CellPress

Review

Object Domain and Modality in the Ventral Visual Pathway

Yanchao Bi,[1,*] Xiaoying Wang,[1] and Alfonso Caramazza[2,3]

The nature of domain-specific organization in higher-order visual cortex (ventral occipital temporal cortex, VOTC) has been investigated both in the case of visual experience deprivation and of modality of stimulation in sighted individuals. Object domain interacts in an intriguing and revelatory way with visual experience and modality of stimulation: selectivity for artifacts and scene domains is largely immune to visual deprivation and is multi-modal, whereas selectivity for animate items in lateral posterior fusiform gyrus is present only with visual stimulation. This domain-by-modality interaction is not readily accommodated by existing theories of VOTC representation. We conjecture that these effects reflect a distinction between the visual features that characterize different object domains and their interaction with different types of downstream computational systems.

Ventral Visual Cortex: Visual or Multi-Modal?

A core assumption of cognitive science and cognitive neuroscience is that the brain processes information at various levels of representation, progressing from those closely tied to stimulus features to increasingly more general and abstract representations. One of the mysteries in this framework is the transition from modality specific representations – those explicable fully in the language of a given modality – to representations that capture other properties of the object – such as, for example, the possibility that a particular shape is appropriate for a certain type of grip. The conjecture we will articulate here is related to this difficult problem in cognitive science and cognitive neuroscience. In particular, we consider the representational distinctions, or the information encoded in such representations, that might give rise to the well-established domain-level organization in higher-order visual cortex (ventral occipital temporal cortex, VOTC), and the general principles that drive this organization.

The nature of the representations computed in this territory is one of the major topics of investigation in cognitive neuroscience. Various types of visual-level dimensions have been proposed and examined to account for the category-preferring distributions [1–7]. This visual-driven framework has recently been challenged by a wave of studies that reported similar domain preference effects in sighted and congenitally blind individuals, for example, for the animate–inanimate distinction, places, bodies, large objects, and tools [8–13]. A commonly shared contention in these articles, highlighted in a recent review article [14], is that 'These findings provide a consistent demonstration of the supra-modal functional organization of specific task-related cortical networks', marking a shift of sentiment about the VOTC, from being part of the visual cortex to being supra-modal and, at least partly, independent from visual experience. In that framework, 'supra-modal' was defined to be 'brain areas [that] are equally recruited and show overlapping patterns of connectivity, mainly directed toward multisensory brain areas, in both sighted and blind individuals and across different sensory modalities'.

However, this is a one-sided reading of the empirical findings. The literature on the effects of visual deprivation on selectivity for various object categories actually paints an intriguing pattern

Trends

A wave of recent studies has reported similar domain preference effects in ventral occipital temporal cortex (VOTC) in sighted and congenitally blind individuals, leading to the contention that object representation in this region is multi-modal.

However, the effects of visual deprivation on object category selectivity paint an intriguing pattern of heterogeneity: selectivity to spatial navigation stimuli and manipulable artifacts found to be robustly multi-modal, whereas selectivity to animate objects reliably present only for sighted individuals when processing visual stimuli.

We formulate a novel conjecture about the nature of representations in VOTC: representation types are partly driven by the nature of the mapping between object visual properties and other object properties, which differ between animate and inanimate objects.

[1]State Key Laboratory of Cognitive Neuroscience and Learning, and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China
[2]Department of Psychology, Harvard University, Cambridge, MA, USA
[3]Center for Mind/Brain Sciences, University of Trento, Rovereto TN, Italy

*Correspondence: ybi@bnu.edu.cn (Y. Bi).

 http://dx.doi.org/10.1016/j.tics.2016.02.002

of heterogeneity, with selectivity robustly observed for some categories even in the absence of visual experience, whereas rarely observed, if at all, for some other categories (e.g., [12]). Relatedly, studies that considered the effect of input modality on category selectivity in sighted individuals have similarly found cross-modal selectivity more robustly for some categories than others in VOTC (e.g., [15,16]). Although it has been argued [14,17–20] that interpretation of these results require caution because potential effects obtained with nonvisual input may be affected (contaminated) by visual imagery, it would have to be further explained how imagery might play different roles for different object domains. The different effects of visual deprivation and input modality on category selectivity remain unexplained by existing theories of object representation in VOTC.

The aim here is to bring attention to this intriguing empirical phenomenon regarding the relationship between object domain and modality in VOTC, and propose a novel conjecture to explain this heterogeneity. The conjecture rests on the observation that the contrasting modality effects for different object domains might reflect a distinction between the visual features that characterize those different domains and their interaction with downstream computational systems and, in particular, action systems.

Domain Specificity Effects across Different Modalities Are Different

When sighted individuals view pictures, various clusters in VOTC are more responsive to certain categories of objects, such as faces, bodies, tools, or places. The overall distribution of category preference follows a broad animate versus inanimate distinction, with a further differentiation within the inanimate domain between manipulable and non-manipulable objects. This results in a tripartite organization, from ventral medial regions (parahippocampal and medial fusiform) showing preference for inanimate items broadly related to navigation, including scenes, places, buildings, and large non-manipulable objects, to lateral regions showing a preference for animate items including faces and animals, to more dorsolateral regions showing preference for bodies and small, manipulable objects [21–24] (see also [8,25] for the broad animate–inanimate pattern distinction).

In the following, we summarize the empirical findings regarding these several major object domain effects in VOTC along this tripartite distinction, bringing together the comparison between sighted visual versus nonvisual input and blind nonvisual cases. Studies using fMRI and positron emission tomography (PET) in which category-selective activations in VOTC in sighted or blind participants were investigated through at least one of the following types of stimuli: object names (including generating mental images of object names), object sounds, haptically presented objects, and objects presented through **sensory substitution devices** (see Glossary). We considered only those cases where category selectivity was tested by contrasting the target category to some other type of object category (control category). Experiments where nonvisual modalities were examined but did not yield positive results were included in an attempt to reduce potential file-drawer problems.

The results are presented in Figure 1A and Table 1. They show that the degree of consistency across input modalities and experience groups (blind versus sighted) differs greatly across object domains. Results for each specific domain effect are described in the following sections, beginning with nonvisual experiments in the sighted, followed by experiments in the blind.

Items Related to Spatial Navigation

It is well established that the medial fusiform gyrus/parahippocampal gyrus is more strongly activated when a sighted person sees pictures of scenes, buildings, or large objects relative to other objects [12,24,26,27]. Such selectivity is highly robust across various visual and nonvisual modalities within sighted individuals and across various nonvisual modalities in blind individuals

Glossary

Connectional fingerprints: the unique set of anatomical or functional connections a cortical region owns, which could be measured as the vector of the cortical region's connection strengths with other cerebral regions.

Functional fingerprints: the unique response properties a cortical region exhibits, which could be measured as vector of the region's response strengths to a variety of stimuli or tasks (e.g., object categories).

Multi-modal: for the purpose of this review, multi-modal was used to mean qualitatively similar task-related activation patterns across multiple sensory modalities even if the overall activation strengths were different.

Resting-state functional connectivity: the synchrony between spontaneous temporal fluctuations of brain activity of distinct brain regions in the absence of external stimuli.

Sensory substitution devices: devices that transform stimuli presented in one sensory modality into stimuli of another sensory modality. For example, shape-to-sound sensory substitution devices transform visual images into sounds through isomorphic mapping from plane coordinates and brightness of visual pixels to timing, frequency, and loudness of sounds.