Factors determining where category-selective areas emerge in visual cortex

Hans P. Op de Beeck*, Ineke Pillet, & J. Brendan Ritchie

Department of Brain & Cognition and Leuven Brain Institute, KU Leuven, Belgium

Corresponding author: hans.opdebeeck@kuleuven.be (H.P. Op de Beeck)

Abstract

A hallmark of functional localization in the human brain is the presence of areas in visual cortex specialized for representing particular categories such as faces and words. Why do these areas appear where they do during development? Recent findings highlight several general factors to consider when answering this question. Experience-driven category selectivity arises in regions that have: (i) preexisting selectivity for properties of the stimulus; (ii) are appropriately placed in the computational hierarchy of the visual system; and (iii) exhibit domain-specific patterns of connectivity to non-visual regions. In other words, cortical location of category selectivity is constrained by what category will be represented, how it will be represented, and why the representation will be used.

Keywords

Occipitotemporal cortex; Categorization; Cognitive neuroscience; Vision Sciences

Where does category selectivity come from?

Some of the most impressive examples of functional localization in the human brain can be found in the broad expanse of lateral and ventral **occipitotemporal cortex** (**OTC,** see Glossary), which is richly populated (Figure 1) with **areas** that exhibit **selectivity** for images of ecologically meaningful **categories** such as faces [1–4], visual word forms [5–7], bodies [8,9], hands [10], scenes [11–13], tools [14], and numerals [15]. Although the location of these areas is often highly predictable, why they emerge in their respective anatomical locations during development continues to be debated. Recent studies from different domains have provided important pieces for solving this puzzle [16–21], and point to three interacting factors that help constrain the cortical location of category-selective areas: (i) prior selectivity for **visual features** related to the stimulus properties; (ii) **computations** at different stages in the information-processing hierarchy of the visual system; and (iii) the early presence of **domain-specific connectivity** between brain regions. In what follows, first the character of these factors is elaborated on. The next sections then show how these factors can be deduced from the functional organization of OTC, and are implicated in recent studies on the emergence of category-selective areas. Finally, in the last section recent proposals are compared and contrasted that, to varying degrees, have emphasized the importance of the three factors.

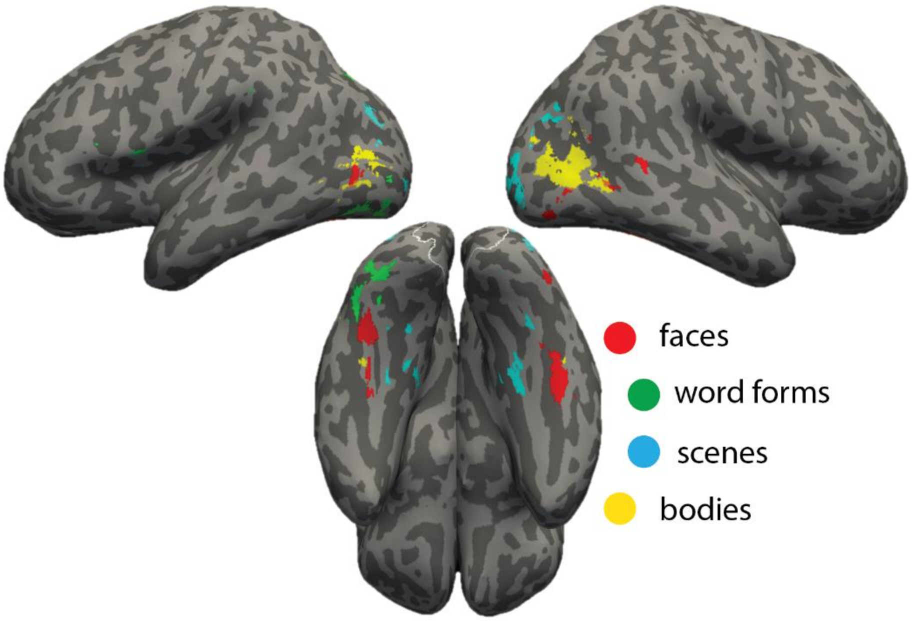

Figure 1: Category-selective areas on the cortical surface. *To illustrate the strength of this selectivity up to the individual level, data are shown of a single representative subject, from an unpublished fMRI experiment viewing stimuli of several categories, contrasting the fMRI BOLD response to the target category against all other categories. Results are projected onto the inflated cortical surface. The top left and right depict lateral views of left and right hemisphere and the bottom depicts a ventral view. The white lines at the occipital pole indicate the position of primary visual cortex (V1) for reference.*

Three factors to consider: What, how, and why

In broad terms, the three factors at the heart of the present review can be connected to what stimulus a category-selective area in OTC represents, how the stimulus is processed and represented, and why.

