What drives the organization of object knowledge in the brain?

Bradford Z. Mahon\textsuperscript{1,2} and Alfonso Caramazza\textsuperscript{3,4}

\textsuperscript{1}Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627, USA
\textsuperscript{2}Department of Neurosurgery, 601 Elmwood Ave, University of Rochester Medical Center, Rochester, NY 14642, USA
\textsuperscript{3}Department of Psychology, William James Hall, 33 Kirkland Street, Harvard University, Cambridge, MA 02138, USA
\textsuperscript{4}Center for Mind/Brain Sciences, University of Trento, Palazzo Fedrigotti, Corso Bettini 31, I-38068 Rovereto (TN), Italy

Various forms of category-specificity have been described at both the cognitive and neural levels, inviting the inference that different semantic domains are processed by distinct, dedicated mechanisms. In this paper, we argue for an extension of a domain-specific interpretation to these phenomena that is based on network-level analyses of functional coupling among brain regions. On this view, domain-specificity in one region of the brain emerges because of innate connectivity with a network of regions that also process information about that domain. Recent findings are reviewed that converge with this framework, and a new direction is outlined for understanding the neural principles that shape the organization of conceptual knowledge.

Category-specificity as a means to study constraints on brain organization

Brain-damaged patients with category-specific semantic impairments have conceptual level impairments that are specific to a category of items, such as animals, fruit/vegetables, nonliving things or conspecifics. Detailed analysis of those patients (Box 1) suggests that conceptual knowledge is organized according to domain-specific constraints [1,2]. According to the domain-specific hypothesis [2], there are innately dedicated neural circuits for the efficient processing of a limited number of evolutionarily motivated domains of knowledge. This interpretation of the neuropsychological phenomenon of category-specific semantic deficits has been extended to interpret results from functional magnetic resonance imaging (fMRI) in healthy subjects [3,4]. Much of the research using fMRI to study category-specificity has focused on the pattern of responses in the ventral visual pathway, which projects from early visual areas to lateral and ventral occipital-temporal regions, and processes object shape, texture and color in ways that are relatively invariant to viewpoint, size and orientation [5–7]. Different regions within the ventral pathway preferentially respond to images of faces, animals, tools, places, written words and body parts [4,6,8–13], see also [13–15].

The existence of consistent topographic biases by semantic category in the ventral stream raises fundamental questions about the principles that determine brain organization [4,10–12,16,17]. To date, the emphasis of research on the organization of the ventral stream has been on the stimulus properties that drive responses in a particular brain region, studied in relative isolation from other regions. This approach was inherited from well-established traditions in neurophysiology and psychophysics where it has been enormously productive for mapping psychophysical continua in primary sensory systems. It does not follow that the same approach will yield equally useful insights for understanding the principles of the neural organization of conceptual knowledge. The reason is that unlike the peripheral sensory systems, the pattern of neural responses in higher order areas is only partially driven by the physical input – it is also driven by how the stimulus is interpreted, and that interpretation does not occur in a single, isolated region. The ventral object processing stream is the central pathway for the extraction of object identity from visual information in the primate brain – but what the brain does with that information about object identity depends on how the ventral stream is connected to the rest of the brain.

Here, we focus on visual object recognition, as this has been the aspect of object knowledge and processing that has been studied in greatest depth; however, similar principles would be expected to apply to other modalities as appropriate. We argue that there are innately determined patterns of connectivity that mediate the integration of information from the ventral stream with information computed by other brain regions. Those channels are at the grain of a limited number of evolutionarily relevant domains of knowledge. We further suggest that what is given innately is the connectivity, and that specialization by semantic category in the ventral stream is driven by that connectivity. The implication of this proposal is that the organization of the ventral stream by category is relatively invariant to visually based, bottom-up, constraints. This approach corrects an imbalance in explanations of the causes of the consistent topography by semantic category in the ventral object-processing stream by giving greater prominence to endogenously determined constraints on brain organization.

