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.

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

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.

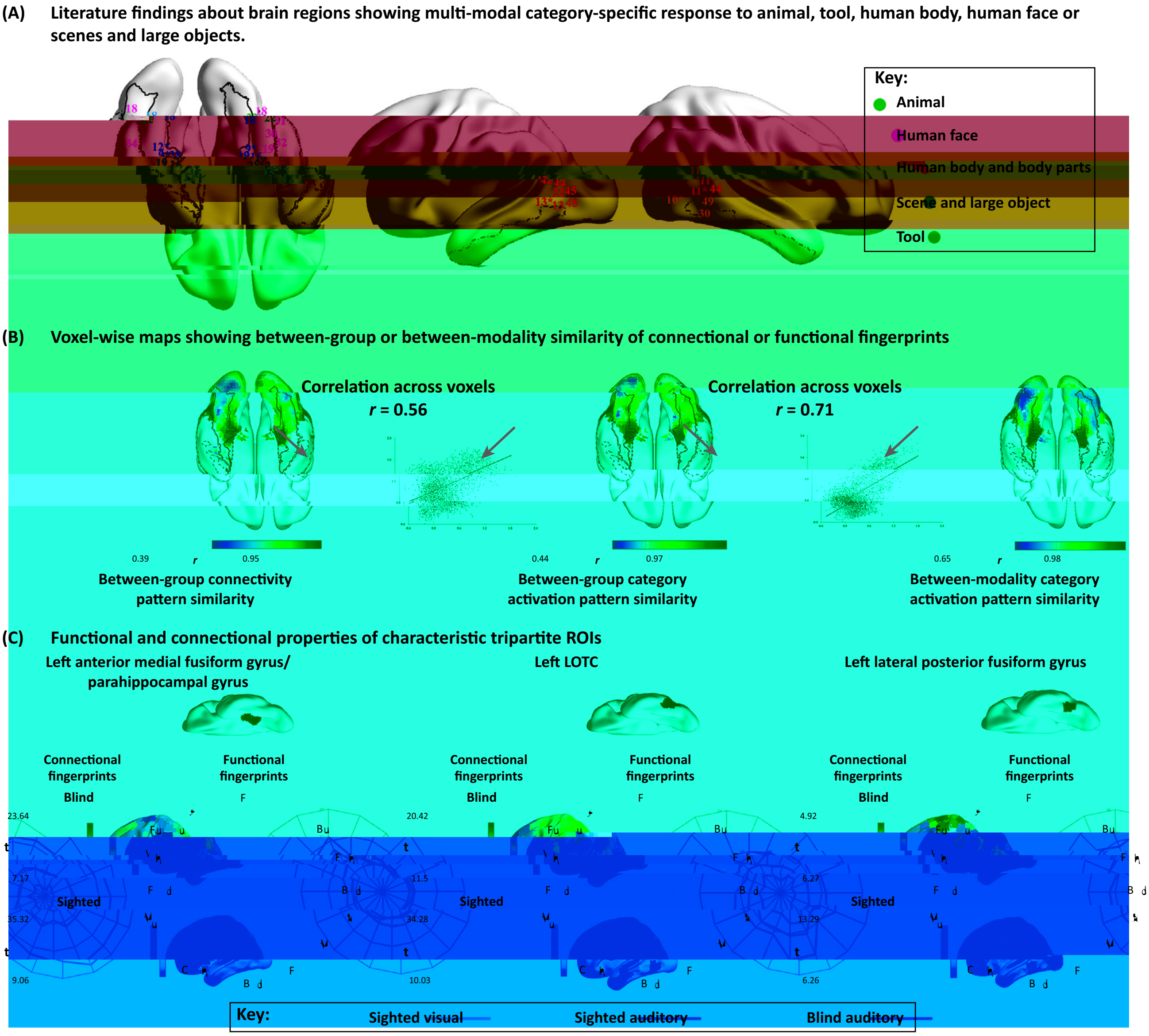
The nature of the representations computed in this territory is one of the major topics of investigation in cognitive neuroscience. Various types of visual-level dimensions have been proposed and examined to account for the category-preferring distributions [1–7]. This visual-driven framework has recently 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

