Published in final edited form as:

Cognition. 2011 February ; 118(2): 211–233. doi:10.1016/j.cognition.2010.10.017.

Integrating Conceptual Knowledge Within and Across Representational Modalities

Chris McNorgan, Jackie Reid, and Ken McRae

University of Western Ontario, London, Canada

Abstract

Research suggests that concepts are distributed across brain regions specialized for processing information from different sensorimotor modalities. Multimodal semantic models fall into one of two broad classes differentiated by the assumed hierarchy of convergence zones over which information is integrated. In shallow models, communication within- and between-modality is accomplished using either direct connectivity, or a central semantic hub. In deep models, modalities are connected via cascading integration sites with successively wider receptive fields. Four experiments provide the first direct behavioral tests of these models using speeded tasks involving feature inference and concept activation. Shallow models predict no within-modal versus cross-modal difference in either task, whereas deep models predict a within-modal advantage for feature inference, but a cross-modal advantage for concept activation. Experiments 1 and 2 used relatedness judgments to tap participants' knowledge of relations for within- and cross-modal feature pairs. Experiments 3 and 4 used a dual feature verification task. The pattern of decision latencies across Experiments 1 to 4 is consistent with a deep integration hierarchy.

Keywords

semantic memory; multimodal representations; binding problem; embodied cognition

Semantic memory contains a great deal of knowledge regarding lexical concepts such as *dog* and *banana*, and as such is important for language processing, perception, reasoning, and action. Concepts referring to living and nonliving things include information such as how something looks, tastes, feels, and sounds, and how it is used. The manner in which this knowledge is represented and organized greatly impacts behavior. It is intuitive to think of this conceptual knowledge in terms of features. For example, how a typical *dog* looks or sounds can be described by features such as *<has legs>*, *<has a tail>*, *<has a nose>*, *<has ears>*, *<barks>*, and so on. Although some models of semantic memory are not based on feature representations – for example, Latent Semantic Analysis (Landauer & Dumais, 1997) – feature-based models, which describe concepts as collections of features at some level of abstraction, dominate the literature.

© 2010 Elsevier B.V. All rights reserved.

Please address all correspondence to: Ken McRae, Department of Psychology, Social Science Centre, University of Western Ontario, London, Ontario, Canada N6A 5C2, kenm@uwo.ca.

Author Notes

This work was supported by Natural Sciences and Engineering Research Council Discovery Grant 0155704, and National Institute of Health Grant HD053136 to KM.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The manner in which types of featural knowledge are neurally organized and integrated differentiates semantic memory models. Many of the remarkable capabilities of the human conceptual system are attributable to a large and highly interconnected network of processing units. As is elaborated below, although activity in one processing unit or region may eventually influence that of many others, it often does so indirectly. The fact that connectivity patterns determine the speed and/or strength of signal propagation between units has a number of important behavioral consequences. Therefore, the organization of conceptual representations directly influences cognitive processing because it determines the manner in which subsystems influence one another, and the temporal dynamics of such influences.

The goal of the present research is to provide the first direct test of two broad, central assumptions that have been made concerning how the brain organizes and uses types of knowledge. We test their behavioral consequences in tasks that are sensitive to the temporal dynamics of semantic processing in a distributed multimodal representational system. These assumptions concern whether modality-specific concepts are integrated using (1) direct connections or a single convergence zone (semantic hub), versus (2) a deeper hierarchy of convergence zones. A convergence zone is a neural region that binds information (Damasio, 1989). Convergence zones can, for example, bind features from a single modality into combinations (as in co-occurring clusters of visual parts) or bind features from multiple modalities (as in clusters of visual parts and correlated functions). Convergence zones can also bind information from lower-level convergence zones (and thus form a hierarchy).

Experiments 1 and 2 use feature relatedness judgments to tap participants' knowledge of relations for within-modal (<has 2 wheels> <has handle bars>) and cross-modal (<used by riding> <has handle bars>) feature pairs. Experiments 3 and 4 use a dual feature verification task with either within-modal (<has pockets> <has sleeves> *coat*) or cross-modal feature pairs (<worn for warmth> <has sleeves> *coat*). We found that within-modal feature relatedness latencies are shorter than cross-modal ones, and verification latencies are shorter given features from two modalities rather than one. These results favor models in which distributed modality-specific conceptual representations are bound together using a deep hierarchy of convergence zones.

To gain insight into the knowledge underlying people's concepts, researchers use tasks in which participants list features such as <has four legs>, <has fur>, <has a tail>, and <barks> for concepts like *dog*. Such features have been useful in accounting for a range of behaviors, from similarity judgments (Tversky, 1977) to theory generation (Ahn, Marsh, Luhmann, & Lee, 2002; McNorgan, Kotack, Meehan, & McRae, 2007). Although features like <is man's best friend> for *dog* reflect encyclopaedic-like knowledge, perhaps acquired linguistically, many features are learned by directly experiencing concepts' referents through the senses. For example, one sees that a dog has four legs, hears that it barks, and feels that it is covered in fur. Thus, many features are strongly associated with particular senses. Feature production norms, therefore, can be used to provide insight into the salience and amount of knowledge that people possess for each sensory modality with respect to individual concepts.

Multimodal versus Amodal Representations

It has long been known that some brain regions are specialized for perception in specific sensory modalities. The question of representational modality concerns the extent to which conceptual organization is tied to perceptual organization. That is, given that perception across the senses is distributed, at least in part, across specialized brain regions, it is possible that people's conceptual representations are organized similarly. One way to contrast models

is to partition them into *amodal* versus *multimodal* theories. Although various amodal theories make different assumptions with respect to what information is stored in semantic memory, all assume that objects or their features are represented in a single homogenous store. For amodal models, the sensory modality through which knowledge is gained is irrelevant to the representation of that knowledge because this information is lost when it is transduced into mental symbol systems. In contrast, multimodal theories posit that concepts are distributed across a wide network of brain areas, and that a concept's features are tied to sensory modalities.

The issue of whether the human conceptual system is multimodal or amodal remains under debate. However, the bulk of recent evidence from a number of lines of research favours the multimodal account. The literature regarding patients with category-specific semantic deficits has long been used to argue for multimodal representations. Warrington and McCarthy's (1987) sensory/functional theory accounts for patterns of category specific impairments of knowledge in patients who have suffered focal or diffuse brain damage, under the assumption that living things and artifacts differentially depend on visual and functional information – an assumption that has been supported and extended by analyses of feature norms (Cree & McRae, 2003; Garrard, Lambon Ralph, Patterson, & Hodges, 2001), and by a number of functional neuroimaging (see Martin, 2007; Martin & Chao, 2001 for reviews) and ERP experiments (Sitnikova, West, Kuperberg, & Holcomb, 2006).

The imaging literature also provides a wealth of evidence extending the sensory/functional theory that supports a distributed multimodal representational system. Goldberg, Perfetti, and Schneider (2006a) used fMRI to tie together previously reported neuroimaging evidence supporting modally bound tactile, colour (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Mummery, Patterson, Hodges, & Price, 1998), auditory (Kellenbach, Brett, & Patterson, 2001), and gustatory representations (Goldberg, Perfetti, & Schneider, 2006b). Goldberg et al. (2006a) found that sensory brain areas for each modality were recruited during a feature verification task that used linguistic stimuli (e.g., *banana-yellow*). These results indicate that the semantic representations activated from linguistic stimuli are modally distributed across brain regions. In summary, a number of complimentary techniques provide converging evidence supporting a distributed multimodal representational system.

Convergence Zones

Though concepts may be distributed neurally across a wide network, our mental experiences of them are not a jumble of features, disjointed across space and time, but instead they are experienced as coherent unified objects. Any model that uses distributed feature representations must account for what is sometimes called the binding problem: How are representational elements integrated into conceptual wholes? Similarly, how are we able to infer one feature from the presence of another, such as the likelihood that something flies if it has feathers? If one makes the additional assumption that semantic representations are modally-distributed, the binding problem becomes further complicated because it raises the question of whether within-modal binding is accomplished differently than cross-modal binding, or differs by modality. Understanding how distributed representations are integrated into conceptual wholes is therefore of central importance to evaluating semantic memory models and understanding brain function.

One solution to the binding problem involves temporal synchrony between firing rates of neurons (von der Malsburg, 1981, 1999). Object representations may be derived from the coincidental firing rates of distributed neural populations, bound together by virtue of firing at a particular rate. The dominant competing solution to the binding problem, described in

some detail by Damasio (1989), relies on the convergence zone, defined as a “record of the combinatorial arrangements [of feature-based sensory or motor activity]” (p. 26). A convergence zone can be thought of as a collection of processing units that receive input from and encode coincidental activity among multiple input units. In connectionist terms, a convergence zone may be likened to a hidden layer (Sejnowski, Kienker, & Hinton, 1986). Because they encode time-locked activation patterns, an important property of convergence zones is that they transform, rather than simply repeat signals, with one consequence being that convergence zones encapsulate information. In this way, successive convergence zones (or iterative feedback through individual convergence zones) may gradually build more complex or abstract representations. Naturally, these binding mechanisms are not mutually exclusive, and it is possible that various systems rely to different degrees on either or both, as might be suggested by evidence supporting both binding mechanisms (Treisman, 1999). We focus on convergence zones because it is clear that the organization of these regions plays an important role in multimodal semantic processing (Patterson, Nestor, & Rogers, 2007; Simmons & Barsalou, 2003) and because we have strong predictions how this organization should influence within- and cross-modal semantic processing.

If the multimodal conceptual system is built atop the highly interconnected perceptual system, one might reasonably assume that a similar pattern of connectivity has developed in the semantic system, and that the same neural regions that serve as sensory convergence zones also act as representational convergence zones. This need not be the case, however. For example, although the perceptual and conceptual systems may share some of the same pathways, there may be practical reasons for two functionally independent systems of convergence zones to have emerged (this may prevent synaesthetic experiences, or prevent top-down processing from inducing hallucinations). Moreover, although there may be processes common to perceptual and conceptual binding, they do differ in important ways. Segregating objects from the background or accommodating partially occluded objects are problems for object parsing in perceptual binding that do not seem to apply (or perhaps apply only analogously) in conceptual binding. Thus, even if there was a consensus about how the cognitive system solves the perceptual binding problem (and there is not), the manner in which multimodal semantic integration occurs would remain an open question.

A number of multimodal semantic theories have been proposed in the last two decades, and each makes somewhat different assumptions about the modalities that are represented and the relationships among them. These models can be broadly grouped into two classes, deep and shallow, based on the assumed hierarchy of convergence zones. Differences in assumed connectivity lead to untested predictions for how modally distributed information is integrated. Thus, tasks that are sensitive to the time course of integration of featural information either within or across modalities constrain models of semantic representation.

Convergence Zone Hierarchies

We use hierarchical depth to describe models with respect to the number and configuration of convergence zones, ranging from the shallowest models with no convergence zones to arbitrarily deep models. Although hierarchical depth is a continuous dimension, an interesting distinction can be made between shallow models that assume zero or one convergence zone, versus deeper models with multiple convergence zones.

Hierarchically Shallow Models—Hierarchically shallow models refer to those in which all semantic integration occurs in the same location. Modally segregated representational stores pass information to one another either through direct connections (and thus lack any convergence zones, as in Figure 1), or through a single convergence zone that integrates information from all representational modalities (Figure 1). A number of proponents of

multimodal semantic representations have put forward shallow hierarchy models. Farah and McClelland's (1991) implementation of Warrington and McCarthy's (1987) sensory/functional theory, depicted in Figure 1, and the attractor network used in Cree, McNorgan, and McRae's (2006) investigation of the roles played by distinguishing and shared features use direct interconnections between processing units, and do not include any distance assumptions. Examples of models employing a single convergence zone include the attractor network described in Cree, McRae, and McNorgan's (1999) simulations of semantic priming effects, Humphreys and Forde's (2001) Hierarchical Interactive Theory (HIT) model, and Patterson et al.'s (2007) semantic hub model (however, see the General Discussion for possible alternative conceptions of Patterson et al.).

