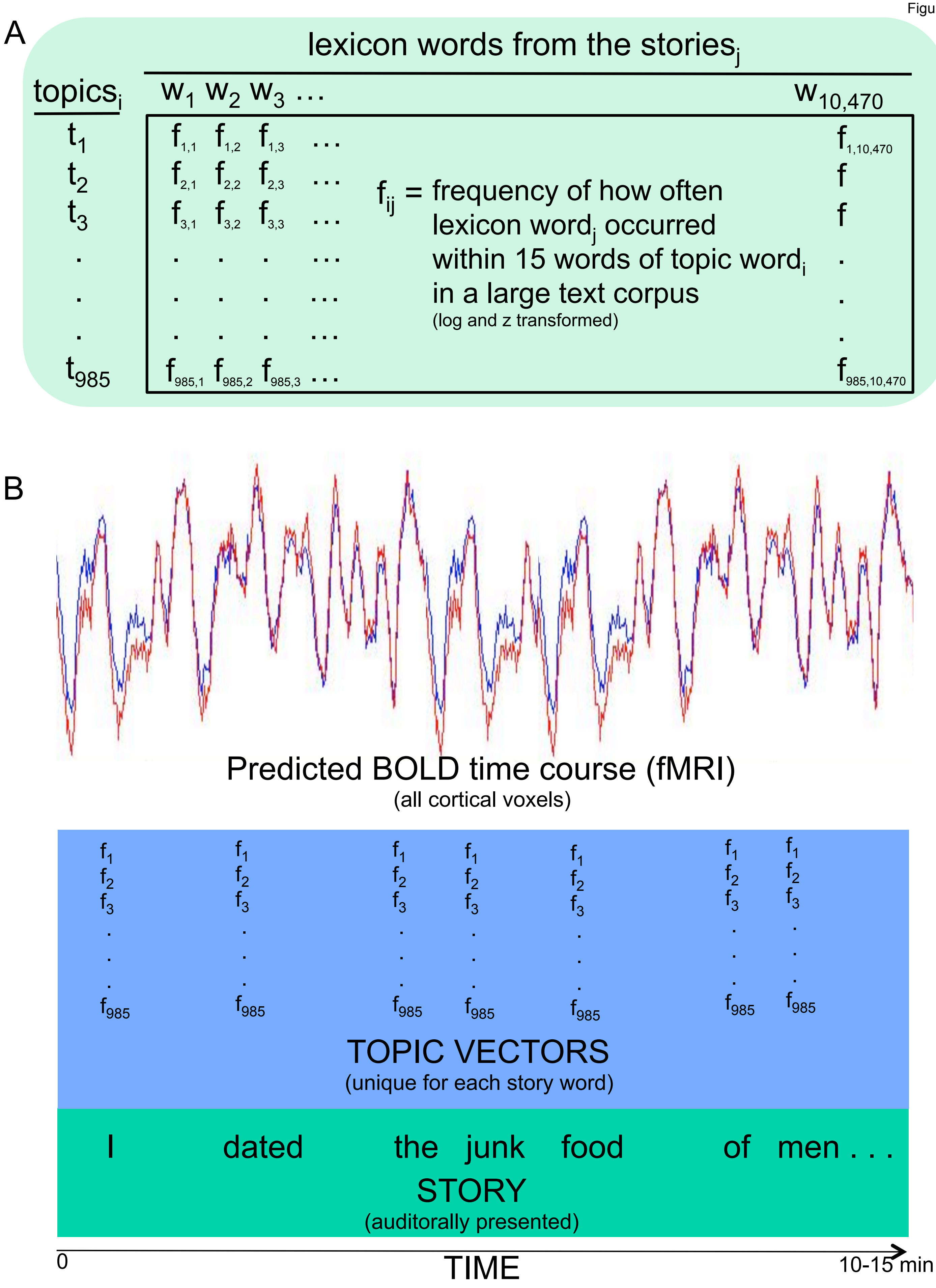

A
lexicon words from the stories$_j$
topics$_i$
w_1 w_2 w_3 …
$w_{10,470}$
t_1
t_2
t_3
t_{985}
$f_{1,1}$ $f_{1,2}$ $f_{1,3}$ …
$f_{2,1}$ $f_{2,2}$ $f_{2,3}$ …
$f_{3,1}$ $f_{3,2}$ $f_{3,3}$ …
$f_{985,1}$ $f_{985,2}$ $f_{985,3}$ …
$f_{1,10,470}$
$f_{985,10,470}$
f_{ij} = frequency of how often lexicon word$_j$ occurred within 15 words of topic word$_i$ in a large text corpus
(log and z transformed)
B
Predicted BOLD time course (fMRI)
(all cortical voxels)
f_1
f_2
f_3
f_{985}
TOPIC VECTORS
(unique for each story word)
I
dated
the
junk
food
of
men . . .
STORY
(auditorally presented)
0
TIME
10-15 min