The distributed domain-specific hypothesis

A domain-specific neural system is a network of brain regions [11] in which each region processes a different type
Box 1. Cognitive neuropsychological evidence for domain-specific constraints

Patients with category-specific semantic deficits can be differentially or even selectively impaired for knowledge of animals, plants, conspecifics or artifacts (for review see [11]). The knowledge impairment cannot be explained in terms of a differential impairment to a sensory or motor-based modality of information. Although discussion and debate continues as to whether non-categorical dimensions of organization can lead to category-specific brain organization, there is consensus that the phenomenon itself is ‘categorical’ (see Figure I for representative patients’ performance in picture naming and answering semantic probe questions).

There are important parallels between the neuropsychological literature on category-specific semantic deficits and the findings from functional neuroimaging and neurophysiology. First, the categories that emerge from the neuropsychological literature map onto the categories that emerge in functional imaging and neurophysiology. This indicates that the different methods and populations are tracking the same underlying property of brain organization. Second, the resistance of category-specific deficits to be explained by dimensions of organization that do not include semantic category [2] parallels the same pattern that has emerged in imaging and neurophysiology [60]. It is clearly the case that the brain is organized by sensory and motor modalities, and it is also the case that different sensory and motor modalities participate to varying extents in the representation of items from different categories. However, the existence of category-specificity in imaging [4], neurophysiology [67] and neuropsychology [11] cannot be explained exclusively by appeal to modality-based principles of organization. This suggests that the dimensions of brain organization that express themselves as phenomena of category-specificity (across methods and populations) are in fact domain-specific constraints on brain organization. Finally, there is emerging neuropsychological evidence for endogenous constraints on brain organization, including the existence of category-specific semantic deficits tested at age 16 years after stroke at 1 day of age [patient Adam, see below; ref 68].

There are also parallels between the patterns of category-specific semantic deficits and psychophysical studies of putatively specialized routes for processing specific classes of visual stimuli. For instance, New and colleagues [69], using a change detection paradigm, demonstrated a significant advantage for living animate stimuli. Thorpe and colleagues [70] have demonstrated extremely rapid and accurate detection of face and animal stimuli. Almeida and colleagues [65] have demonstrated that conceptual information about manipulable objects can be extracted from stimuli that are putatively not processed by the ventral visual pathway. These and other findings could indicate experimental ways of isolating domain-specific networks.

Figure I. Representative patients with category-specific semantic deficits. Patients with category-specific semantic deficits may have selective impairments for naming items from one category of items compared to other categories (top panel). Those patients may also have categorical impairments for answering questions about all types of object properties (i.e., visual/perceptual and functional/associative; bottom panel). For further discussion and references to the patients shown here, see [11].

of information about the same domain or category of objects [2,18]. The types of information processed by different parts of a network can be sensory, motor, affective or conceptual. The range of potential domains or classes of items that can have dedicated neural circuits is restricted to those with an evolutionarily relevant history that could have biased the system toward a coherent organization. A second important characteristic of domain-specific systems is that the computations that must be performed over items from the domain are sufficiently ‘eccentric’ [19] so as to merit a specialized process. In other words, the coupling across different brain regions that is necessary for successful processing of a given domain is different in kind from the types of coupling that are needed for other domains of knowledge.

For instance, the need to integrate motor-relevant information with visual information is present for tools and
The role of visual experience

According to the distributed domain-specific hypothesis, the organization by category in the ventral stream is not only a reflection of the visual structure of the world, it also reflects the structure of how ventral visual cortex is connected to other regions of the brain [11,23,33]. However, visual experience and dimensions of visual similarity are also crucial in shaping the organization of the ventral stream [34,35] – after all, the principal afferents to the ventral stream come from earlier stages in the visual hierarchy [36].

Although some authors have recently discussed nonvisual dimensions that could be relevant in shaping the organization of the ventral stream [4,6,7], many accounts differentially weight the contribution of visual experience in their explanation of the causes of category-specific organization within the ventral stream. Several hypotheses have been developed, and we merely touch on them here to illustrate a common assumption: that the organization of the ventral stream reflects the visual structure of the world, as interpreted by domain-general processing constraints. Thus, the general thrust of those accounts is that the visual structure of the world is correlated with semantic category distinctions in a way that is captured by how visual information is organized in the brain. One of the most explicit proposals is that there are weak eccentricity preferences in higher order visual areas that are inherited from earlier stages in the processing stream. Those eccentricity biases interact with our experience of foveating some classes of items (e.g. faces) and viewing others in the relative periphery (e.g. houses) [37]. Another class of proposals is based on the suppositions that items from the same category tend to look more similar than items from different categories, and similarity in visual shape is mapped onto the ventral occipital–temporal cortex [17].