Hierarchically Deep Models—Hierarchically deep models are those for which connective distance differs. Initial convergence zones generally integrate information from nearby representational units for a single modality, whereas others with successively larger receptive fields integrate multimodal information from more distant brain areas. In the most clearly hierarchical models, this information is passed forward from earlier convergence zones (Damasio, 1989; Simmons & Barsalou, 2003).

Quasi-hierarchical Models—Plaut (2002) implemented a computational model that uses a single hidden layer to make within- and cross-modal connections, and is therefore, strictly speaking, a shallow model. There is, however, an influence of connective distance: features within a modality are connected by relatively short proximal connections, whereas those across modalities are connected by relatively long distal connections. Given that neurons predominantly form short connections (Jacobs & Jordan, 1992), Plaut stated that the "literal implementation of [the model] is implausible" (p. 626). Thus, connective distance in Plaut's model corresponds to differing numbers of interposing processing units, which leads to one of two assumptions. The first is that longer connections comprise chains of signal repeaters; that is, each unit in the chain simply passes along an unmodified signal. However, neurons typically receive multiple connections, which in turn allows them to integrate, perform computations on, and modify signals. Thus, the second (and we believe more plausible) assumption is that increasing physical distance introduces additional integrative units, which, in essence, corresponds to introducing convergence zones. For this reason, we classify Plaut's model as quasi-hierarchical.

Amodal Models—In amodal models, because information is not functionally segregated by sensorimotor modality, convergence zones are not strictly required. One possibility is that associations among an object's features are coded via direct connections reinforced through statistical learning (Tyler & Moss, 2001). Note that the absence of modality information does not preclude convergence zones. Concept names, for example, could be assumed to encapsulate features of the concepts they represent, thereby integrating featural information. However, regardless of whether a particular amodal model includes convergence zones, the distinction between within-modal and cross-modal feature pairs would not influence the tasks used in the present experiments because factors like correlational strength between features were equated in our experiments.

Arguments for Deep and Shallow Convergence Zone Hierarchies

A number of theoretical considerations favour a shallow integration hierarchy. Multimodal semantic models have been criticized as lacking parsimony (Riddoch, Humphreys, Coltheart, & Funnell, 1998), therefore models specifying multiple hierarchical convergence zones would seem to be even less parsimonious. Furthermore, many semantic phenomena have been simulated using networks lacking convergence zones (Cree et al., 2006; Farah & McClelland, 1991), implying that a deep hierarchy of convergence zones may not be

necessary. Patterson et al. (2007) contend that generalized impairments that accompany semantic dementia are best explained by a single semantic hub that integrates information from all modalities, and Rogers et al. (2004) present a computational implementation of this idea that simulates a number of aspects of behavioral phenomena exhibited by semantic dementia patients.

On the flip side, there are anatomical constraints that seem to suggest a hierarchically deep organization. First, the volume of the human skull precludes the degree of connectivity required for the shallowest models that lack any convergence zones (Plaut, 2002). Bidirectional communication within a bank of n processing units requires the order of n^2 direct connections but just $n(\log n)$ connections to higher order integrating units capable of pattern separation. One might reasonably assume that the savings in the number of connections would favour a hierarchical organization. Second, candidate brain regions for a single convergence zone should have reciprocal projections to all modalities, and ablation of such an area should preclude any sort of multimodal conceptualization. Damasio (1989) argues that the only such region is the hippocampus, and because bilateral ablation of this structure does not lead to a catastrophic loss of the ability to conceptualize, it is unlikely that semantic integration occurs within a single convergence zone. On the other hand, one could argue that the sort of generalized impairments that accompany semantic dementia constitute a progressive breakdown of the conceptual system. Because this disease is accompanied by degeneration of anterior temporal lobes, Patterson et al. (2007) argue that this region is the locus for a single semantic hub. Third, the arrangement of cells, such as in visual cortex, into functionally distinct layers with progressively larger receptive fields may occur elsewhere in the brain, including those supporting conceptual processing, and would implement the sort of deep hierarchy suggested by Damasio and Simmons and Barsalou (2003).

The preceding discussion highlights a number of arguments favouring each of the two major assumptions regarding hierarchical organization. Both assumptions have been incorporated in models that have been used to explain a number of behavioral phenomena. Furthermore, the literature that speaks to the brain's connectivity, which would appear to be the best source of insight into constraints on the brain's ability to integrate information, does little to resolve the matter. A number of brain regions, including perirhinal cortex (Bussey, Saksida, & Murray, 2002), anterior temporal cortex (Patterson et al., 2007), frontal and prefrontal cortex (Fuster, Bodner, & Kroger, 2000; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006), and left inferotemporal cortex (Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004) have been put forward as critical structures for learning relationships among features from multiple modalities. However, it is unclear whether these areas represent a network of regions that act as a single convergence zone in a shallow system, or a hierarchy of convergence zones.

In summary, there is a sizeable literature that strongly supports the idea that semantic integration is accomplished across multiple brain regions. Because neural connectivity patterns appear to be consistent with both shallow and deep models, the existing literature is far from conclusive regarding the manner and location(s) of semantic integration.

Proximity-Sensitive Integration

The physical relationships among modality-specific representational areas and their convergence zones are assumed to influence the time course of information integration. Neurally proximal areas should generally communicate with one another in less time than distal areas. The aspect of proximity that we explore concerns the fact that neurons may (and commonly do) communicate with one another indirectly. Thus, rather than measure the spatial distance between two neurons, one might instead count the number of synapses that

connect them. Because the transmission of information at the synapse is not instantaneous, communication between a directly connected pair of neurons may be faster than between an indirectly connected but physically closer pair of neurons. When thinking about distance in terms of number of connections, one might also describe communication time between two processing units in terms of processing steps, with each step reflecting a unit of time required for communication across a synapse. Because shallow and deep integration hierarchies differ in this regard, it is clear that a model's integration hierarchy should influence the predicted time course of semantic processing. Moreover, as explained below, this influence may differ by task.

We make two critical assumptions, neither of which favours either shallow or deep-hierarchy accounts. We first assume that reading a feature name such as *<has a blade>* activates the underlying modality-specific representation. It has repeatedly been demonstrated in the imaging literature that feature names such as *<has a blade>* activate their underlying modality-specific neural representation (Chao & Martin, 1999; Goldberg et al., 2006a; Simmons, Martin & Barsalou, 2005; Simmons, Ramjee, Beauchamp, McRae, Martin, & Barsalou, 2007). One might argue that, for example, reading a feature name activates the corresponding modality-specific representation only after a delay. Thus, all stimuli in our experiments, because they are presented as words, would belong to the same modality, such as an amodal or 'lexical' modality, depending on the position one takes. Importantly, in the experiments reported herein, if this was correct, null effects would be predicted.

We further assume that activity spreads outward via neural connections and promotes activation of other representations, from which further activation spreads outwards, and so forth. In this way, if a convergence zone forms a path between representational stores, then verbally presented features can effectively prime subsequently presented ones, either individually as in feature-to-feature inferences (inferring that if something has wheels, it is used by riding), or entire clusters of features as in feature-to-concept activation (classifying a small animal as a skunk on the basis of its size, shape, colouration, and gait). The manner in which features are integrated should thus influence integration-based decision latencies. To investigate and constrain semantic memory models, the present research uses feature-to-feature (Experiments 1 and 2) and feature-to-concept judgments (Experiments 3 and 4) on correlated feature pairs from a large set of feature production norms (McRae, Cree, Seidenberg & McNorgan, 2005).

Experiment 1

Experiment 1 used within- and cross-modal feature pairs in a speeded relatedness decision task to test the role that modality plays in feature inference. Relatedness decisions are a fairly transparent measure of people's knowledge of the relations between object features, and therefore tap the types of processes used during feature inference (McNorgan et al., 2007). Feature inference involves determining the probability of whether some feature B exists for an object, given that the object is known to possess feature A, and may be accomplished independently of object categorization. For example, given that a novel object *<has a blade>*, one might infer that it is also *<used for cutting>*, or that it *<has a handle>*, without any other supporting evidence. To make this inference, the activation of knowledge of the visual form of a blade must propagate along neural circuitry to activate one's knowledge of handles or functional knowledge of the action of cutting. Shallow- and deep-hierarchy assumptions make different predictions about the time course of within- and cross-modal feature inference.

Figure 2 is a simplified shallow model with a single convergence zone and two representational modalities. Integrating two pieces of information from either a single modality (integrating two form features, such as *<has a handle>* and *<has a blade>*, Figure 2a) or cross-modally (integrating a function and a form feature, such as *<used for cutting>* and *<has a blade>*, Figure 2b) requires two steps: one step from a representational unit to the convergence zone, and a second from the convergence zone to the other feature. Therefore, in a timed task that is sensitive to processing steps, this model predicts no difference between within-modal and cross-modal integration. The speed with which this inference is made could reflect differences in the degree to which features are related, perhaps owing to a more efficient or dedicated neural pathway between the structures that encode this information. However, it should take no longer to infer *<has a blade>* from either *<has a handle>* or *<used for cutting>*, all else being equal.

Figure 3 is a simplified hierarchically deep model incorporating both within-modal and cross-modal convergence zones. Within-modal integration (Figure 3a) requires two steps: one step between each representational unit and the unimodal convergence zone. This architecture predicts that cross-modal integration (Figure 3b) should take more time because four processing steps are required: one from each representational unit to their corresponding unimodal convergence zone, and one from each unimodal convergence zone to the cross-modal convergence zone. Therefore, assuming that additional processing steps require additional time, it should take longer to infer the form feature *<has a blade>* from the cross-modal functional feature *<used for cutting>* than from the within-modal form feature *<has a handle>*, all else being equal.

Finally, in amodal models, because features are not segregated according to sensorimotor knowledge type, there is no within- versus cross-modal distinction, and therefore no predicted within- versus cross-modal differences, all other factors being equal.

Method

Participants—Twenty-one University of Western Ontario undergraduates received \$10 for participating in the speeded relatedness task. All participants in all of the studies reported herein were native English speakers and had either normal or corrected-to-normal visual acuity.

Materials—Correlated feature pairs were chosen from McRae et al.'s (2005) feature production norms. These norms include 541 concepts spanning a number of categories, including animals, tools, vehicles, clothing, musical instruments, and fruits and vegetables. Participants in McRae et al.'s norming task listed features they believed belonged to each concept. The norms include all features provided by at least 5 of the 30 participants who produced features for each concept.

A 541 concepts by 2,526 features matrix was constructed, where each matrix element corresponded to the number of participants listing a specific feature for a specific concept. Thus, each feature was represented by a 541-element vector so that a Pearson correlation could be computed between each feature pair. For the present study, spurious correlations were avoided by considering only pairs involving the 340 features that were listed for more than 3 concepts. The Pearson correlation between each feature pair was squared to obtain shared variance between features. Consistent with other research involving these production norms (McRae, Cree, Westmacott, & de Sa, 1999; McRae, de Sa, & Seidenberg, 1999), the threshold for inclusion in the analyses presented in this section was arbitrarily set at a shared variance of 5%, which serves as an estimate of the minimum degree of statistical relatedness required to be a psychologically real association. As in our previous research, correlations were calculated across all concepts, rather than within category. We felt this was an

appropriate approach for two reasons. First, the category to which a concept belongs is often unclear; a concept may belong to several categories at different levels of abstraction (*dog* as a carnivore, a mammal, or an animal) or may not belong to any obvious category at all (*ashtray*). Second, the tasks used in Experiments 1 and 2 involve decisions about feature relatedness without explicit regard to a particular category. We assume that the magnitude of the correlation between two features is related to the degree to which the underlying neural representations of these features are functionally connected, and therefore is a critical variable to control. It would be unclear which correlation should be used if feature pairs had unique values for each category.