Figure 2

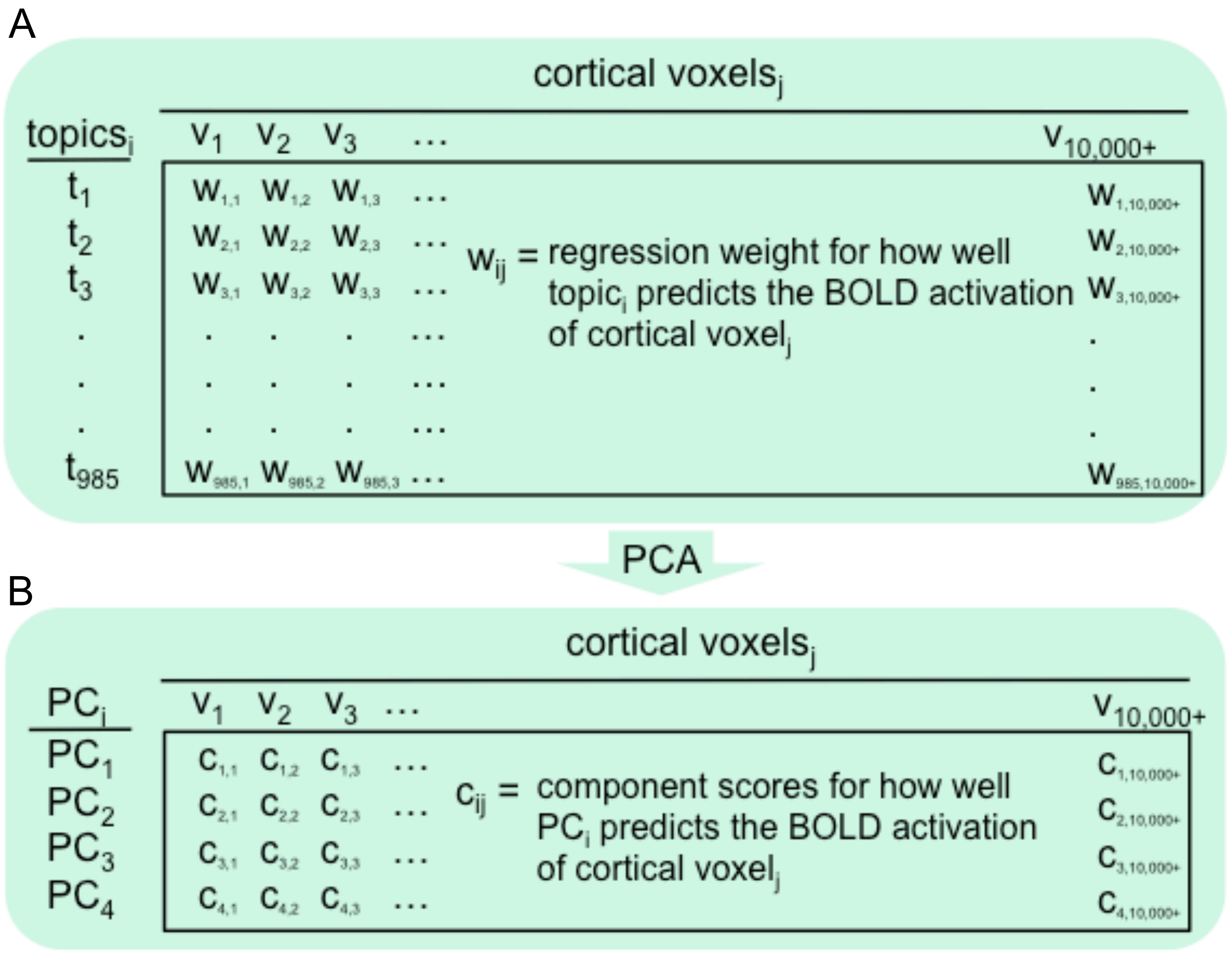

A
cortical voxels$_j$
topics$_i$
v_1 v_2 v_3 … $v_{10,000+}$
t_1 $w_{1,1}$ $w_{1,2}$ $w_{1,3}$ … $w_{1,10,000+}$
t_2 $w_{2,1}$ $w_{2,2}$ $w_{2,3}$ … $w_{2,10,000+}$
t_3 $w_{3,1}$ $w_{3,2}$ $w_{3,3}$ … $w_{3,10,000+}$
t_{985} $w_{985,1}$ $w_{985,2}$ $w_{985,3}$ … $w_{985,10,000+}$
w_{ij} = regression weight for how well topic$_i$ predicts the BOLD activation of cortical voxel$_j$
PCA
B
cortical voxels$_j$
PC$_i$
v_1 v_2 v_3 … $v_{10,000+}$
PC$_1$ $c_{1,1}$ $c_{1,2}$ $c_{1,3}$ … $c_{1,10,000+}$
PC$_2$ $c_{2,1}$ $c_{2,2}$ $c_{2,3}$ … $c_{2,10,000+}$
PC$_3$ $c_{3,1}$ $c_{3,2}$ $c_{3,3}$ … $c_{3,10,000+}$
PC$_4$ $c_{4,1}$ $c_{4,2}$ $c_{4,3}$ … $c_{4,10,000+}$
c_{ij} = component scores for how well PC$_i$ predicts the BOLD activation of cortical voxel$_j$