It has also been proposed that a given category could require differential processing relative to other categories, for instance in terms of expertise [38], visual crowding [39] or the relevance of visual information for categorization [40]. Other accounts appeal to ‘feature’ similarity and distributed feature maps [41]. Finally, it has been suggested that multiple, visually based, dimensions of organization combine super-additively to generate the boundaries among category-prefering regions [12].

Common to all of these accounts is the assumption that visual experience provides the necessary structure, and that a visual dimension of organization happens to be highly correlated with semantic category.

Although visual information is important in shaping how the ventral stream is organized, recent findings indicate that visual experience is not necessary in order for the same, or similar, patterns of category-specificity to be present in the ventral stream. In an early position emission tomography study, Buchel and colleagues [42] showed that congenitally blind subjects show activation for words (presented in Braille) in the same region of the ventral stream as sighted individuals (presented visually). Pietrini and colleagues [43] used multi-voxel pattern analysis to show that the pattern of activation over voxels in the ventral stream was more consistent across different exemplars within a category than exemplars across categories. More recently, we [44] have shown that the same medial-to-lateral bias in category preferences on the ventral surface of the occipital–temporal cortex that is present in sighted individuals is present in congenitally blind subjects. Specifically, nonliving things, compared to animals elicit stronger activation in medial regions of the ventral stream (Figure 1).

Although these studies on category-specificity in blind individuals represent only a first-pass analysis of the role of visual experience in driving category-specificity in the ventral stream, they indicate that visual experience is not necessary in order for category-specificity to emerge in the ventral stream. This fact raises an important question – if visual experience is not needed for the same topographical biases in category-specificity to be present in the ventral stream, then, what drives such organization? One possibility, as we have suggested, is innate connectivity between regions of the ventral stream and other regions of the brain that process affective, motor and conceptual information.

Connectivity as an innate domain-specific constraint

A crucial component of the distributed domain-specific hypothesis is the notion of connectivity. The most obvious candidate to mediate such networks is white matter connectivity. However, it is important to underline that functional networks need not be restricted by the grain of white matter connectivity and, perhaps more importantly, task- and state-dependent changes could bias processing toward different components of a broader anatomical brain network. For instance, connectivity between lateral
and orbital prefrontal regions and the ventral occipital–temporal cortex [45,46] is crucial for categorization of visual input. It remains an open question whether multiple functional networks are subserved by this circuit, each determined by the type of visual stimulus being categorized. For instance, when categorizing manipulable objects, connectivity between parietofrontal somatomotor areas and prefrontal cortex could dominate, whereas when categorizing faces other regions could express stronger functional coupling to those same prefrontal regions. Such a suggestion would generate the expectation that whereas damaging prefrontal-to-ventral stream connections could result in difficulties categorizing all types of visual stimuli, disruption of the afferents to the prefrontal cortex from a specific category-prefering area could lead to categorization problems selective to that domain. The neural basis of the connectivity that supports domain-specific neural systems is, admittedly, in need of further development and articulation. Below, we will return to expectations that can be drawn from this explanation.

Evidence for innate constraints

The signature of innate structure is similarity across individuals, both within a species and potentially across
species. 'Innate' does not imply 'present-from-birth', although present-from-birth strongly suggests an innate contribution. Maturation in the context of the right types of experience could be necessary for the expression of innate structure, and interactions between innate and experiential factors can jointly constrain outcome [47]. This is particularly the case for mental processes, as there would be nothing to process without the content provided by experience. Several lines of evidence show that genetic variables capture similarity in functional brain organization as it relates to the presence of domain-specific neural circuits.