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



Trends in Cognitive Sciences

Figure 1. Schematic Summary of Effects of Visual Experience on Category Selectivity in Ventral Occipital Temporal Cortex (VOTC). (A) Brain areas showing category selectivity in blind and sighted nonvisual experiments in VOTC. The studies are reported with the reference numbers in Table 1 and the reference list, with asterisks indicating findings with blind participants. Animal (green): [22]; human face (magenta): [18,19,30–32,34]; human body and body parts (orange): [10*,11*,30,49]; scene and large object (blue): [9*,12*,15,16,18,19,28]; Tool (red): [13*,22,42,44,45,48]. The position of each study is derived from the peak coordinates reported. Only positive findings can be shown here; studies where category selectivity was not observed in nonvisual modalities can be found in Table 1. (B) Comparison of the functional and connectional fingerprints of the VOTC voxels in the blind and sighted groups. The brain maps show voxel-wise similarity of resting-state functional connectivity patterns between blind and sighted groups, of category response patterns between blind and sighted nonvisual experiments, and of category response patterns between sighted visual and nonvisual modalities, respectively. Warmer colors indicate greater similarity, which are primarily found in the left anterior medial fusiform gyrus/parahippocampal gyrus and lateral occipital temporal cortex (LOTC). The lateral posterior fusiform gyrus showed colder colors, indicating less similarity. The scatter plots demonstrate the correspondence between maps across voxels. The contours mark the fusiform gyrus (solid), parahippocampal gyrus (dashed), and inferior temporal gyrus (dash-dot). (C) Properties of the characteristic tripartite VOTC region of interests (ROIs) in blind and sighted individuals. For each ROI, the connectional fingerprint maps show the resting-state functional connectivity strength between each region with the seed ROI (t values); the functional fingerprint map shows the response strengths to the 16 object categories (β values). Panels (B) and (C) are adapted and reprinted with permission from the Society for Neuroscience [51].

Table 1. Summary of the Category Selectivity in VOTC Results in Sighted Nonvisual and Blind Nonvisual Experiments, Organized by the Tripartite Distribution

Anatomical Region	Functional Selectivity	Sighted Nonvisual		Blind	
		Selectivity Observed	Selectivity Not Observed	Selectivity Observed	Selectivity Not Observed
Anterior medial fusiform/ parahippocampal	Scenes and large objects	Auditory word [12] Written word [15,28] Visual imagery [18,19][a] Haptic [9][a] Sound [16][a]		Haptic [9][a] Auditory word [12]	
Lateral occipital temporal cortex (LOTC)	Tools	Auditory word [13,48] Written word [22,48] Sound [42,44], [45][a]	Written word [38][a]	Auditory word [13]	
	Body	Haptic Writteni/F373804.06432..566900-.5669267.1035553.3101443.9551rg1Tf6.9738006.9738185.555855.1874T443.9551rg[j/F3rry			

(i.e., robust across task input and experience), including haptic exploration to Lego scenes (both blind and sighted), imagery generation of buildings, semantic judgment or size judgment of visual or auditory names of places or large non-manipulable objects (both blind and sighted), or listening to sounds associated with landmarks [9,12,15,16,18,19,28].

Animate Items

Lateral posterior fusiform is known to be more strongly activated by pictures of animate items such as faces and animals relative to other objects [22,24,29]. Various studies investigated face-preferring responses in nonvisual tasks in sighted individuals and in blind groups, such as generating visual images in response to auditory cues or descriptions, performing semantic judgment on printed names of famous people, and haptically exploring face masks [15,18,19,30–34]. With sighted individuals performing imagery, haptic, or word tasks, contradictory findings that yield no clear consensus were obtained. For the blind group, congenitally blind participants did not show face selectivity in fusiform gyrus in haptic tasks [31,33], whereas late blind participants did [31].

For non-human animals, using nonvisual object input, selectivity in lateral posterior fusiform was even more fragile. Nearly all papers that included tests of animal selectivity in sighted individuals that used written or spoken object names [8,12,22,35–41] or object sounds [16,42–45] failed to find selectivity for animals relative to other objects.

Items Related to Bodies and Manipulable Objects

The lateral occipital temporal cortex (LOTC) is known to show preference to images of body parts and small manipulable objects, which has been interpreted to reflect the bodily motor components associated with these stimuli [46,47]. The preference for tools in LOTC was robust in nearly all studies considered, including when sighted subjects responded to printed or auditory object names [13,22,48] and to object sounds [42,44,45]. Similar tool selectivity in LOTC was also observed in congenitally blind individuals [13].

For preference to body parts in LOTC, studies have been carried out testing verbal or haptic stimuli in sighted people and haptic stimuli or sensory substitution device-generated sound stimuli in congenitally blind participants. Haptic exploration of body parts in both sighted and blind individuals, as well as body shape conveyed through sensory substitution devices in blind subjects, consistently activated at least part of LOTC [10,11,30,49]. The results using words are less consistent [30,49,50].