Because the present investigation concerns communication between modality-specific brain regions, these norms are useful because Cree and McRae (2003) classified all features into nine knowledge types that are linked to modality-specific (or information-type specific) neural processing regions. Thus, for example, some authors have associated retrieval of object colour knowledge (as opposed to retrieval of other sorts of visual knowledge) with ventral regions of the posterior temporal lobes (Chao & Martin, 1999; Martin et al., 1995), and so features that describe an object's colour were assigned the visual-colour knowledge type. Three knowledge types corresponded to visual information (visual-colour, visual-parts and surface features, and visual-motion), four corresponded to other primary sensory-processing channels (smell, sound, tactile, and taste), one corresponded to functional/motor information regarding the ways in which people interact with objects (function). Finally, we excluded taxonomic features, which correspond to information regarding the categories to which concepts belong. We also excluded encyclopaedic features that correspond to types of information that either cannot be confidently mapped to particular neural regions, or, in a few cases, are not part of a coherent group of sufficient size to be of practical use (see Cree & McRae, 2003, for a detailed discussion of the considerations underlying selecting and assigning these knowledge types). The knowledge type assignments were based on the features themselves, and did not take object category into account. For example, though it has been demonstrated that people's knowledge of animal motion and artefact motion are represented in adjacent but somewhat distinct brain areas (Martin, 2007), we treated all motion features as being of the same knowledge type. One reason relates to the fact that our features are collapsed across categories, and it is therefore impossible to determine whether *<flies>* activates the representation characteristically associated with a *sparrow* or an *airplane*. A second reason is that the tasks in the experiments that follow either do not refer to particular categories (Experiments 1 and 2) or else are judgments about the same target (i.e., only one category is under consideration, Experiments 3 and 4).

The correlated feature pairs were divided into types based on sensorimotor modality. For example, form-function pairs involve a form feature and a function feature (*<has a blade>*, *<used for cutting>*), and form-form pairs involve two form features (*<has a blade>*, *<has a handle>*). We examined the patterns of correlations for each pairtype using mean shared variance, total number of correlated pairs, and sum of shared variance. Across all of these measures, correlations involving visual-form and function features tend to dominate. This is partly the result of the prevalence of these knowledge types in the production norms. People tend to list more features describing how objects look and how they are used, and so there are a greater number of potential correlations in which they may participate. In addition, some authors have argued for a special status of these knowledge types in conceptual representations (Tyler & Moss, 2001), and a rich literature examines how categories may differentially depend on form and function (Chao & Martin, 2001; Warrington & McCarthy, 1987). Thus, it seemed appropriate to begin our investigation with these knowledge types.

Concerning the validity of the modality assignment procedure, one might reasonably argue that the putative modality assignments for some features may be ambiguous (e.g., *<is fuzzy>*

as a tactile, rather than visual-form and surface feature). However, it is important to note that in the present experiments, analyses contrast within- versus cross-modal conditions, rather than focusing on individual modalities per se (form, colour, function, and so forth). We therefore argue that this potential objection is not relevant to the interpretation of our results for two reasons. First, although we do admit that the modality assignment procedure used for the McRae et al. (2005) norms can be ambiguous in some cases, they have been generally coherent across other investigations of modality-specific representations using neuroimaging and behavioural methodologies (Kalénine et al., 2009; Zahn et al., 2009). Second, for the purposes of this research, incorrect or ambiguous modality assignments either do not affect whether a given pair is within- or cross-modal (for example, *<is fuzzy>* and *<used by throwing>* is a cross-modal pair, whether *fuzzy* is treated as a visual or a tactile property), or they reduce our experimental power by intermixing items from the two experimental conditions.

Twenty form-form and 20 function-form feature pairs were selected (see Appendix A). Because the task measures the time required to judge perceived feature relatedness, the groups were matched on a number of variables that index the functional connectivity between the underlying neural representations of feature knowledge, or were expected to influence reading time and perceived relatedness (see Table 1). The mean percentage of shared variance is the mean of the squared Pearson correlations between the feature production frequency vectors created from our norms. The number of shared concepts is the number of concepts in which both features appear together, and reflects the ease of generating a concept possessing both features. This variable was matched in addition to shared variance because shared variance was calculated on feature production values, thus a pair of features may appear together in many concepts but have a lower proportion of shared variance than another pair appearing in fewer concepts. Moreover, concepts can be described as clusters of intercorrelated (i.e., n-way correlated) features (McRae et al., 1997). Because the feature pairs were matched with respect to the concepts in which they co-occurred or appeared individually, this ensured that the feature pairs also participated in the same higher-order correlations. The length in characters of the first and second feature indicates the number of characters, including spaces, in the feature name, and is assumed to influence reading time. The number of unique concepts listed for the first and second features is the number of concepts within our norms in which the one feature appears without the other. This indexes the likelihood that, if a feature was to prompt the retrieval of a concept, that concept would not provide evidence of feature pair co-occurrence. Finally, the number of concepts per feature is the number of concepts appearing in our norms for which the feature was listed by at least five participants.

There were no differences between form-form and function-form pairs on eight of the ten variables. In addition, each group had similar distributions of higher ($r^2 > .25$), medium ($.15 \leq r^2 \leq .25$) and lower ($r^2 < .15$) correlated pairs, such that the cross-modal group had 5, 3, and 12 and the within-modal group had 4, 2 and 14 high, medium and lower correlated pairs, respectively. Because the task was intended to measure judgments of relatedness between features without reference to any particular concepts, we avoided distinguishing features (i.e., features that are true of one or two concepts, such as *<moos>*). Rather, within- and cross-modal pairs included features that were true of an average of approximately 16 concepts, and the number of concepts per feature did not differ between conditions. This is an important point because the task was designed to tap people's knowledge at the level of individual features, rather than of particular concepts. Note that the shorter length of the first feature of form-form pairs was not a critical issue for two reasons. First, decision latencies were measured from the onset of the presentation of the second feature. Second, the 500 ms SOA was expected to provide participants ample time to read the first feature, regardless of

length (see Cree et al., 2006). Note that equating for word length of the second feature ensured that there was no effect of reading time on decision latency.

Relatedness Ratings—Twenty-two participants not taking part in the on-line task produced off-line relatedness ratings to ensure that differences in decision latencies were not attributable to differences in perceived relatedness when time pressure was not an issue. Participants rated the relatedness of each pair (i.e., judged how well each pair “goes together in common living and/or non-living things”) on a 7-point scale, ranging from 1 (“not at all related”) to 7 (“very highly related”). This decision could be based on an estimation of how often the features co-occur, but other criteria were certainly possible, such as whether a single object possessing both features came to mind quickly, or how many different types of things possessed both features. Regardless of the strategy employed, the decision seems to necessitate some sort of Bayesian inference. There was no time limit.

Comparisons between relatedness ratings were conducted using modality (within vs. cross) as the independent variable. Modality was within participants (t_1) but between items (t_2). Cross-modal pairs were judged to be more strongly related ($M = 5.7$, $SE = 0.23$) than within-modal pairs ($M = 5.4$, $SE = 0.20$), though this difference was significant by participants, $t_1(21) = 3.62$, $p < .002$, but not by items, $t_2(38) = 0.71$, $p < .5$. This difference should lead, if anything, to the facilitation of cross-modal decision latencies.

For the speeded binary decision task, an equal number of ‘yes’ and ‘no’ trials were used to avoid biasing the response. Thus, in addition to the 40 related pairs, we constructed 20 form-form and 20 function-form pairs that could not be construed as co-occurring in common objects (e.g., *<has branches>* and *<has sharp fangs>*). For practice items, we constructed an additional 15 related and 15 unrelated pairs using visual, functional, and other types of features. For example, *<lives in aquariums>* and *<swims>* conveyed information about location and motion. No feature used in the filler or practice pairs appeared in the experimental trials.

Procedure—Participants were tested using PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993) on a Macintosh PowerMac 8600 computer, equipped with a 17-inch colour monitor. Response latencies were recorded using a CMU button box that measured the time in milliseconds between the onset of the second feature of each pair and the button press. Participants responded “yes” by pressing a button with the index finger of their dominant hand, and “no” using the index finger of their non-dominant. They received written and verbal instructions concerning how relatedness decisions were to be made, as well as examples of related (*<is crunchy>* and *<grows in gardens>*) and unrelated pairs (*<covered in felt>* and *<used in salads>*). They were instructed to silently read each feature and respond as quickly and accurately as possible as to whether the paired features were related – that is, to indicate “whether the features appear together in common living or non-living things”. There was one stimulus list, and pairs were presented in random order.

Each trial proceeded as follows. First, a fixation point (+) appeared in the center of the screen for 500ms. The fixation point was then replaced by the first feature for 500ms, after which time the second feature appeared on the line below the first one so that both were present until the participant responded. Participants received the 30 practice trials followed by the 80 experimental trials. Each session took approximately 15 minutes.

Design—The dependent variables were decision latency and the square root of the number of errors (Myers, 1979). The independent variable was modality (within vs. cross). Modality was within participants (t_1) but between items (t_2).

Results and Discussion

Mean decision latencies were significantly faster for form-form ($M = 888$ ms, $SE = 51$ ms) than for function-form pairs ($M = 1032$ ms, $SE = 50$ ms), $t_1(20) = 5.94$, $p < .00001$, $t_2(38) = 1.93$, $p = .06$. Error rates were not expected to differ because both pair-types were judged to be at least moderately related in the off-line relatedness rating task, with mean relatedness ratings greater than 5 out of 7 for both conditions. The error rates for form-form pairs ($M = .05$, $SE = .01$) did not significantly differ from that for function-form pairs ($M = .07$, $SE = .01$), $t_1(20) = 2.04$, $p > .05$, $t_2(38) = 0.77$, $p > .4$.

Despite being judged off-line as less strongly related, relatedness decision latencies were shorter for within-modal pairs. Because numerous variables including strength of correlation were equated, and differences in off-line relatedness ratings favoured cross-modal pairs, this suggests a cross-modal advantage if anything. Instead, the reverse pattern was found. These results are consistent with the assumption of a deep integration hierarchy in which modally distributed information is first integrated in single-modality convergence zones, which then feed into cross-modal convergence zones. The latency advantage for within-modal items is inconsistent with the assumption of a shallow integration hierarchy and with the predictions of amodal models, both of which predict no modality effect.

One potential concern regarding Experiment 1 is that, although the length of the second feature was controlled, multiple word stimuli are potentially problematic for two reasons. The form features frequently began with the five-character (including the space between words) phrase “*has a*”, whereas the functional features frequently began with the 8-character phrase “*used for*”. This introduces two problems, the first of which concerns timing. The tendency for functional features to begin with longer phrases potentially produces longer reading times on average for these features. Although this was not believed to be a problem for Experiment 1 because the SOA was chosen to exceed the expected reading time for the first feature, it is possible that the SOA did not provide enough time to read some of the initially-presented features in cross-modal pairs. A second potential issue is that initial phrases are repeated between within-modal features for 12 of 20 pairs (<*has fins*>, <*has gills*>), although the remaining 8 contained different initial phrases (<*has a lid*>, <*made of glass*>). Initial phrases were never repeated in cross-modal pairs. Experiment 2 addresses these potential issues in two ways. First, the same features were used as targets for both the within- and cross-modal pairs, thus automatically controlling for a number of variables. Second, multiword items such as <*has a handle*> were divided into stem (*has a*) and content (*handle*) components which were presented sequentially. Decision latency was measured from the onset of the content component, nullifying any potential advantage of repeated stems within a pair.

Experiment 2

Experiment 2 replicated Experiment 1 using a more rigorously controlled set of items, and a modified presentation paradigm. We expected to find the same results, which supported hierarchically deep models.

Method

Participants—Thirty-eight University of Western Ontario undergraduates received either course credit or \$10 for their participation.