**Twin studies**

Two recent reports highlight greater neural or functional similarity between monozygotic twin pairs than between dizygotic twin pairs (for discussion see [48,49]). The strength of these studies is that experiential contributions are held constant across the two types of twin pairs. In a fMRI study, Polk and colleagues [50] studied the similarity between twin pairs in the distribution of responses to faces, houses, pseudowords and chairs in the ventral stream. The authors found that face and place-related responses within face and place selective regions, respectively, were significantly more similar for monozygotic than for dizygotic twins. In another study, Wilmer and colleagues [51] studied the face recognition and memory abilities [52] in monozygotic and dizygotic twin pairs. The authors found that the correlation in performance on the face recognition task for monozygotic twins was more than double that for dizygotic twins. This difference was not present for control tasks of verbal and visual memory, indicating selectivity in the genetic contribution to behavioral abilities (see also [53]).

**Congenital prosopagnosia**

Further evidence for a genetic contribution to face recognition abilities comes from congenital prosopagnosia, a developmental disorder in which individuals can have selective impairments for recognizing faces [54]. A recent study by Thomas and colleagues [55] found that congenital prosopagnosia was associated with reduced structural integrity of the inferior longitudinal fasciculus, which projects from the fusiform gyrus to anterior regions of the temporal lobe. Reduced structural integrity was also observed for the inferior fronto-occipital fasciculus which projects from the ventral occipital–temporal cortex to frontal regions. Such observations of reduced integrity of major white matter tracts linking the posterior occipital–temporal cortex with other brain regions underlines the strength of a network-level analysis in understanding the constraints that shape the organization of knowledge in the ventral stream.

**Non-human primates**

An expectation on the view that innate constraints shape category-specificity in the ventral stream is that such specificity, at least for some categories, can also be found in non-human primates. It is well known, using neurophysiological recordings, that preferences for natural object stimuli exist in the inferior temporal (IT) cortex of monkeys [35,56], comparable to observations with similar methods in awake human subjects [15]. More recently, functional imaging with macaques [57] and chimpanzees [58] suggests that at least for the category of faces, comparable clusters of face preferring voxels can be found in the temporal cortex in monkeys, as are observed in humans.

Such common patterns of neural organization for some classes of items in monkeys and humans could, of course, be entirely driven by dimensions of visual similarity, which are known to modulate responses in the IT cortex [59]. However, even when serious attempts have been made to explain such responses to dimensions of visual similarity, taxonomic structure emerges over and above the contribution of known visual dimensions. For instance, Kriegeskorte and colleagues [60] used multi-voxel pattern analysis to compare the similarity structure of a large array of different body, face, animal, plant and artifact stimuli in the monkey IT cortex and human occipital–temporal cortex.

The similarity among the stimuli was measured in terms of the similarity of the patterns of brain responses they elicited, separately on the basis of the neurophysiological data (monkeys) [56] and fMRI data (humans). The similarity structure that emerged revealed a tight taxonomic structure common to monkeys and humans, and which could not be reduced to known dimensions of visual similarity.

**Next steps**

Specialization of function in the brain is clearest at the level of primary sensory and motor areas that have a physical organization in the brain that projects topographically onto a psychophysical dimension such as retinotopy, tonotopy or somatotopy. At the other end of the continuum, there are aspects of human cognition that have eluded neat parcellation in the brain, such as the neural instantiation of the abstract and recursive systems that make human thought and metacognition possible. Somewhere in the middle are conceptual representations -- they interface with and draw on the sensory and motor systems and at the same time require the flexibility characteristic of symbolic representations [61]. We have outlined a framework for understanding the causes of category-specific organization in the brain that is based on the hypothesis that there are innate patterns of connectivity that constrain the distribution of category-specific neural regions. This proposal fully embraces a hierarchical view of the organization of conceptual knowledge [3]: the organization of the ventral stream reflects the final product of a complex tradeoff of pressures, some of which are expressed locally within the ventral stream and some of which are expressed as connectivity to the rest of the brain. Our suggestion is that connectivity to the rest of the brain is the first, or broadest, principle according to which the ventral stream comes to be organized by semantic category.