This review shows that the results for sighted and blind participants in experiments that used nonvisual stimuli allow the following empirical generalization: highly convergent results are obtained for sighted and congenitally blind participants; not all object domain selectivity is equally robust across modalities in both subject populations. Although the effects for large objects and scenes (spatial navigation stimuli) and manipulable artifacts show a robust **multi-modal** nature, the effects for animate objects were only observed in sighted individuals and only when processing visual stimuli, with little evidence for domain effects in blind participants or in sighted participants in experiments with nonvisual stimuli. Thus, domain-selective VOTC is neither uniformly multi-modal (amodal) nor uniformly unimodal.

We recently provided direct support for the pattern of object domain by modality and experience effects in VOTC gleaned from the various studies reviewed here [51] (Figure 1B,C). We systematically compared category-related responses and **resting-state functional connectivity** patterns between congenitally blind and sighted individuals across the whole VOTC. We obtained voxel-wise, large-scale, continuous maps of the degree to which **connectional** and **functional 'fingerprints'** of ventral visual cortex depend on visual experience or input. There was close agreement between connectional and functional maps, pointing to a strong interdependence of connectivity and function in VOTC. We observed that although visual input and experience (or the absence thereof) had a pronounced effect on the response and connectivity profiles of early visual cortex, their effects on higher-order 'visual' cortex were not homogeneous. Specifically, the functionality and connectivity of lateral posterior fusiform gyrus was strongly affected, whereas those of the anterior medial (showing strongest

Box 1. How Well Do Existing Theories Work?

Is the phenomenon of domain-by-modality interaction explained by current theories about VOTC representation?

Object Visual Property Hypothesis

Sophisticated vision theories have been developed about how different neural patches in the ventral visual pathway come to prefer specific object categories. The theories emphasize the role of types of visual information that are robustly associated with object categories, such as eccentricity, size, rectilinear, or curvature features [1–3,6,7,61]. Accounts of this type do not explain why selectivity for inanimate domains is strongly independent of visual experience and input, whereas selectivity for animate objects seems to depend on visual experience and input. Specifically, it is unclear how the proposed visual properties (e.g., eccentricity) lead either to the multi-modal or modality-specific nature of these patches of VOTC. These hypotheses of the factors that determine domain specificity are not incompatible with the observed modality heterogeneity of VOTC, but neither do they provide an explanation for the observed phenomenon.

Connectivity–Constraint Hypothesis

It has been proposed that domain specificity in VOTC is driven in large part by structural and/or functional connections with other brain regions that process nonvisual properties of the corresponding categories [54,62–65]. However, this hypothesis, too, is silent on why domain-preferring regions in VOTC are differentially affected by visual experience and input type.

The interaction of domain by modality showing visual specificity for animate objects in VOTC is reminiscent of explanations of category-specific semantic deficits for living things, which proposed that such deficits reflect damage to a visual semantic system, which is especially important for living categories [4,37,66–69]. Whatever the merits of this hypothesis (see [70,71] for critical analysis), the observation that visual properties are relatively more salient for living versus non-living things does not explain why domain-preferring regions in VOTC differ in terms of whether they respond only to visual object stimuli or to multiple modalities and types of stimuli (e.g., tactile input, spoken words).

action value (e.g., the handle of a cup), making such representations accessible through different modalities and, hence, multi-modal.

This conjecture is based on the assumption that, particularly for inanimate objects, shape properties severely constrain their motor and function representations; that is, how we interact with them and what they are used for. For example, a thin longish shape (blade) could afford the motor act and function of cutting; a flattish solid shape could afford the motor act and function of pounding. Small artifacts can be manipulated, whereas large artifacts generally indicate a fixed location [52,53] and involve whole-body movements such as approaching or going around them. Because of this, 'visual' representations of inanimate objects are typically parsed to reflect those visual characteristics that are relevant for physical interaction. In other words, at this level of visual representation the information encoded is not arbitrary object parts but shape properties that map naturally onto action-based systems [54,55]. Such articulated correspondence across modalities makes the VOTC representation (and probably other related modality representations) directly addressable from different input systems and hence multi-modal, at least in this sense.