Materials—We selected 18 related form-form and 18 related function-form pairs (see Appendix B). The pairs were yoked to create function-form-form triples, such that one form feature was the second feature of both a within- and cross-modal pair. For example, <*used*

for storing food> <*has windows*> <*has doors*> is a triplet containing a form-form (<*has windows*> <*has doors*>) and a function-form pair (<*used for storing food*> <*has doors*>).

The pairs comprising each triplet were selected to be as similar as possible with respect to several variables expected to influence decision latencies. In addition to the variables identified in Experiment 1 as potential influences on decision latencies, the pairs were matched on distinctiveness, which is the inverse of the number of concepts in which the feature appears in our norms, and thus indexes the likelihood with which the feature could cue a particular basic level concept. Because stem and content words were presented separately and we measured decision latencies from the onset of second feature's content word, content word length influences reading times. Therefore, content word length in characters, including spaces, was equated. In addition, each group had similar distributions of higher ($r^2 > .25$), medium ($.15 \leq r^2 \leq .25$) and lower ($r^2 < .15$) correlated pairs, such that the cross-modal group had 10, 3, and 5 and the within-modal group had 9, 3 and 6 higher, medium and lower correlated pairs, respectively. One of the 18 yoked triplets was dropped from the analyses that follow because, although judged to be moderately related in the offline relatedness rating task ($M = 3.1$ out of a possible score of 7), more than 40% of participants judged the within-modal pair (<*has a motor*> <*has sails*>) to be unrelated in the speeded task. Item characteristics for the remaining 17 sets of yoked feature pairs are summarized in Table 2. Because the second features were identical for the within- and cross-modality groups, the groups are automatically matched on all variables concerning them. Importantly, within- and cross-modal pairs did not differ significantly on any variable, other than the number of concepts in which the features appeared together, and this difference favoured cross-modal pairs.

Relatedness Ratings—Forty University of Western Ontario students not participating in the main task provided off-line relatedness ratings. Each participant rated half of the within- and half of the cross-modal pairs, and saw only one pair from each triplet. The procedure and analyses were identical to the relatedness ratings in Experiment 1.

As in Experiment 1, the perceived relatedness of the cross-modal pairs ($M = 4.3$, $SE = 0.1$) was greater than that of the within-modal pairs ($M = 4.1$, $SE = 0.1$), and this was significant by participants, $t_1(39) = 2.64$, $p < .02$, but not by items, $t_2(16) = 0.39$, $p > .7$. Again, this difference would facilitate decision latencies of cross-modal pairs, if anything.

The sets of yoked pairs were divided equally between two experimental lists, such that one list contained the within-modal pairs for half of the sets, and the cross-modal pairs for the other half. The remaining items appeared in the second list. An equal number of unrelated filler pairs (nine form-form, and nine function-form) were constructed using features not appearing among the related pairs, and were used in both lists.

All feature names were divided into stem and content components. The stem comprised the initial segment and included prepositions, conjunctions, and verbs (*has a*, *used for*, *is*). The content component comprised the final one or two words that carried much of the feature's unique meaning. Because experimental and filler pairs were matched with respect to the number of times each stem appeared, $\chi^2(8) = 11.21$, $p > .15$, and an equal number of within- and cross-modal pairs appeared among the experimental and filler items, the stems did not cue the response. Participants therefore needed to wait for and process the content component to make the relatedness judgment, and response latencies were measured with respect to the onset of the content component of the second feature.

Procedure—Participants were tested using E-Prime (Psychology Software Tools Inc., 2002) on an AMD Athlon 64 3200+ personal computer, equipped with a 17-inch colour monitor. The instructions were identical to Experiment 1.

Each trial proceeded as follows: First, a blank white screen was presented for 2000 ms, followed by a vertically and horizontally centered fixation cross ('+') for 250 ms. For purposes of displaying the stem and content components of the first and second features, the screen was divided into vertically and horizontally centered quadrants, though the quadrant boundaries were not visible (see Figure 4). Each quadrant was justified opposite to its position (i.e., the top-left quadrant was lower-right justified, etc.). The left quadrants were used for presenting the stem (*used by*) and the right quadrants were used for presenting the content word(s) of each feature name. The first feature's stem was presented in the top left quadrant immediately following the removal of the fixation cross. After 300 ms, its content component (*riding*) was then presented in the top right quadrant so that the first feature (*used by riding*) was displayed in the top half of the screen, centered on the boundary between stem and content. After 700 ms, the second feature's stem (*has*) appeared in the lower left quadrant. Finally, after 300 ms, its content (*handlebars*) appeared in the lower right quadrant, so that the second feature (*has handlebars*) was displayed in the lower half of the screen immediately below the first. Both features remained on the screen until the participant responded. The use of quadrants avoided cuing the length (and possibly the identity) of the content words, and allowed them to always appear in the same screen position. Decision latencies were recorded using a button box that measured with millisecond accuracy the time between the onset of the presentation of the second feature's content component and the button press.

Participants received five lead-in trials immediately followed by 36 experimental trials. The experiment took less than 15 minutes to complete.

Design—Analyses of variance were conducted using participants (F_1) and items (F_2) as random variables. The dependent variables were decision latency and the square root of the number of errors (Myers, 1979). The independent variable was modality (within vs. cross), which was within both participants and items. List was included as a between-participants dummy variable and item rotation group as a between-items dummy variable to stabilize variance that may result from rotating participants and items over lists (Pollatsek & Well, 1995). Effects involving these dummy variables are not reported.

Results and Discussion

Incorrect trials and those with decision latencies greater than 3 standard deviations above the grand mean were removed from the analysis (3% of the trials). As indicated previously, one related pair (<*has a motor*> <*has sails*>) had an error-rate over 40%. Presumably this occurred because the motor on vehicles such as sailboats is not at all salient for people without first-hand experience, and so participants considered these features to be mutually exclusive. Thus, we excluded it and its corresponding yoked pair (<*used for cruising*> <*has sails*>) from the analyses.

For the remaining 17 pairs, relatedness decision latencies were greater for cross-modal ($M = 1087$ ms, $SE = 44$ ms) than for within-modal pairs ($M = 1013$ ms, $SE = 37$ ms), which was significant by participants, $F_1(1, 36) = 10.16$, $p < .004$, but not by items, $F_2(1, 15) = 2.36$, $p > .1$. Participants were highly accurate, and within-modal ($M = .07$, $SE = .01$) and cross-modal error rates ($M = .09$, $SE = .02$) did not differ, both F 's < 1 .

The results of Experiments 1 and 2 suggest that multimodal feature representations activate one another relatively quickly in a hierarchically deep integration structure. Stimuli were

selected to be as closely matched as possible on several factors that are assumed to influence relatedness decision latencies, and any differences tended to favour the cross-modal items. Nonetheless, there was an advantage for within-modal pairs in both experiments.

One limitation of Experiments 1 and 2 is that an advantage was found for within-modal pairs. This introduces a potential issue in that cross-modal processing may involve some form of switching and thus incur a performance cost, reflected in longer decision latencies for cross-modal pairs. Modality-switch costs have been demonstrated in the perception literature (Spence, Nicholls, & Driver, 2000), and in the concepts literature using feature verification (Pecher, Zeelenberg, & Barsalou, 2003). Thus, it is unclear whether the within-modal advantage in Experiments 1 and 2 results from a deep integration hierarchy or modality switching costs. To alleviate this concern, Experiment 3 used a task in which deep models predict a latency advantage for cross-modal trials.

A second potential issue is that functional features appeared only in cross-modal pairs, and it is possible that functional information takes longer to access. For example in Barsalou's (1999) perceptual symbol systems account, retrieval of an object's function (*<used for opening cans>*) would involve a situated perceptual simulation, or mental re-enactment of the sensorimotor elements of that function (perhaps visualizing the process of grabbing a can, applying a can-opener to the can, and then turning the key that causes the opener to follow and cut the rim of the can). Because function typically unfolds over time, it is reasonable to predict that retrieval of this information takes more time than retrieving relatively static form information, such as an object's shape (though retrieval of some form information in the perceptual symbol systems framework may involve a mental rotation from an object's canonical orientation, Edelman & Bühlhoff, 1992). Though function features were always presented first, it is possible that they imposed additional processing demands that carried over during the processing of the second feature.

The within-modal speed advantage observed in Experiments 1 and 2 could therefore be interpreted as an advantage for pairs involving only form features over pairs involving functions. One reasonable solution would be to carry out an experiment contrasting within-modal function-function correlated pairs with cross-modal form-function pairs. While designing such an experiment, however, we discovered that the nature of the functional features appearing in the norms made it impossible to use a similar empirically-based approach to item selection. The first problem is that many functional features tend to participate in the same relationships, and it is therefore difficult to select a controlled set of items of sufficient size without including some features multiple times. The second, and more theoretically troublesome, problem is that many functions appearing in the norms could be argued to be semantically similar. For example, in the pair *<used for hitting>* and *<used for hammering>*, hammering could be seen as a type of hitting. A within-modal advantage found using such items would allow the alternative explanation that it arose from the greater semantic similarity between these pairs compared to cross-modal form-function pairs. Because of these issues, it was impossible to create a sufficient set of items. Therefore, we did not conduct such an experiment.

A final issue concerns whether these results extend to other representational modalities. Because only visual form and functional features were used in Experiments 1 and 2, it remains unclear whether similar results would emerge using other modalities. Experiment 3 was designed to deal with all of these concerns.

Experiment 3

People possess a rich knowledge of many objects that spans multiple representational modalities. Accordingly, when identifying an object, people generally do so using only a

subset of the information they possess about it, and yet they are able to retrieve other knowledge about it. For the sake of brevity, we use the term, “concept activation” as a short form for activating concepts given one or more features. One way to think about this process is in terms of pattern completion. In a system using distributed representations, an object’s identity and relevant features can be deduced from a subset of features through pattern completion. For example, if one determines that an object *<has a blade>* and *<has a handle>*, a number of other features may be inferred; the blade of the object is probably *<made of metal>* and *<used for cutting>*. Over time, the pattern of activated features comes to resemble the representation for some class of objects, allowing the object to be categorized. In this example, one may coarsely identify the object as some sort of manipulable tool, or more precisely as some sort of knife. As with the feature inference task, shallow- and deep-hierarchy assumptions make different predictions about the time course of within- and cross-modal concept activation.

Figure 5 illustrates a simplified hierarchically shallow model of a concept, consisting of just six features distributed across two modalities. We assume that concept identification requires that some proportion of the concept’s features reach an activation threshold. Suppose in the first case that two features from the same modality are presented, as when a person is told that an object *<has a handle>* and *<has a blade>* (two form features, Form A and Form B, as in Figure 5a). Activation from the two feature units spreads in the first processing step to the common convergence zone, and to all associated features in the second step. In this way, the concept, or a reasonable portion of it, is activated. The cross-modal case, illustrated in Figure 5b, is similar. For example, a person might be told that an object *<has a blade>* (form) and is *<used for cutting>* (function). Again, two processing steps are required to activate the remaining features in a shallow model. Activation of the two feature units spreads to the common convergence zone in the first step, and then to all associated features in the second. Thus, regardless of whether the presented information corresponds to one or multiple modalities, shallow hierarchy models predict the same number of processing steps, and presumably therefore the same amount of time, to activate a concept.

Figure 6 illustrates a simplified hierarchically deep model. Again, in the first scenario, two features from the same modality are presented (two form features, Form A and Form B, as in Figure 6a). After one processing step, activation has spread to the modality-specific convergence zone, and after two steps, it has begun to activate the remaining within-modal feature, Form C, and the top-level cross-modal convergence zone. It requires two additional processing steps, as activation passes first to the function convergence zone, and then to the three function features, Function A, B and C, before all of the concept’s features are activated.

Consider now the case in which a cross-modal pair is presented (Form A and Function A; Figure 6b). After one step, activation has spread from both features to their respective modality-specific convergence zones. After the second step, both within-modal convergence zones pass activation to the cross-modal convergence zone. In addition, each within-modal convergence zone passes activation back to correlated units within their respective modalities. Thus, in the hierarchically deep model, cross-modal feature pairs activate more of the network faster than do within-modal feature pairs because each within-modal convergence area allows the parallel activation of clusters of correlated features within its modality.