Figure 3

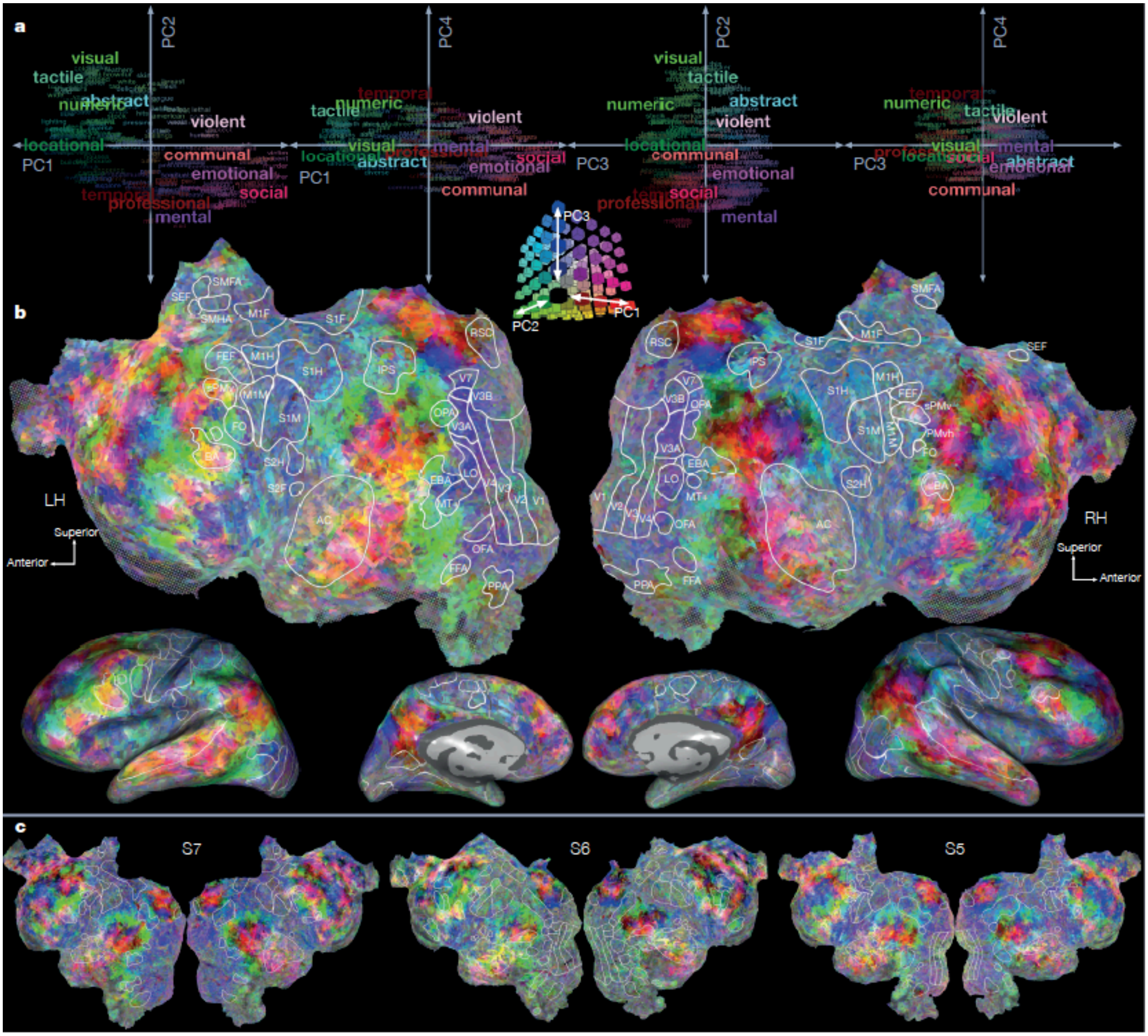

a
PC2
PC1
visual
tactile
abstract
numeric
violent
locational
communal
emotional
social
temporal
professional
mental
PC4
PC3
PC1
PC2
PC3
b
SMFA
SEF
SMHA
M1F
S1F
FEF
M1H
S1H
IPS
RSC
sPMv
M1M
S1M
FO
BA
S2H
S2F
AC
EBA
LO
MT+
V7
V3B
OPA
V3A
V4
V3
V2
V1
OFA
FFA
PPA
PMvh
LH
RH
Superior
Anterior
c
S7
S6
S5