The first factor relates to the visual features of the *stimulus* itself**.** Typically, stimuli that belong to the same category tend to share many visually distinguishable features. Furthermore, people tend to position stimuli of the same category in their visual field in similar ways when they look at them. For example, all human faces tend to have a similar general shape with stereotyped positions of mouth and eye components (when viewed upright). Humans and primates also tend to fixate faces, which results in face stimuli being positioned in the foveal portion of the visual field when viewed [16,22].

The second factor relates to the *computations* required to extract the information that is relevant for categorization and individuation of exemplars from a category. Sensory processing requires a gradual extraction of relevant stimulus properties through a hierarchy of processing stages [23], as illustrated powerfully by the recent success of deep artificial neural networks [24]. In earlier stages computations enable extraction of simple features such as local line orientation with small receptive fields. In further stages extracting more complex combinations of these simpler features becomes possible, which finally results in a representation of category exemplars with a high degree of invariance for identity-preserving image transformations, such as rotation in the depth plane. In the example of object shape specifically, the processing might shift from local, part-based processing in earlier stages to more holistic representations in later stages with a high degree of correspondence to perceptual judgments of shape [25].

The third factor relates to the cognitive *domain* associated with the category. For example, faces are an important source of socially relevant signals indicating the identity, gender, emotions, moods, and mental states of conspecifics, which people track to successfully navigate social interactions. Similarly, word forms are the basis for the unique human capacity for non-verbal text-based communication.

The following sections review how these three factors help to describe the functional organization of OTC, and even explain how this functional organization emerged throughout development.

The functional organization of OTC

While often studied individually, it is a mistake to think of category-selective areas in isolation from each other. Rather OTC contains richly interconnected and overlapping topographies in which all category-selective areas are situated [26]. In addition to the local peaks of strong category selectivity for the subset of categories mentioned earlier, OTC is associated with weak and distributed selectivity for object categories more generally [27,28]. Furthermore, areas tend to cluster based on superordinate relationships between categories. For example, in ventral OTC areas responsive to “animate” stimuli, like faces, appear more laterally compared to those responsive to “inanimate” stimuli, like scenes [29]. Other superordinate principles have also been suggested, including for real-world size [30], color [31], and body partonomics [32]. Notably, even the areas with peak selectivity tend to represent this hierarchical structure by showing graded selectivity rather than only having a response for the preferred category [33]. Proposals about how category selectivity emerges have to recognize the complexity of this functional organization. Metaphorically speaking, a category-selective area is not a lone functional peak in an otherwise flat functional landscape—like a neural Kilimanjaro. Rather, it is one of many hilltops in a Yellowstone-like complex of which the origin can only be understood by investigating the larger-scale environment. Below three important characteristics of the OTC landscape are mentioned that connect to the three factors that serve as determinants for the emergence of category selectivity.

Cohabitation with visual feature maps

Besides category selectivity, OTC also contains **topographic maps** of other visual features, such as retinotopy, spatial frequency, and shape. For example, the visual cortex contains multiple retinotopic maps: neurons are organized accordingly with the visual field; in full, these neurons provide a map of the visual field or retina. The strength of this mapping is weaker in OTC compared to primary visual cortex, but at least some bias for retinal positions is still present [34,35]. Three properties are common to all these maps. First, the feature maps exist independently of category selectivity, for example, they have been demonstrated with stimuli that do not differ in category membership. Second, category selectivity cannot be reduced to selectivity for the features: even if the features are controlled, category selectivity is still observed, for a category is more than the sum of its featural parts [26,36]. Third, despite the fact that features and categories cannot be reduced to one another, there are interesting correlations between the two (Box 1).

Consider two examples. First, retinotopic maps exist in OTC independent of variations in category membership, and also, category selectivity cannot be reduced to effects of retinotopy. Face- and word-selective responses tend to be located in regions that prefer foveal stimuli, while scene-selective responses are instead biased towards more eccentric parts of retinotopic maps, and general object-selective responses fall in between these two extremes [34,37–40]. This correlation seems meaningful from a behavioral point of view, because face and word perception mostly involve foveal vision while scene perception does not [41].