Finally, lacking any modality information, amodal models predict no within- versus cross-modal speed advantage, all other factors being equal.

Using knowledge of within- or cross-modal features to retrieve concepts is a common event and is therefore an interesting test paradigm for a couple of reasons. First, like feature inference, it is a basic conceptual process in which we often engage, and therefore, elucidating the manner in which it occurs is of importance. Second, and most importantly in the present context, whereas deep hierarchy models predict a within-modal advantage for feature inference, they predict a cross-modal advantage for activating a concept from partial information. Finding a cross-modal advantage would therefore address the possible explanation for Experiments 1 and 2 that processing cross-modal information may be generally disadvantaged.

Experiment 3 tests the predictions for hierarchically deep and shallow models for a task involving concept activation from incomplete information. Moreover, we used functional features in both within- and cross-modal conditions, and features from other knowledge types (smell, taste, sound, etc.). A dual feature verification task was used wherein two sequentially presented features were followed by a concept name. Under the assumption that verbally presented features activate other features prior to the presentation of a concept name, and that this activation facilitates feature verification, the hierarchically deep and shallow models predict different patterns for concepts preceded by within- and cross-modal pairs. Hierarchically shallow models predict no difference between within- and cross-modal decision latencies, whereas hierarchically deep models predict a cross-modal advantage because two features should activate clusters of correlated features for two modalities in parallel. Amodal models again predict no effect of modality.

McRae et al.'s (1999) investigation of the role of feature correlations in concept activation suggests that the advantage for cross-modal pairs predicted by hierarchically shallow models will rely on clusters of intercorrelated features in each modality in which activation may occur in parallel. Accordingly, we divided the target concepts into those with low and high intercorrelational density, which indexes the degree to which a concept's features are intercorrelated. Intercorrelational density is the sum of the percentage of shared variance across all of a concept's significantly correlated feature pairs (because it is a sum, it is no longer truly a percentage), as calculated from McRae et al.'s (2005) norms. We defined low and high density concepts as those with a standardized intercorrelational density less than -1 and greater than 1 , respectively. Although effects of target density do not distinguish among shallow, deep, or amodal theories, because high density concepts possess more cohesive clusters of intercorrelated features on which concept activation depends, this factor was included in the analyses. All feature-based models that allow for influences of correlated features predict a latency advantage for high-density items. However it is additionally possible within hierarchically deep models that target density might interact with pair type, such that greater facilitation may be found for high-density than for low-density concepts, whereas shallow and amodal models do not make this prediction.

Method

Participants—Twenty-six University of Western Ontario undergraduates received \$10 for their participation. Three additional participants were dropped because their mean response latency (2 participants) or error rate was greater than three standard deviations above the grand mean.

Materials—Standardized intercorrelational density was calculated for all concepts in the norms. Thirty-six sets of the form {<A> <C> TARGET} were selected such that {<A> } were correlated within-modal features (<has a sheath> <has a blade>), {<C> } were correlated cross-modal features (<used for cutting> <has a blade>), and {<A> <C>} were listed for the target concept (*sword*) in the norms (see Appendix C). Of these 36

item sets, half contained low-density and half contained high-density targets. A between-items analysis of target concept density confirmed that the mean intercorrelational density of low density targets ($M = 95$, $SE = 13$) was significantly lower than that of high density targets ($M = 635$, $SE = 58$), $t(34) = 9.06$, $p < .00001$.

To control for factors that might influence decision latencies, items were selected such that characteristics of the first feature, and the relationships between it and the second feature and target concept were equated (see Table 3). As described below, two low density items were dropped because of error rates of approximately 50% on both within- and cross-modal trials, so the equated statistics in Table 3 are for the remaining items only. Items were pairwise matched on length in characters (including spaces), percent of shared variance, the number of concepts in which the features occur together, and the number of concepts per feature. Because the features were associated with particular concepts, items were additionally matched on production frequency (the number of people who listed each feature for that concept) to control for the likelihood of the feature bringing to mind the target concept (Ashcraft, 1978). Additionally, because the initially presented features are assumed to activate the target concept by activating its features, and the shared variance among these features influences the speed with which this occurs (McRae et al., 1997), we matched items on intercorrelational strength, which is the sum of the shared variance between the presented features and other features of the concept. This ensured that differences in feature verification latencies were not attributable to differences between the groups with respect to the degree to which the initially presented features are correlated with other features of the concept. Because they are common to both members of a yoked pair, relations between the second feature and target concept were automatically equated. All knowledge types in the norms, with the exception of encyclopaedic and taxonomic features, were represented. Target concepts included both living and non-living concrete objects.

The 36 yoked within- and cross-modal items were assigned pseudo-randomly to two lists, such that each included 18 within- and 18 cross-modal trials, and no two yoked items appeared in the same list. Because each member of a yoked pair shared a target concept, half of the items in each list had high-density targets, and half had low-density targets. Both lists contained 36 filler items in which it was not true that both features were true of the target concept. The filler items were divided into thirds in which either the first feature, the second, or neither was true of the target concept. Filler items used features and medium-density target concepts not appearing among the experimental items. The use of correlated features in the filler items ensured that there was at least one concept in which both features occurred for both experimental and filler trials, and therefore participants needed to wait for the presentation of the target concept to respond accurately. No feature or target concept appeared more than once in either list.

Procedure—Participants were tested using E-Prime (Psychology Software Tools Inc., 2002) on an AMD Athlon 64 3200+ personal computer, equipped with a 17-inch colour monitor. Each trial proceeded as follows. First, a blank white screen was presented for 1500 ms, followed by a vertically and horizontally centered fixation cross ('+') for 500 ms. The first feature was displayed immediately above the position of the fixation cross, followed 1000 ms later by the second feature immediately below the first. Both features remained on the screen for an additional 1000 ms, after which time the target concept was presented in upper case letters on the line immediately below the second feature. The 1000 ms SOA for each feature ensured that participants had sufficient time to read the longest multiword features. The two features and the target concept remained on the screen until the participant responded. Responses were collected using a button-box that recorded the time between the onset of the target concept and the button press with millisecond accuracy. Participants

responded “yes” by pressing a button with the index finger of their dominant hand and “no” by pressing another button using the index finger of their non-dominant hand.

There were 16 lead-in practice trials comprising a mixture of yes and no trials immediately followed by 72 experimental trials. The experiment took about 15 minutes to complete.

Design—Analyses of variance were conducted with decision latency and the square root of the number of errors as the dependent variables. The independent variables were modality (within vs. cross) and concept density (low vs. high). Modality was within participants (F_1) and items (F_2), whereas density was within participants but between items. List and item rotation group were again included as dummy variables.

Results and Discussion

Two low-density items with error rates approaching 50% were dropped from the analysis. Because these items' error rates were similar for within- and cross-modal conditions, the errors were attributed to ambiguity of, or unfamiliarity with, the relationship of second feature to the target concepts (*beets* and *screws*). Decision latencies greater than 3 standard deviations above the grand mean were replaced by the cutoff value (4% of the trials).

Mean verification latencies and error rates are presented in Table 4. Critically, feature verification latencies were shorter for cross-modal ($M = 773$ ms, $SE = 29$ ms) than for within-modal items ($M = 849$ ms, $SE = 34$ ms), $F_1(1,24) = 15.68$, $p < .0007$, $F_2(1,30) = 7.83$, $p < .009$. Verification latencies were shorter for high-density ($M = 770$ ms, $SE = 29$ ms) than for low-density concepts ($M = 851$ ms, $SE = 33$ ms), which was significant by participants, $F_1(1,24) = 26.80$, $p < .0002$, but marginal by items, $F_2(1,30) = 3.46$, $p < .08$. Concept intercorrelational density did not interact with feature modality, $F_1(1,24) = 2.80$, $p > .1$, $F_2 < 1$.

Because it was hypothesized that a greater cross-modal advantage might be apparent for high-density items than for low-density items, planned comparisons were conducted between the modality conditions for both levels of density. For low density targets, the 52 ms advantage for cross-modal pairs was significant by participants, $F_1(1,45) = 4.62$, $p < .04$, but not by items, $F_2(1,30) = 2.11$, $p > .1$. For high density targets, the 99 ms advantage for cross-modal pairs was significant, $F_1(1,45) = 17.42$, $p < .0003$, $F_2(1,30) = 6.46$, $p < .02$. Thus, it appears that cross-modal pairs more quickly activated concepts than did within-modal pairs for concepts with relatively few and many clusters of intercorrelated features, though this difference was numerically but not statistically greater for higher density concepts.

Participants were generally quite accurate. There was no difference in error rates between within and cross modal pairs, $F_1 < 1$, $F_2(1,30) = 1.02$, $p > .3$. Error rates were marginally lower for high density than for low density concepts, $F_1(1,24) = 3.44$, $p < .08$, $F_2(1,30) = 3.57$, $p < .07$. The two factors did not interact, $F_1(1,24) = 1.48$, $p > .2$, $F_2 < 1$.

One potential concern was that some within-modal feature pairs might be construed as redundant for some items. For example, it is perhaps the case that little additional information is provided in pairing <eaten in sandwiches> with <is edible>, as the former implies the latter. One alternative explanation of these results, therefore, is that people are slower at responding to items for which less information is provided. The above analyses were repeated with the removal of seven yoked pairs of items identified as including features with potentially overlapping meaning. A reanalysis of the item characteristics found the stimuli to be matched on all variables of interest, with the exception of the feature length (in characters) for low-density items, which favoured within-modal ($M = 10.2$, $SE = 0.8$)

over cross-modal items ($M = 14.4$, $SE = 1.5$), $t(12) = 2.29$, $p < .05$. Despite this bias, feature verification latencies were again shorter for cross-modal ($M = 779$ ms, $SE = 31$ ms) than for within-modal items ($M = 875$ ms, $SE = 33$ ms), $F_1(1,24) = 24.29$, $p < .0001$, $F_2(1,23) = 10.14$, $p < .004$. Verification latencies were shorter for high-density concepts ($M = 794$ ms, $SE = 32$ ms) than for low-density concepts ($M = 860$ ms, $SE = 32$ ms), which was significant by participants, $F_1(1,24) = 12.43$, $p < .002$, but not by items, $F_1(1,23) = 1.45$, $p > .2$. Concept intercorrelational density interacted with feature modality by participants, $F_1(1,24) = 7.00$, $p < .02$, but not by items, $F_2 < 1$. Error rates did not differ between conditions, nor was there an interaction in any analysis, all $F_s < 1$.

Finally, planned comparisons showed that for low density targets, the 55 ms difference (within-modal: $M = 887$ ms, $SE = 33$ ms; cross-modal: $M = 832$ ms, $SE = 32$ ms), was marginal by participants, $F_1(1,24) = 4.05$, $p < .06$, and nonsignificant by item, $F_2(1,23) = 2.35$, $p > .1$. In contrast, for high density targets, there was a significant 139 ms difference (within-modal: $M = 864$ ms, $SE = 33$ ms; cross-modal: $M = 725$ ms, $SE = 36$ ms), $F_1(1,25) = 25.86$, $p < .0001$, $F_2(1,23) = 9.26$, $p < .006$. Thus, cross-modal pairs again more quickly activated concepts than did within-modal pairs, particularly for relatively densely intercorrelated concepts.