Figure 4

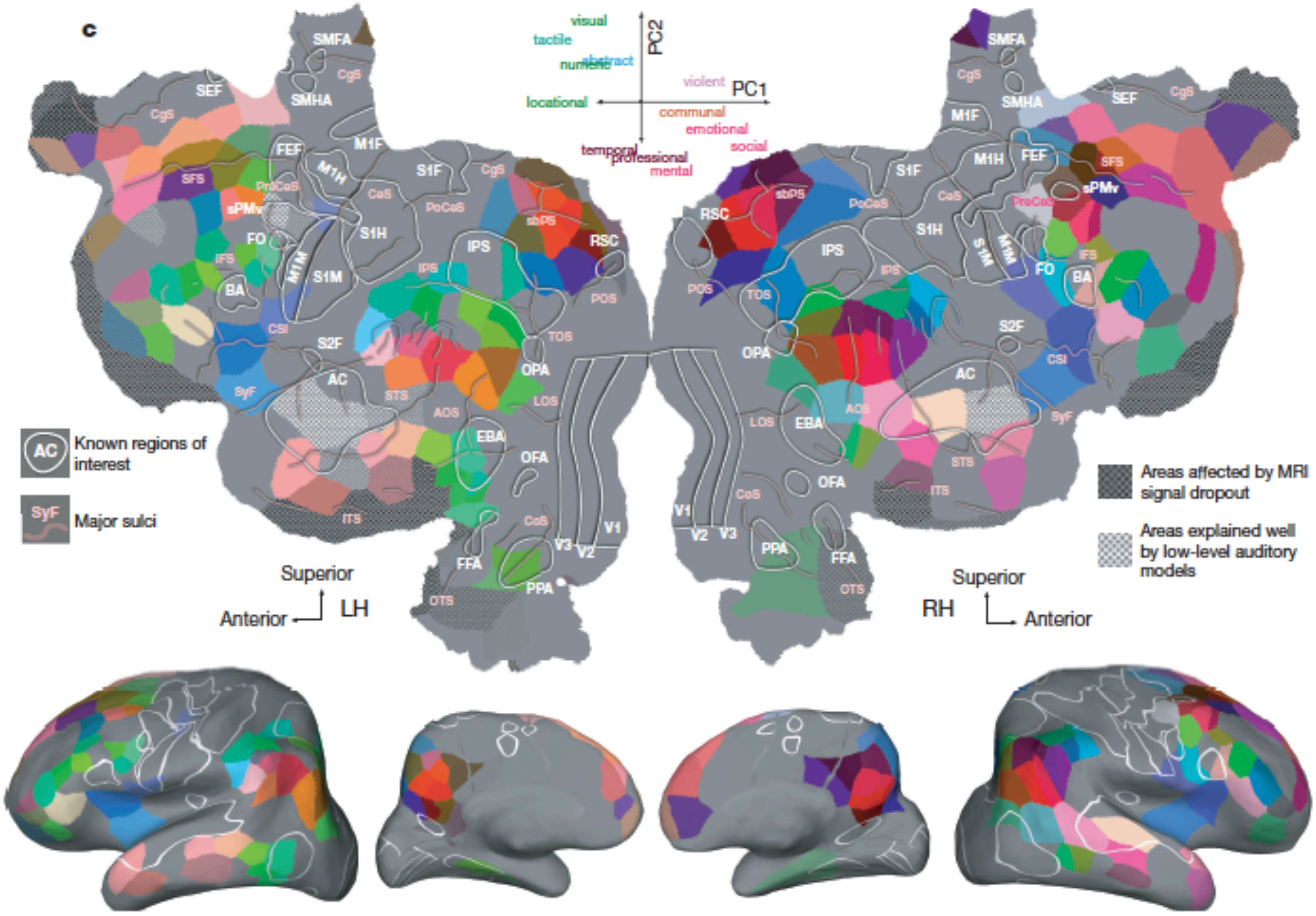

c
visual
tactile
numeric
abstract
PC2
violent
PC1
locational
communal
emotional
social
temporal
professional
mental
SMFA
CgS
SEF
SMHA
M1F
FEF
M1H
SFS
PrCeS
sPMv
FO
IFS
BA
M1M
S1M
S1H
S1F
CeS
PoCeS
IPS
sbPS
RSC
POS
TOS
CSI
S2F
AC
SyF
STS
AOS
OPA
LOS
EBA
OFA
ITS
CoS
V1
V2
V3
FFA
PPA
OTS
AC
Known regions of interest
SyF
Major sulci
Areas affected by MRI signal dropout
Areas explained well by low-level auditory models
Superior
Anterior
LH
RH
Superior
Anterior