Although there is striking overlap in the semantic categories that can dissociate under conditions of brain damage and which show consistent topographic organization in the ventral stream (Box 1), there is some divergence between the lesion locations in patients with category-specific deficits and the patterns of neural activation observed with fMRI. In particular, focal lesions to category-prefering
regions within the ventral stream do not invariably lead to category-specific semantic deficits. This suggests that what is damaged in patients with category-specific semantic deficits are the broader neural circuits that are specialized for the impaired domain of knowledge. Damage to multiple regions within that domain-specific neural circuit could lead to a category-specific deficit by disrupting or disorganizing the broader network. Furthermore, damage to regions that serve to integrate processing across the whole domain, such as the anterior temporal lobes [62,63] for the domains of animals and conspecifics, could particularly disrupt functioning throughout the broader network.

A second direction for research that is encouraged by the distributed domain-specific hypothesis is to characterize the patterns of both anatomical and functional connectivity within domain-specific neural circuits. The expectation is that there will be a tight coupling between patterns of connectivity and the locations of category-prefering regions. In this regard, it is important to note that regions expressing connectivity with category-specific regions within the ventral stream are not necessarily ‘downstream’ from visual object recognition, and do not necessarily represent ‘more developed’ or ‘more processed’ information than what is computed in the ventral stream. Stimuli are processed through multiple routes in parallel, such as subcortical processing of emotional face stimuli [20,21] and dorsal stream processing of manipulable objects [64,65]. Thus, one exciting possibility is that fast but coarse analysis of the visual input that bypasses the geniculate striate pathway could ‘cue’ or ‘bias’ processing within the ventral stream according to the content of the stimulus to be processed [45], analogous to attentional modulation of early visual responses.

A third way in which the distributed domain-specific hypothesis can be tested is to explore the connectivity of all the categories that show selective responses in the ventral stream. For instance, an expectation that could be generated is that stimuli from different domains, such as hands and tools, can live next to each other in the ventral stream because both would be predicted to have connectivity to the somatomotor cortex. In other words, the way in which representations are organized in the ventral stream should follow patterns of connectivity, such that they are organized according to similarity metrics represented in other parts of the brain, rather than (only) by dimensions of visual similarity.

Perhaps the most pressing issue that must be addressed by the distributed domain-specific hypothesis is whether connectivity drives specialization by category, as we have proposed, or whether specialization of function is present independently of connectivity, and the connectivity emerges later. One way to empirically address this is to test individuals who are blind since birth. Sensory deprivation will remove the influence of local constraints, presumably expressed over short-range bottom-up connections from earlier visual regions, but would not be expected to fundamentally alter the ‘longer range’ connections. Combining detailed analysis of connectivity in such individuals with analysis of the location of category-prefering regions in the ventral stream could ground inferences about whether connectivity in fact drives the location of category preferences in the ventral stream. In particular, the regions specialized for printed words could offer a means to test this issue, as there is no motivation for presuming specialization of function to be innately present for printed words in the human brain. Because there are regions that are consistently specialized for printed words, the expectation would be that this specialization is driven by connectivity between the ventral stream and regions of the brain involved in linguistic processing. The prediction can be made that subject-by-subject variation in the location of the visual word form area (tested with Braille) in congenitally blind individuals will match up with subject-by-subject variation in connectivity between that region of the ventral stream and other language processing regions of the brain.

The core of our proposal, that specialization in a region of the brain is driven, in part, by constraints on how that information will ultimately be used in the service of behavior, is not new. It is well established that visual processing bifurcates into a dorsal stream for object-directed action and spatial processing and a ventral stream for the extraction of object identity [66]. The two visual system models place important restrictions on plasticity of function within the visual system. Analogously, the distributed domain-specific hypothesis places new limits on plasticity of function within the ventral object processing stream, and suggests that the key to describing those limits lies in the patterns of connectivity between the ventral stream and other category-specific brain regions.

References

7 Cant, J.S. et al. (2009) fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream, Exp. Brain Res. 192, 391–405
8 Allison, T. et al. (1994) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors, Cereb. Cortex 4, 544–554
49 Zhu, Q. et al. (2010) Heritability of the specific cognitive ability of face perception. *Curr. Biol.* 20, 137–142
53 Zhu, Q. et al. (2010) Heritability of the specific cognitive ability of face perception. *Curr. Biol.* 20, 1–6