One final thing to note is that there are other ways we could have measured the relationship between the feature pairs and the clusters of intercorrelated features within target concepts. The item groups were based on the intercorrelational density of the target concept, an index of the correlated feature clusters within the concept. However, McRae et al. (1997;1999) showed that intercorrelational strength (i.e., the sum of the squared correlations between a feature and each other feature of a particular concept) influences feature verification latencies. For example, *<hunted by people>* is more strongly intercorrelated with other features of *deer* than of *duck*, and verification latency was shorter for *deer*. Because the cross-modal advantage for concept activation depends on these features activating feature clusters that overlap with the concept, the data were reanalyzed using two alternative methods. First, we classified items into high and low intercorrelational strength on the basis of the mean intercorrelational strength of the presented features, using the same logic underlying McRae et al. (1997;1999). The second method incorporated both intercorrelational strength and intercorrelational density by multiplying these two values together to create a composite variable. For both measures, the placement into high and low groups remained the same for the vast majority of items, and consequently, the pattern of results after re-analyses of the verification latency data did not qualitatively differ from those presented above.

Experiment 3 advances our understanding of how concept activation is influenced by the sensorimotor modality of available object features by providing novel insight into the underlying neural mechanisms that allow integration of distributed semantic representations into coherent concepts. Importantly, there was a clear cross-modal latency advantage that is predicted by deep hierarchy models. These results are inconsistent with the predictions of shallow hierarchy and amodal models. Although modality and density did not reliably interact, the cross-modal advantage was somewhat stronger for high density concepts.

Experiment 4

As explained above, deep hierarchy models predict modality effects only for tasks for which processing speed should be the primary determinant of performance. With additional processing time, overall levels of semantic activation for both modality conditions should approach the same levels and furthermore allow participants to employ strategic processing. Experiment 3 used a 1000 ms SOA between the first and second feature, and between the

second feature and the target concept to ensure that participants would have enough time to read the multiword features. Though the various item counterbalancing measures and the nature of the task should have reduced the efficacy of any participant strategies, it could nonetheless be argued that the 1000 ms SOA used in Experiment 3 might be a bit long. Therefore, Experiment 4 replicated Experiment 3 using 500 ms SOAs.

Method

Participants—Thirty-six University of Western Ontario undergraduates participated for course credit, or received \$10 for their participation. Three were dropped because their mean response latency was greater than three standard deviations above the grand mean.

Materials—The materials were identical to those used in Experiment 3.

Procedure—Experiment 4 used the same procedure as that used in Experiment 3, with the exception of a 500 ms SOA between the first and the second feature, and between the second feature and the target concept.

Results and Discussion

One low-density item (*beets*) with an error rate approaching 50% was dropped from the analyses, and decision latencies greater than 3 standard deviations above the grand mean were replaced by the cutoff value (5% of the trials). The remaining items differed from those summarized in Table 4 by only one item, and thus the modality conditions remained matched on all characteristics.

Mean feature verification latencies and error rates are presented in Table 5. Again, verification latencies were shorter for cross-modal ($M = 792$ ms, $SE = 33$ ms) than for within-modal items ($M = 857$ ms, $SE = 37$ ms), $F_1(1,34) = 9.33$, $p < .005$, $F_2(1,31) = 4.36$, $p < .05$. Verification latencies were shorter for high-density ($M = 815$ ms, $SE = 30$ ms) than for low-density concepts ($M = 877$ ms, $SE = 31$ ms), which was significant by participants, $F_1(1,34) = 8.99$, $p < .006$, but not by items, $F_2(1,31) = 1.94$, $p > .1$. Concept intercorrelational density did not interact with feature modality, $F_1(1,34) = 4.12$, $p < .06$, $F_2 < 1$.

Planned comparisons showed that for low density targets, the 37 ms advantage for cross-modal pairs was not significant, $F_1(1,58) = 2.17$, $p > .1$, $F_2 < 1$. For high density targets, the 92 ms cross-modal advantage was significant, $F_1(1,58) = 13.42$, $p < .0007$, $F_2(1,31) = 4.63$, $p < .04$. Thus, we replicated the finding that, although there was no significant interaction, the modality effect was somewhat stronger for higher density concepts.

Participants were quite accurate. There was no difference in error rates between within and cross modal pairs, $F_1(1,34) = 3.12$, $p > .08$, $F_2(1,31) = 1.68$, $p > .2$. Error rates did not differ between high and low density concepts, $F_1(1,34) = 2.13$, $p > .1$, $F_2(1,31) = 2.08$, $p > .1$. The interaction was nonsignificant, $F_1(1,34) = 3.28$, $p < .08$, $F_2 < 1$.

We also reanalyzed the data after excluding the seven items identified in Experiment 3 as including features with potentially overlapping meaning. The overall pattern of results was identical to that of Experiment 3. Feature verification latencies remained shorter for cross-modal ($M = 802$ ms, $SE = 32$ ms) than for within-modal items ($M = 876$ ms, $SE = 39$ ms), $F_1(1,34) = 8.75$, $p < .007$, $F_2(1,24) = 4.57$, $p < .05$. Verification latencies remained shorter for high-density ($M = 823$ ms, $SE = 35$ ms) than for low-density concepts ($M = 855$ ms, $SE = 34$ ms), again significant by participants, $F_1(1,34) = 4.86$, $p < .04$, but not by items, $F_2 < 1$.

Concept intercorrelational density interacted with feature modality by participants, $F_1(1,34) = 9.33, p < .005$, but not by items, $F_2(1, 24) = 2.14, p > .1$.

Finally, planned comparisons showed that for low density targets, there was a nonsignificant 20 ms difference (within-modal: $M = 865$ ms, $SE = 38$ ms; cross-modal: $M = 845$ ms, $SE = 36$ ms), both F s < 1 . In contrast, for high density targets, there was a significant 129 ms difference (within-modal: $M = 887$ ms, $SE = 44$ ms; cross-modal: $M = 758$ ms, $SE = 32$ ms), $F_1(1,62) = 17.49, p < .0002, F_2(1,24) = 6.60, p < .02$. Thus, cross-modal pairs again more quickly and/or fully activated concepts than did within-modal pairs, particularly for relatively densely intercorrelated concepts.

General Discussion

The present research used complementary behavioural tasks to investigate the neural architecture underlying integration of multimodal semantic representations. In Experiments 1 and 2, using a feature relatedness task, we found a within-modal advantage that is predicted by hierarchically deep models. However, necessary aspects of these experiments' design yielded open issues. The first concern was that functional features, which may take longer to retrieve, appear only in the slower cross-modal condition, and so it was unclear whether the results reflect a general disadvantage for processing functional features. The second was that cross-modal items may have incurred a modality-switching cost and correspondingly longer decision latencies. Finally, we wished to extend the results to knowledge types other than visual form and function. Experiments 3 and 4 addressed all three of these concerns using a double feature verification task, and again the results supported a deep integration hierarchy.

Theoretical Implications and Relation to Existing Literature

The results across the four experiments lead us to make two major claims about the manner in which semantic information is organized and integrated. We discuss how each of these claims relates to existing models and empirical findings in turn.

Integration occurs in a deep hierarchy—The first claim is that multimodal semantic integration occurs in multiple convergence zones organized in a deep integration hierarchy. The results across tasks support this claim and are thus clearly compatible with models such as those of Damasio (1989) and Simmons and Barsalou (2003) that explicitly specify such an organization. In deep hierarchy models, higher-level multimodal convergence zones receive input from lower-level within-modal convergence zones that are proximal to the sensorimotor representational regions that they integrate. We speculate that within-modal convergence zones may correspond to the anterior shift described by Thompson-Schill (2003), which is the association between conceptual processing and activity in brain regions just anterior to perceptual areas. Thus, this phenomenon may reflect activity among focused integration units as they provide re-entrant activation to perceptual areas in what Barsalou (1999) would call a simulation of the concept.

We explored two consequences of this organization. First, because information encoded by modality-specific representations is integrated in proximal modally-tuned convergence zones, this facilitates the rapid spread of activation within a modality. For a task relying primarily on feature-level processing (in contrast with concept-level processing), this implies a speed advantage for processing within-modal feature pairs.

Our results are consistent not only with prototypical deep hierarchy models, but also with Plaut's (2002) graded distance-only model because it too explicitly claims that within-modal integration uses proximal integration units. One feature of Plaut's model is that within-

modal stimulus pairs generate more similar representations, in terms of overlapping ‘semantic’ hidden units, than do cross-modal stimulus pairs. Experiments 1 and 2 are essentially semantic priming tasks; thus, because semantic priming is largely driven by similarity, the model may be able to simulate our findings. As discussed in the Introduction, distance can be measured both spatially and in terms of number of connections. Because the human brain is biased to form short connections (Jacobs & Jordan, 1992), longer distances (such as those that form cross-modal connections in Plaut’s model) likely correspond to longer chains of connections. Nodes along this chain are unlikely to be simple signal repeaters, and more plausibly integrate information and pass it forward, which may be functionally equivalent to introducing convergence zones. Therefore, Plaut’s distance model can be viewed as a form of a deep hierarchy model. As implemented, longer (weaker) connections to cross-modal convergence zones (i.e., those units connecting different modalities) should lead to faster within-modal activation of single units as in Experiments 1 and 2.

The second property of a deep hierarchical organization we explored is that, because information from multiple representational modalities is integrated within distal multimodal convergence zones which take as their inputs the output of earlier within-modal convergence zones, passing information across modalities takes longer than within modalities. For a task relying primarily on processing concepts with representations spanning multiple sensorimotor modalities, we observed a predicted speed advantage for processing cross-modal feature pairs.

The cross-modal advantage for dual feature verification in Experiments 3 and 4 appears, at first glance, to be inconsistent with Barsalou, Pecher, Zeelenberg, Simmons, and Hamann (2005), who found a within-modal advantage for dual-feature verification. However, a closer examination of subtle but important differences in item presentation reveals how the results are indeed completely consistent. In Barsalou et al., concept and property names were presented simultaneously, with the concept name appearing where the fixation stimulus had been, and feature names appearing beside one another four lines below. This presentation method encourages participants to first process the concept name, and then the features. In Experiments 3 and 4, the target concept was presented after the sequentially presented features. Thus, although all experiments used dual feature verification tasks, they reflect different processes. Feature verification latencies in our task depended on the speed with which both features activated other features of the concept to (partially) complete its pattern, and as described earlier, deep hierarchy models predict shorter latencies for cross-modal items. We believe that Barsalou et al.’s task actually involves the same sort of processing that underlies our feature inference task in Experiments 1 and 2. In their task, we assume that processing the concept name partially activates both of the features before they have been read. Furthermore, as in our Experiments 1 and 2, reading one feature should further activate the second before it is read. Because deep hierarchy models predict faster within-modal feature-to-feature activation, and verification latencies in Barsalou et al.’s task depended on the speed with which both features were processed, we believe not only that our cross-modal advantage for dual feature verification can be reconciled with Barsalou et al., but also that their results further support our account of the within-modal advantage in Experiments 1 and 2.

We note a number of explicitly shallow hierarchy models that have been successfully used in the past to make various theoretical arguments about semantic organization, and thus merit discussion in light of the support we find for deep hierarchy models. Humphreys and Forde’s (2001) HIT model was designed to capture some of the properties of Farah and McClelland’s (1991) fully interconnected network in a hierarchical model in which processing from a top-level layer cascades back to earlier functional and sensory

representational units. Both are shallow models as presented, and would therefore need to be extended. Farah and McClelland's model was designed to show that category specific deficits can arise out of a segregation of sensory and functional knowledge, and was not intended to be an argument about other aspects of the underlying neural architecture. The HIT model was designed to confirm that the same patterns hold in a system that allows reciprocal activation (or "cascading") from a higher-level structure that acts as a convergence zone. Their model included only three types of knowledge: structural descriptions of the general visual form of objects, functional and inter-object associative information that they called "semantic" knowledge, and name representations. The main focus of Humphreys and Forde's argument was the importance of the role of reentrant activation from higher-level integration sites, and the role of the various sensorimotor modalities in the model was secondary. Thus, they sketched alternative versions illustrating the role of cascading activation in different tasks relying primarily on different modalities, for example, the role of auditory knowledge when identifying a guitar by its characteristic sound. The present results show how the mechanism of cascading reentrant activation might work in a deep hierarchy model capable of accounting for the phenomena that the HIT model was designed to address, and further refines the model by demonstrating that this cascading may function differently within and across modalities.