Figure 5

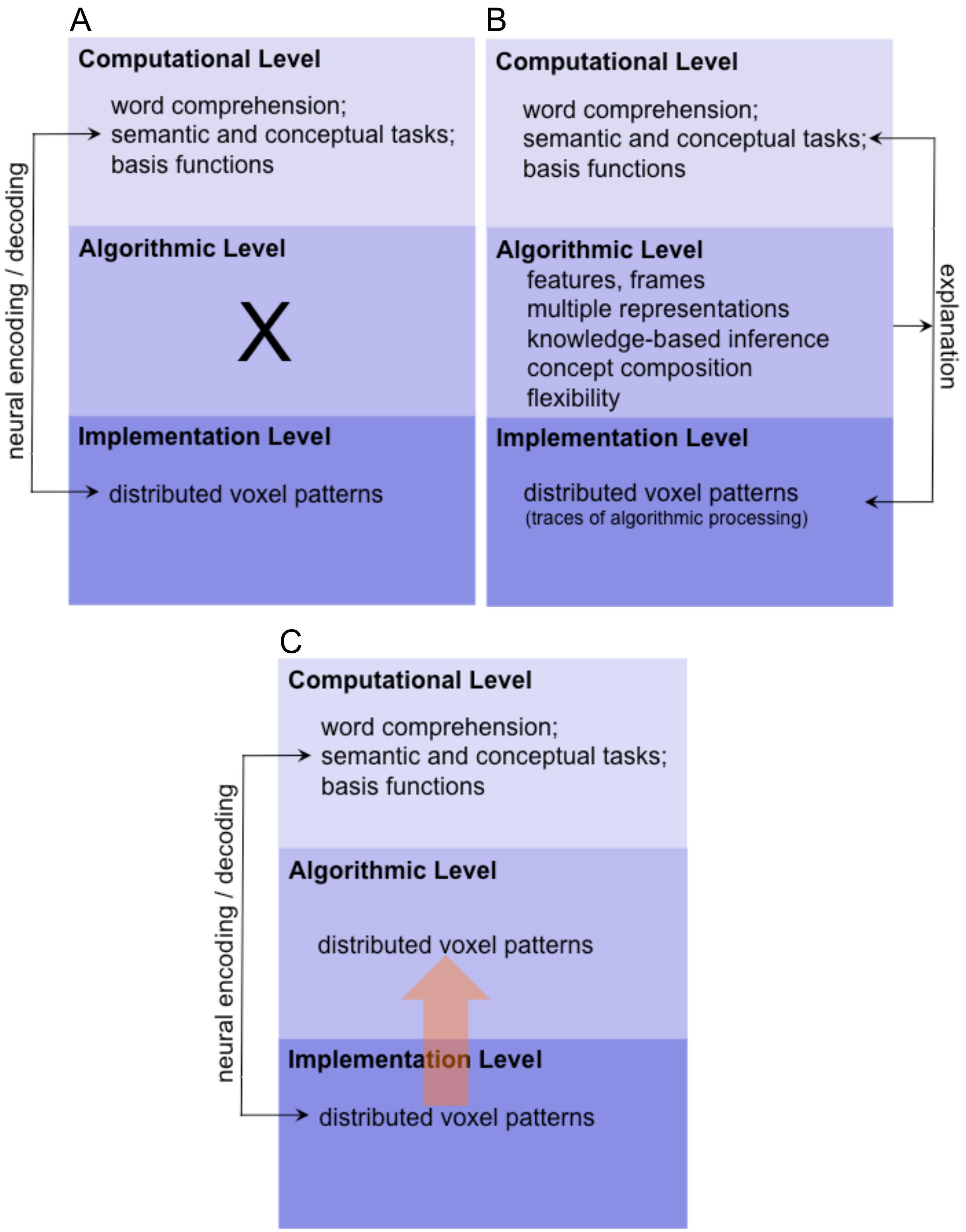
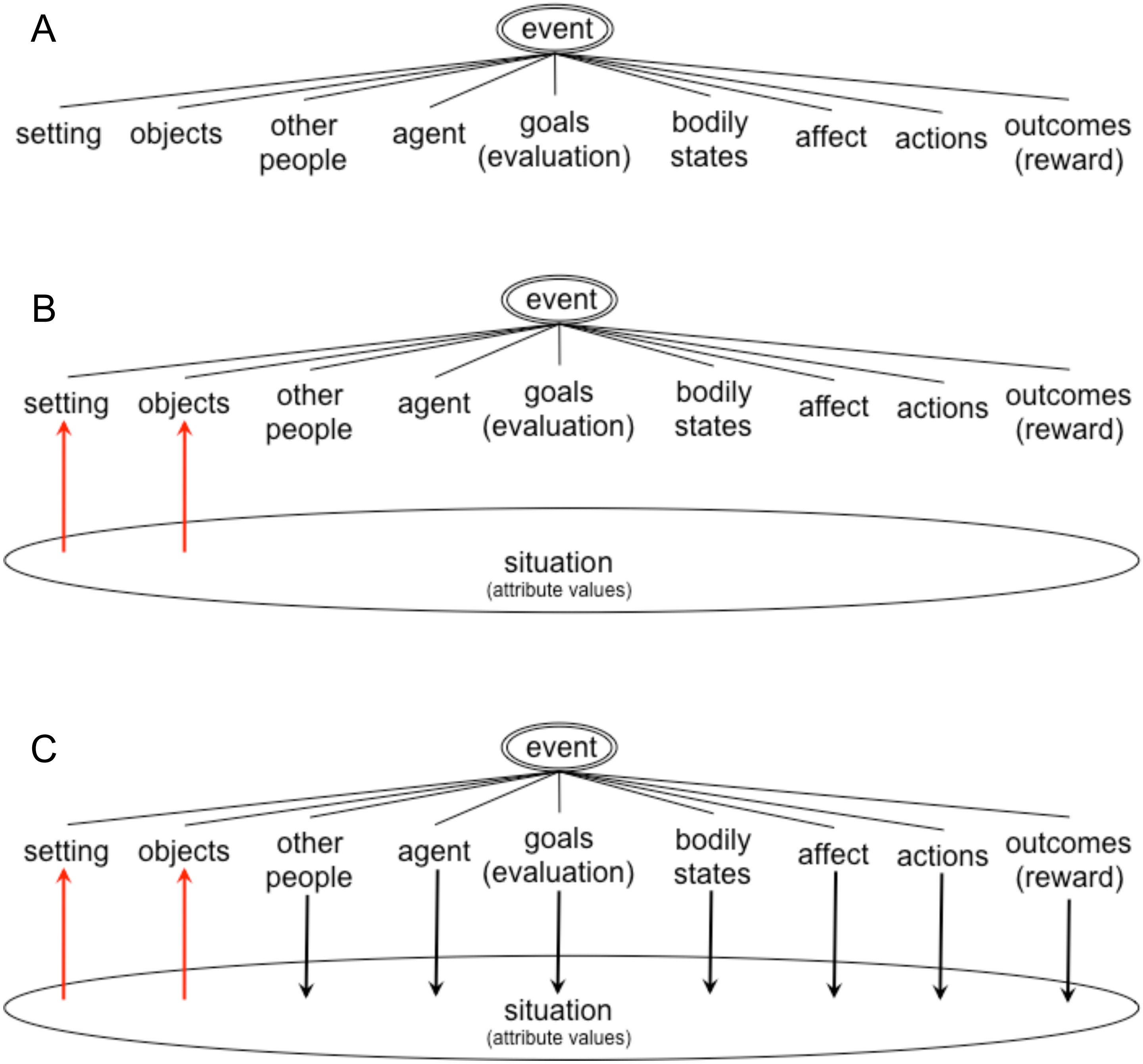

A
Computational Level
word comprehension;
semantic and conceptual tasks;
basis functions
Algorithmic Level
X
Implementation Level
distributed voxel patterns
neural encoding / decoding
B
Computational Level
word comprehension;
semantic and conceptual tasks;
basis functions
Algorithmic Level
features, frames
multiple representations
knowledge-based inference
concept composition
flexibility
Implementation Level
distributed voxel patterns
(traces of algorithmic processing)
explanation
C
Computational Level
word comprehension;
semantic and conceptual tasks;
basis functions
Algorithmic Level
distributed voxel patterns
Implementation Level
distributed voxel patterns
neural encoding / decoding

Figure 6


A
event
setting
objects
other people
agent
goals (evaluation)
bodily states
affect
actions
outcomes (reward)
B
event
setting
objects
other people
agent
goals (evaluation)
bodily states
affect
actions
outcomes (reward)
situation
(attribute values)
C
event
setting
objects
other people
agent
goals (evaluation)
bodily states
affect
actions
outcomes (reward)
situation
(attribute values)