For animate items the picture is rather different. The relationship between visual shape properties (at least at the level of size or shape) and how we interact motorically with an animal is much less articulated. An elongated animal could either be a harmless, even beneficial friend (e.g., earthworm) or a dangerous enemy (venomous snake). Two spiders with similar shape may elicit different types of actions – one can be benign and 'useful' for humans, yet another can be deadly and requires opposite reactions. The visual shape similarity between a wolf and a dog is greater than that between a dog and a cat, yet very often the latter two elicit a more similar reaction (petting instead of fleeing) from human observers than the former two. Item size is not indicative of how to react either. Both small animals (leech or black widow spider) and large animals (e.g., tiger) may elicit a same 'flee' action, just as both small animals (rabbit) and large animals (deer) may elicit a same 'catch-for-food' act. That is, the relationship is more 'holistic' and arbitrary – for example, potential danger is not associated with any particular shape feature. Owing to the lack of articulated correspondence between visual shape properties and properties of other modalities, the visual representations are more 'isolated' from other inputs for animals. As a consequence, fusiform 'feature' representations do not interact directly with other sensory–motor systems, making them relatively inaccessible from other input modalities, and hence their more strictly visual representational format.

Intriguingly, for some categories selectivity is found in multiple subregions within this tripartite distribution. For instance, body parts not only induce selective responses in LOTC [56–58] but also in lateral posterior fusiform [57–60]. The different roles played by these two regions in body part recognition remain to be worked out. However, according to our conjecture, these multiple subregions represent different information types, which interact with different downstream processing systems. The representation of body parts in LOTC is appropriately formatted for interaction with the motor system, whereas that in lateral posterior fusiform is appropriate for further processing within the visual processing system itself. This view predicts that body selectivity in lateral posterior fusiform gyrus is only found for visual stimuli, whereas body selectivity in LOTC is relatively more immune to modality changes. The unimodal/multi-modal functional specialization in domain-preferring VOTC may be related to recent findings about the anatomical heterogeneity within this region, which show that the lateral and medial fusiform have distinct anatomical properties (Box 2).

Concluding Remarks

We began this review article by drawing attention to an empirical phenomenon regarding the effect of visual experience (and stimulus input) on object representations in higher-order visual cortex, showing that contrary to recent claims that this territory is multi-modal, there is a clear animate/inanimate dimension along which effects of modality differ. This empirical pattern motivates a novel conjecture about the nature of representations in VOTC: 43151.6535Tm