Whether the present results are consistent with Patterson et al.'s (2007) semantic hub theory depends on precisely how their claims are interpreted. The most straightforward interpretation of Patterson et al. appears to be that the anterior temporal lobe is the sole convergence zone through which semantic memory is routed. As the authors state when contrasting their view with that of Damasio and colleagues who have argued for multiple convergence zones, "By contrast, the distributed-plus hub view proposes that, in addition to direct neuroanatomical pathways between different sensory, motor and linguistic regions, the neural network for semantic memory requires a single convergence zone or hub that supports the interactive activation of representations in all modalities, for all semantic categories." (p. 977). Thus, this reading of their model corresponds to a shallow hierarchy model and is inconsistent with the present results. Note, however, that our results have no bearing on Patterson et al.'s claim that the anterior temporal lobe is the top-level multimodal convergence zone, and is critical for integrating semantic information.

Under another reading of Patterson et al. (2007), our results are consistent with their semantic hub model. If pathways between modalities implement information integration, and thus act as convergence zones, this would more clearly cast it as a deep-hierarchy model, distinguished from other such models by Patterson et al.'s commitment to a specific anatomical location for the top-level convergence zone.

Finally, the present results are inconsistent with our own implemented models. For example, Cree et al. (2006) simulated influences of distinctive versus shared semantic features in a connectionist model that used direct connections among its feature units. That is, there were no convergence zones (which could be implemented as sets of hidden units in connectionist models). In fact, our implemented models have not even instantiated modality-specificity, though we have argued elsewhere that sensorimotor modality is an important organizing force in how the brain represents information (Cree & McRae, 2003; McRae, 2004). Thus, the present results suggest that our future models should include intramodal hidden unit clusters that feed into a higher level convergence zone.

Finally, it is difficult to see how theories of temporal-based binding mechanisms, which argue that features across various modalities are bound by virtue of firing in synchrony rather than in convergence zones ("combination coding cells"; von der Malsburg, 1999) might predict our results. For example, assuming conceptual binding occurs when its

constituent features are synchronously activated, the cross-modal feature verification advantage counterintuitively suggests that distant brain regions synchronize more quickly than proximal brain regions. Thus, although our results do not preclude a temporal binding mechanism, they do suggest that such a mechanism certainly is not the only manner in which representations may be bound, and that the organization of the connections over which these signals are synchronized plays an important factor. One possibility is that both convergence zones and temporal synchronization play different but complementary roles in conceptual binding. For example perceptual binding requires various objects to be parsed – that is, object features in a complex scene must be grouped in such a way that object representations do not interfere with one another. Although conceptual binding may avoid this problem to some extent through processing concepts sequentially, temporal binding provides a natural mechanism for a common set of representational units, integrated over a distributed set of convergence zones, to maintain multiple concepts in memory representational units without interference.

Retrieval of concept knowledge activates modality-specific representations—

A secondary, but no less important, claim is that concept knowledge is embodied in a representational system in which sensorimotor modality (i.e., the perceptual apparatus through which we learned the information) is a relevant property of our semantic representations. The present data are inconsistent with amodal representational systems (Caramazza, 1991; Tyler & Moss, 2001), which have no basis on which to predict that features activate one another more quickly by virtue knowledge type. Given no modality distinction in these models, semantic processing is largely driven by statistical properties such as feature correlation and distinctiveness. Because these factors are assumed also to play a role in multimodal models, the stimuli in our experiments were matched on them, making it difficult for our results to be explained by amodal models. The present research therefore adds to the growing body of literature that challenges the argument that our representational system stores only abstract amodal representations.

Note that this same argument applies to an alternative explanation of our results. Our experiments used feature and concept names, rather than, for example, pictures of parts and auditory clips of sounds. One might argue, therefore, that all processing occurred in a lexical system. This explanation of our results seem unlikely for two reasons. First, numerous imaging studies have shown that modality-specific areas are activated from feature and concept names (Goldberg et al., 2006a; see Martin, 2007, for a review). For example, during feature verification using printed word stimuli concerning object-associated visual, sound, touch, and taste features, regions involved in sensory processing in each modality were activated (Goldberg et al., 2006a). As another example, Hauk, Johnsrude, and Pullvermüller (2004) showed that reading words denoting tongue (*lick*), finger (*pick*), and leg (*kick*) actions activated regions in premotor cortex that were activated when participants made tongue, finger, and leg movements. Thus, it is likely that our feature and concept names activated modality-specific regions. Second, if one dismissed this evidence, then accounting for the present results would require appealing to lexical and/or statistical variables. However, we equated for the relevant variables, including numerous factors regarding feature-feature and feature-concept relations. Therefore, if feature names did not activate modality-specific representations, null effects should have been found in all four experiments, but they were not.

Perceptual and Conceptual Integration

We suggested in the Introduction that the question of how multimodal conceptual knowledge is integrated into unified concepts is the conceptual analog of the perceptual binding problem (Triesman, 1996); object perception requires a mechanism through which

input streams across multiple modalities are incorporated into a single perceptual object. A critical assumption underlying the present research and the body of research supporting distributed multimodal semantic representations is that the brain regions specialized for perceptual processing in each modality are also used in processing semantic information from the corresponding modality. From this perspective, the primary difference between perception and semantic processing is in the source of the inputs driving processing. In the case of perception, processing is initially driven by environmental inputs (though top-down processing introduces learned information not present in the environment into the processing stream). In contrast, conceptual processing may be primarily driven by internal representations acting as inputs (though environmental cues provide context that may guide the retrieval of information and provide additional information). If concepts and percepts are indeed two sides of the same coin, then the present research and the investigation of multimodal perception mutually inform one another.

Because we used words as stimuli, rather than pictures or sounds, our studies clearly involve multimodal integration in the conceptual system. Nonetheless, in a perceptual symbols system framework (Barsalou, 1999), retrieval of the underlying meaning is assumed to induce a pattern of activation in the primary sensory areas similar to that experienced during perception, albeit presumably less vividly. If this is so, then the same neural circuits could be used in both conception and perception, and therefore be subject to the same processes. The support for hierarchically deep representational models would thus suggest a similar organization for the routing of multimodal information in perceptual processing.

Conclusion

In the present research, we used complementary behavioural tasks to test assumptions regarding the neural architecture of semantic memory. Our studies provide clear evidence for the existence of a deep hierarchy in a multimodal distributed semantic memory system.

APPENDIX A

Within- and cross-modal pairs for Experiment 2

	Feature A	Feature B
Function-Form	worn for warmth	made of wool
	used in bands	made of brass
	used for storage	has doors
	used for killing	has a trigger
	worn by women	made of material
	used for passengers	has an engine
	worn on feet	made of leather
	used for transportation	has 4 wheels
	eaten in pies	has seeds
	worn by men	has pockets
	worn around neck	made of gold
	used for cutting	has a blade
	used by riding	has a seat
	hunted by people	has 4 legs
	eaten in salads	has leaves
	used for holding things	made of plastic

	Feature A	Feature B
Form-Form	eaten by people	has gills
	used for cooking	is electrical
	used for cleaning	has a handle
	used for eating	made of ceramic
	has fins	has gills
	has a lid	made of glass
	has sleeves	made of cotton
	made of brick	has a roof
	has a blade	has a wooden handle
	has pockets	has buttons
	has blades	is electrical
	has windows	has a roof
	has 4 wheels	has an engine
	has a tail	has hooves
	has buttons	made of material
	made of wood	has drawers
	has scales	has teeth
	has a seat	has wheels
	has a tail	has legs
	is flat	is rectangular
	has seeds	is round
	has doors	has windows
	has fur	has whiskers
	has horns	has hooves

APPENDIX B

Yoked within- and cross-modal pairs used in Study 3

Within-Modal Feature A	Cross-Modal Feature A	Target Feature B
has two wheels	used by riding	has handle bars
has a stinger	used for travel	has wings
has a clasp	used for gifts	made of gold
has buckles	worn for walking	made of leather
has hinges	used by opening	has a lock
has sections	used for juice	has pulp
made of diamonds	worn around neck	made of silver
has buttons	worn by men	has cuffs
has pockets	worn for covering	has zippers
made of fur	worn for rain	has a hood
has a metal blade	used for gardening	has a handle
has heels	worn on feet	has soles
made of rubber	used for games	is round

Within-Modal Feature A	Cross-Modal Feature A	Target Feature B
has a roof	used for living in	made of brick
has cushions	used for relaxing	has armrests
has pedals	used for cargo	has wheels
has doors	used for storage	has shelves

APPENDIX C

Yoked within- and cross-modal pairs and target concepts used in Experiment 3

Within-Modal Feature A	Cross-Modal Feature A	Feature B	Target Concept
has seeds	tastes sweet	is round	ORANGE
fires	used by the police	is loud	PISTOL
has a thumb	worn on hands	made of wool	GLOVES
has zippers	worn for covering legs	has buttons	TROUSERS
used for war	has a trigger	used for killing	GUN
talks	eats seeds	sings	PARAKEET
has feet	waddles	has a beak	DUCK
has laces	worn on feet	made of leather	BOOTS
has pockets	worn for warmth	has sleeves	COAT
eaten in sandwiches	has fins	is edible	TUNA
tastes tart	has peel	tastes sour	LEMON
eats rodents	has talons	flies	FALCON
has a blade	used for digging	has a wooden handle	HOE
used for watching television	has armrests	used for sleeping	COUCH
is soft	used for relaxing	is comfortable	SOFA
has a pit	is juicy	has skin	PLUM
used for passengers	has an engine	used for transportation	JET
has a collar	worn by women	made of cotton	BLOUSE
has a cushion	used by sitting on	has legs	STOOL
put on ceilings	has lightbulbs	used by hanging	LAMP
is smooth	used by throwing	is hard	STONE
worn on heads	has a strap	worn for protection	HELMET
is flat	used for carrying things	made of plastic	TRAY
used for construction	has a pointed end	used for holding	SCREWS
is damp	is dark	is cold	BASEMENT
tastes different flavours	eaten by baking	tastes good	PIE
has fur	used in experiments	has whiskers	MOUSE
has a plug	is hot	is electrical	TOASTER
has taps	used for washing	has a drain	BATHTUB
has fangs	slithers	has scales	PYTHON
used for music	is gold	used in bands	SAXOPHONE
has shelves	used for storing food	has doors	CUPBOARD
is pointed	used for writing	is thin	PEN

Within-Modal Feature A	Cross-Modal Feature A	Feature B	Target Concept
produces sound	has strings	produces music	GUITAR
is square	used for storage	is rectangular	BOX
eaten as pickles	has leaves	eaten in salads	BEETS

References

- Ahn W, Marsh JK, Luhmann CC, Lee K. Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition* 2002;30:107–118.
- Allport, DA. Distributed memory, modular systems and dysphasia. In: Newman, SK.; Epstein, R., editors. *Current Perspectives in Dysphasia*. Churchill Livingstone; Edinburgh, UK: 1985. p. 32-60.
- Ashcraft MH. Feature dominance and typicality effects in feature statement verification. *Journal of Verbal Learning and Verbal Behavior* 1978;17:155–164.
- Barsalou LW. Perceptual symbol systems. *Behavioral and Brain Sciences* 1999;22:577–660. [PubMed: 11301525]
- Barsalou, LW.; Pecher, D.; Zeelenberg, R.; Simmons, WK.; Hamann, SB. Multimodal simulation in conceptual processing. In: Ahn, W.; Goldstone, R.; Love, B.; Markman, A.; Wolff, P., editors. *Categorization inside and outside the lab: Essays in honour of Douglas L. Medin*. American Psychological Association; Washington, DC: 2005. p. 249-270.
- Bussey TJ, Saksida LM, Murray EA. Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *European Journal of Neuroscience* 2002;15:365–374. [PubMed: 11849302]
- Cohen JD, MacWhinney B, Flatt M, Provost J. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research, Methods, Instruments, & Computers* 1993;25:257–271.
- Cree GS, McNorgan C, McRae K. Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory & Cognition* 2006;32:643–658.
- Cree GS, McRae K. 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* 2003;132:163–201. [PubMed: 12825636]
- Cree GS, McRae K, McNorgan C. An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science* 1999;23:371–414.
- Damasio A. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 1989;33:25–62. [PubMed: 2691184]
- Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A. Neural systems behind word and concept retrieval. *Cognition* 2004;92:179–229. [PubMed: 15037130]
- Edelman S, Bülthoff H. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research* 1992;32:2385–2400. [PubMed: 1288015]
- Farah MJ, McClelland JL. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General* 1991;120:339–357. [PubMed: 1837294]
- Fuster JM, Bodner M, Kroger JK. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 2000;405:347–351. [PubMed: 10830963]
- Garrard P, Lambon Ralph MA, Hodges JR, Patterson K. Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology* 2001;18:125–174. [PubMed: 20945209]
- Goldberg RF, Perfetti CA, Schneider W. Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience* 2006a;26:4917–4921. [PubMed: 16672666]
- Goldberg RF, Perfetti CA, Schneider W. Distinct and common cortical activations for multimodal semantic categories. *Cognitive, Affective and Behavioral Neuroscience* 2006b;6:214–222.