References

1. Levy, I. *et al.* (2001) Center–periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539
2. Arcaro, M.J. *et al.* (2009) Retinotopic organization of human ventral visual cortex. *J. Neurosci.* 29, 10638–10652
3. Nasr, S. *et al.* (2014) Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J. Neurosci.* 34, 6721–6735
4. Rogers, T.T. *et al.* (2005) Fusiform activation to animals is driven by the process, not the stimulus. *J. Cogn. Neurosci.* 17, 434–445
5. Gauthier, I. *et al.* (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573
6. Hasson, U. *et al.* (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34, 479–490
7. Konkle, T. and Oliva, A. (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74, 1114–1124
8. Mahon, B.Z. *et al.* (2009) Category-specific organization in the human brain does not require visual experience. *Neuron* 63, 397–405
9. Wolbers, T. *et al.* (2011) Modality-independent coding of spatial layout in the human brain. *Curr. Biol.* 21, 984–989
10. Striem-Amit, E. and Amedi, A. (2014) Visual cortex extrastriate body-selective area activation in congenitally blind people 'Seeing' by using sounds. *Curr. Biol.* 24, 687–692
11. Kitada, R. *et al.* (2014) The brain network underlying the recognition of hand gestures in the blind: the supramodal role of the extrastriate body area. *J. Neurosci.* 34, 10096–10108
12. He, C. *et al.* (2013) Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage* 79, 1–9
13. Peelen, M.V. *et al.* (2013) Tool selectivity in left occipitotemporal cortex develops without vision. *J. Cogn. Neurosci.* 25, 1225–1234
14. Ricciardi, E. *et al.* (2013) Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture?. *Neurosci. Biobehav. Rev.* 41, 64–77
15. Fairhall, S.L. *et al.* (2014) Person- and place-selective neural substrates for entity-specific semantic access. *Cereb. Cortex* 24, 1687–1696
16. Adam, R. and Noppeney, U. (2010) Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. *Neuroimage* 52, 1592–1602
17. Ricciardi, E. and Pietrini, P. (2011) New light from the dark: what blindness can teach us about brain function. *Curr. Opin. Neurol.* 24, 357–363
18. Ishai, A. *et al.* (2000) Distributed neural systems for the generation of visual images. *Neuron* 28, 979–990
19. O'Craven, K.M. and Kanwisher, N. (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023
20. Ishai, A. (2010) Seeing faces and objects with the 'mind's eye'. *Arch. Ital. Biol.* 148, 1–9
21. Konkle, T. and Caramazza, A. (2013) Tripartite organization of the ventral stream by animacy and object size. *J. Neurosci.* 33, 10235–10242
22. Chao, L.L. *et al.* (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* 2, 913–919
23. Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45
24. Kanwisher, N. (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11163–11170
25. Kriegeskorte, N. *et al.* (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141
26. Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* 392, 598–601
27. Epstein, R. (2005) The cortical basis of visual scene processing. *Vis. Cogn.* 12, 954–978
28. Fairhall, S.L. and Caramazza, A. (2013) Category-selective neural substrates for person- and place-related concepts. *Cortex* 49, 2748–2757
29. Grill-Spector, K. and Weiner, K.S. (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548
30. Kitada, R. *et al.* (2009) Functional specialization and convergence in the occipito-temporal cortex supporting haptic and visual identification of human faces and body parts: an fMRI study. *J. Cogn. Neurosci.* 21, 2027–2045
31. Goyal, M.S. *et al.* (2006) Tactile perception recruits functionally related visual areas in the late-blind. *Neuroreport* 17, 1381–1384
32. Kitada, R. *et al.* (2010) Brain networks involved in haptic and visual identification of facial expressions of emotion: an fMRI study. *Neuroimage* 49, 1677–1689
33. Kitada, R. *et al.* (2013) Early visual experience and the recognition of basic facial expressions: involvement of the middle temporal and inferior frontal gyri during haptic identification by the early blind. *Front. Hum. Neurosci.* 7, 7
34. Kilgour, A.R. *et al.* (2005) Haptic face identification activates ventral occipital and temporal areas: an fMRI study. *Brain Cogn.* 59, 246–257
35. Devlin, J.T. *et al.* (2005) Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia* 43, 69–74
36. Devlin, J.T. *et al.* (2002) Anatomic constraints on cognitive theories of category specificity. *Neuroimage* 15, 675–685
37. Price, C.J. *et al.* (2003) How is the fusiform gyrus related to category-specificity? *Cogn. Neuropsychol.* 20, 561–574
38. Devlin, J.T. *et al.* (2002) Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia* 40, 54–75
39. Rogers, T.T. *et al.* (2006) Anterior temporal cortex and semantic memory: reconciling findings from neuropsychology and functional imaging. *Cogn. Affect. Behav. Neurosci.* 6, 201–213
40. Lambert, S. *et al.* (2004) Blindness and brain plasticity: contribution of mental imagery? An fMRI study. *Cogn. Brain Res.* 20, 1–11
41. Lambert, S. *et al.* (2002) Neural substrates of animal mental imagery: calcarine sulcus and dorsal pathway involvement – an fMRI study. *Brain Res.* 924, 176–183
42. Doehrmann, O. *et al.* (2008) Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* 46, 2776–2786
43. Engel, L.R. *et al.* (2009) Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage* 47, 1778–1791
44. Lewis, J.W. *et al.* (2005) Distinct cortical pathways for processing tool versus animal sounds. *J. Neurosci.* 25, 5148–5158
45. Tranel, D. *et al.* (2005) Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *J. Cogn. Neurosci.* 17, 1293–1305
46. Bracci, S. and Peelen, M.V. (2013) Body and object effectors: the organization of object representations in high-level visual cortex reflects body–object interactions. *J. Neurosci.* 33, 18247–18258
47. Bracci, S. *et al.* (2012) Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *J. Neurophysiol.* 107, 1443–1456
48. Noppeney, U. *et al.* (2006) Two distinct neural mechanisms for category-selective responses. *Cereb. Cortex* 16, 437–445
49. Costantini, M. *et al.* (2011) Haptic perception and body representation in lateral and medial occipito-temporal cortices. *Neuropsychologia* 49, 821–829
50. Le Clec'H, G. *et al.* (2000) Distinct cortical areas for names of numbers and body parts independent of language and input modality. *Neuroimage* 12, 381–391
51. Wang, X. *et al.* (2015) How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *J. Neurosci.* 35, 12545–12559
52. Mullally, S.L. and Maguire, E.A. (2011) A new role for the parahippocampal cortex in representing space. *J. Neurosci.* 31, 7441–7449

53. Troiani, V.