- Goodale MA, Milner AD. Separate visual pathways for perception and action. *Trends in Neurosciences* 1992;15:20–25. [PubMed: 1374953]
- Green AE, Fugelsang JA, Kraemer DJM, Shamosh NA, Dunbar KN. Frontopolar cortex mediates abstract integration in analogy. *Brain Research* 2006;1096:125–137. [PubMed: 16750818]
- Hauk O, Johnsrude I, Pulvermuller F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 2004;41:301–307. [PubMed: 14741110]
- Humphreys GW, Forde EME. Hierarchies, similarity and interactivity in object recognition: “Category specific” neuropsychological deficits. *Behavioral and Brain Sciences* 2001;24:453–509. [PubMed: 11682799]
- Jacobs RA, Jordan MI. Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience* 1992;4:323–336.
- Kalénine S, Peyrin C, Pichat C, Segebarth C, Bonthoux F, Baciú M. The sensory-motor specificity of taxonomic and thematic conceptual relations: A behavioural and fMRI study. *Neuroimage* 2009;44:1152–1162. [PubMed: 18977304]
- Kellenbach ML, Brett M, Patterson K. Large, colorful, or noisy? Attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. *Cognitive, Affective & Behavioral Neuroscience* 2001;1:207–221.
- Landauer TK, Dumais ST. A solution to Plato’s problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 1997;104:211–240.
- Martin A. The representation of object concepts in the brain. *Annual Review of Psychology* 2007;58:25–45.
- Martin A, Chao LL. Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology* 2001;11:194–201. [PubMed: 11301239]
- Martin A, Haxby JV, Lalonde FM, Wiggs CL, Ungerleider LG. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 1995;270:102–105. [PubMed: 7569934]
- McNorgan C, Kotack RA, Meehan DC, McRae K. Feature-feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition* 2007;33:418–431.
- McRae K, Cree GS, Seidenberg MS, McNorgan C. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods* 2005;37:547–559. [PubMed: 16629288]
- McRae K, Cree GS, Westmacott R, de Sa VR. Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology* 1999;53:360–373. [PubMed: 10646207]
- McRae K, de Sa V, Seidenberg M. On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General* 1997;126:99–130. [PubMed: 9163932]
- Mummery CJ, Patterson K, Hodges JR, Price CJ. Functional neuroanatomy of the semantic system: Divisible by what? *Journal of Cognitive Neuroscience* 1998;10:766–777. [PubMed: 9831743]
- Myers, JL. *Fundamentals of experimental design*. Allyn & Bacon; Boston, MA: 1979.
- Paivio, AU. *Mind and its evolution: A dual coding theoretical approach*. Erlbaum; Mahwah, NJ: 2007.
- Parker, A.; Easton, A. Cross-modal memory in primates: The neural basis of learning about the multisensory properties of objects and events. In: Calvert, GA.; Spence, C.; Stein, BE., editors. *The Handbook of Multisensory Processes*. MIT Press; Cambridge, MA, US: 2004. p. 333–342.
- Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience* 2007;8:976–987.
- Pecher D, Zeelenberg R, Barsalou LW. Verifying different-modality properties for concepts produces switching costs. *Psychological Science* 2003;14:119–124. [PubMed: 12661672]
- Plaut DC. Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology* 2002;19:603–639. [PubMed: 20957556]
- Pollatsek A, Well AD. On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 1995;21:785–794.

- Riddoch MJ, Humphreys GW, Coltheart M, Funnell E. Semantic systems or system? Neuropsychological evidence re-examined. *Cognitive Neuropsychology* 1988;5:3–25.
- Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, Patterson K. The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review* 2004;111:205–235. [PubMed: 14756594]
- Sejnowski TJ, Kienker PK, Hinton GE. Learning symmetry groups with hidden units: Beyond the perceptron. *Physica* 1986;22D:260–275.
- Simmons KW, Barsalou L. The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology* 2003;20:451–486. [PubMed: 20957580]
- Simmons KW, Martin A, Barsalou LW. Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex* 2005;15:1602–1608. [PubMed: 15703257]
- Simmons WK, Ramjee V, Beauchamp M, McRae K, Martin A, Barsalou LW. A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 2007;45:2802–2810. [PubMed: 17575989]
- Sitnikova T, West WC, Kuperberg GR, Holcomb PJ. The neural organization of semantic memory: Electrophysiological activity suggests feature-based segregation. *Biological Psychology* 2006;71:326–340. [PubMed: 16129544]
- Spence C, Nicholls MER, Driver J. The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics* 2000;63:330–336. [PubMed: 11281107]
- Taylor PCJ, Walsh V, Eimer M. Combining TMS and EEG to study cognitive function and cortico–cortico interactions. *Behavioral Brain Research* 2008;191:141–147.
- Thompson-Schill SL. Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychologia* 2003;41:280–292. [PubMed: 12457754]
- Tranel D, Grabowski TJ, Lyon J, Damasio H. Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *Journal of Cognitive Neuroscience* 2005;17:1293–1305. [PubMed: 16197684]
- Triesman A. The binding problem. *Current Opinion in Neurobiology* 1996;6:171–178. [PubMed: 8725958]
- Tversky A. Features of similarity. *Psychological Review* 1977;84:327–352.
- Tyler L, Moss HE. Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences* 2001;5:244–252. [PubMed: 11390295]
- von der Malsburg, C. The correlation theory of brain function. In: Domany, E.; van Hemmen, JL.; Schulten, K., editors. *Models of neural networks II*. Springer; Berlin: 1994. p. 95–119. 1981 Reprinted in
- von der Malsburg C. The what and why of binding: The modeler’s perspective. *Neuron* 1999;24:95–104. [PubMed: 10677030]
- Warrington EK, McCarthy RA. Categories of knowledge: Further fractionations and an attempted integration. *Brain* 1987;110:1273–1296. [PubMed: 3676701]
- Zahn R, Moll J, Iyengar V, Huey ED, Tierney M, Krueger F, Grafman J. Social conceptual impairments in frontotemporal lobar degeneration with right anterior temporal hypometabolism. *Brain* 2009;132:604–616. [PubMed: 19153155]

Figure 1.

(a) A hierarchically shallow model with two directly interconnected modalities. (b) A hierarchically shallow model with two modalities connected via a single convergence zone. (c) A hierarchically deep model containing three modalities, each possessing a unimodal convergence zone that feeds forward to bimodal and trimodal convergence zones.

Figure 2.

Predictions for Experiments 1 and 2: Within- and cross-modal feature-feature activation in a shallow hierarchy requires the same number of processing steps.

Figure 3.
Predictions for Experiments 1 and 2: Feature-feature activation in a deep hierarchy requires fewer processing steps within-modally than cross-modally.

Figure 4.
Presentation sequence for stem and content components of stimuli in Experiment 2.

Figure 5. Predictions for Experiments 3 and 4: Feature-concept activation in a shallow hierarchy requires the same number of processing steps from within-modal pairs as from cross-modal pairs.

Figure 6. Predictions for Experiments 3 and 4: Feature-concept activation in a deep hierarchy requires additional processing steps from within-modal pairs than from cross-modal pairs.

Table 1

Equated variables in Experiment 1

Factor	Function-Form			Form-Form			p
	M	SE	M	SE	t(38)		
Mean % Shared Variance	18.5	3.0	18.7	4.3	-0.04		0.9
Number of Shared Concepts	5.5	0.7	4.9	0.8	0.63		0.5
Length in characters (1 st feature)	17.7	0.4	11.4	0.7	7.32		< .001
No. Concepts Unique to Feature (1 st feature)	10.3	3.4	11.5	3.8	-0.24		0.8
Concepts per feature (1 st feature)	15.8	3.7	16.3	4.0	-0.1		0.9
Length in characters (2 nd feature)	13.7	0.6	13.0	0.7	0.57		0.5
Number of Unique (2 nd feature)	10.9	3.0	11.5	3.2	-0.14		0.8
No. Concepts Unique to Feature (2 nd feature)	16.4	3.2	16.3	3.7	0.01		0.9

Table 2

Equated variables in Experiment 2

Factor	Function-Form		Form-Form		t(16)	p
	M	SE	M	SE		
Mean % Shared Variance	29.4	5.0	27.2	4.7	0.73	>.4
Number of Shared Concepts	2.6	0.2	2.2	0.1	2.38	.03
Content Word Length (1 st feature)	6.6	0.4	6.2	0.5	0.78	>.4
Number of Unique (1 st feature)	2.5	0.7	2.5	0.9	0.00	>.9
Concepts per Feature (1 st feature)	5.2	0.9	4.8	0.9	0.57	>.5
Distinctiveness (1 st feature)	0.25	0.03	0.32	0.04	-1.77	>.1

Table 3

Equated variables in Experiments 3 and 4

Factor	Within-Modal		Cross-Modal		ρ^a	p
	M	SE	M	SE		
High Density						
% Shared Variance	18.3	2.8	17.1	2.8	0.40	> .7
Production Frequency	11.9	1.3	10.0	1.1	1.07	> .3
Paired Concepts	4.4	0.7	4.8	0.9	-0.59	> .5
Concepts per Feature	6.9	1.6	8.1	0.2	-0.55	> .5
Intercorrelational Strength	243.0	37.7	286.0	35.0	-1.32	> .2
Length	11.6	1.3	12.6	0.9	-0.59	> .5
Low Density						
Mean % Shared Variance	22.1	6.2	20.8	6.0	0.43	> .6
Production Frequency	10.6	1.2	11.8	1.3	-0.87	> .3
Paired Concepts	2.8	0.2	2.9	0.2	-0.37	> .7
Concepts per Feature	6.4	1.3	5.8	1.0	-0.42	> .6
Intercorrelational Strength	230	37	209	37	0.79	> .4
Length	11.3	1.1	14.1	1.3	-1.49	> .15

^a paired *t*-tests, *df* = 17 for high density items, and 15 for low density items

Table 4

Feature verification latencies and error rates for Experiment 3

Factor	<i>Latency (ms)</i>		<i>Error Rate</i>	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
High Density				
Within-Modal	825	31	.05	.02
Cross-Modal	725	32	.05	.02
Low Density				
Within-Modal	883	39	.11	.02
Cross-Modal	826	30	.07	.02

Table 5

Feature verification latencies and error rates for Experiment 4

Factor	<i>Latency (ms)</i>		<i>Error Rate</i>	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
High Density				
Within-Modal	844	38	.10	.02
Cross-Modal	752	33	.06	.02
Low Density				
Within-Modal	870	41	.12	.02
Cross-Modal	833	36	.11	.02