Second, the responses to different shape stimuli in OTC reflect the perceived shape similarity of objects, and this correspondence is greater in areas where categorical divisions are stronger [42–44]. The shape responses are also correlated with the shape of the preferred category. For example, face-selective neurons and areas prefer roundish and convex shapes and PPA prefers straight lines [2,45–48]. Overall, the correlations with visual feature maps are a reminder that the stimulus and its visual properties matter for understanding the functional organization of OTC.

Computational hierarchy

The OTC landscape includes the later visual processing steps beyond early retinotopic areas. Within OTC, there is a further hierarchical organization. Categories can be associated with more than one area [49–52]. Initial proposals thought in terms of pairs of areas, with a hierarchy from lateral occipital to ventral occipitotemporal cortex, but the posited number of areas with a similar category preference has increased thanks to improvements in spatial resolution. For example, the fusiform face area FFA, initially conceptualized as one area, decomposes into multiple areas (FFA1 and FFA2) when scanning at a finer spatial resolution [50,51]. Data from human imaging and monkey electrophysiology support a hierarchical model with an increase of the complexity, invariance, and perceptual subjectivity of the represented features [49,53–61]. Areas higher up in the computational hierarchy, corresponding to more anterior regions in ventral OTC in human cortex, represent stimuli such as faces and words in a more holistic, less part-based manner and with higher levels of invariance for image transformations. More complex computational diagrams have also been proposed, with up to six distinct cortical and subcortical recurrent systems involving OTC regions [62].

Domain-specific connectivity

Structural and functional connectivity varies between category-selective areas in two ways [63–65]. First, connectivity within OTC tends to connect areas with the same category preference. For example, face-selective areas connect to other face-selective areas [66–68]. Second, and most importantly in the present context, connectivity between category-selective areas and regions outside the visual system also varies and tends to separate along cognitive domains. Face-selective areas are connected with right anterior temporal cortex, superior temporal sulcus, medial prefrontal cortex, and amygdala [69,70], which include regions involved in person memory, social cognition, and emotional processing. In contrast, the visual word form area (VWFA) is strongly connected with language regions in the left hemisphere [71,72]. The domain-specific connectivity is strong: category selectivity in a functional scan can be predicted as well from connectivity as it can from category selectivity in another functional scan [63,69] .

The emergence of category-selective areas

The combined importance of the three proposed factors is well-supported by recent studies on the development of category-selective areas (Figure 2), many of which focus on the domains of faces and words [1,5]. The contrast between these two areas is useful for a number of reasons [73,74]. First, faces and word forms have radically different visual features (just compare an image of a face to the word *face*); second, the specialization for these two domains is the result of two kinds of experience-driven learning—automatic exposure early in life for faces versus explicit education-based instruction for words; and third, they emerge at different times both in terms of individual development, and phylogenetically. Below several observations from this recent work are summarized. The next section will show that no single factor can account for all these observations.

Faces: From weak to strong selectivity through experience

Recent studies have investigated face selectivity in human infants in the first few months, and even days, of life. With fMRI differential responses were found in ventral OTC of 4 – 6 month old infants for faces versus scenes (Figure 2A) [21]. Even more strikingly, frequency tagged EEG responses in newborns (< 4 days old) proved to differ between upright and inverted face-like dot arrays (Figure 2B) [19]. While impressive, these results do not provide clear evidence for face selectivity shortly after birth. First, there was no clear preference for faces over objects in ventral OTC of infants (right panel in Figure 2A); Second, there was no difference in response for newborns between upright and scrambled arrays. Nevertheless, there is at least a less specific preference for faces compared to very different stimuli such as scenes and inverted arrays, which might be related to the presence of visual feature maps from the first months or days of life [75]. The possible existence of visual feature maps in infants is further supported by the finding of retinotopically specified functional connectivity in visual cortex, which is already present in infant monkeys [76].

Experience with faces seems essential for the emergence of actual face selectivity. In the most relevant recent study, two monkeys were raised without face exposure from birth (Figure 2C) [16]. When scanned at 8-9 months old they showed normal selectivity for other object categories such as hands, and retinotopic organization for eccentricity, but no face patches, even despite later exposure to faces. In normal monkeys, face selectivity was found in cortical regions with a preference for stimuli shown at low eccentricity, but no such selectivity was present in the face-deprived monkeys.

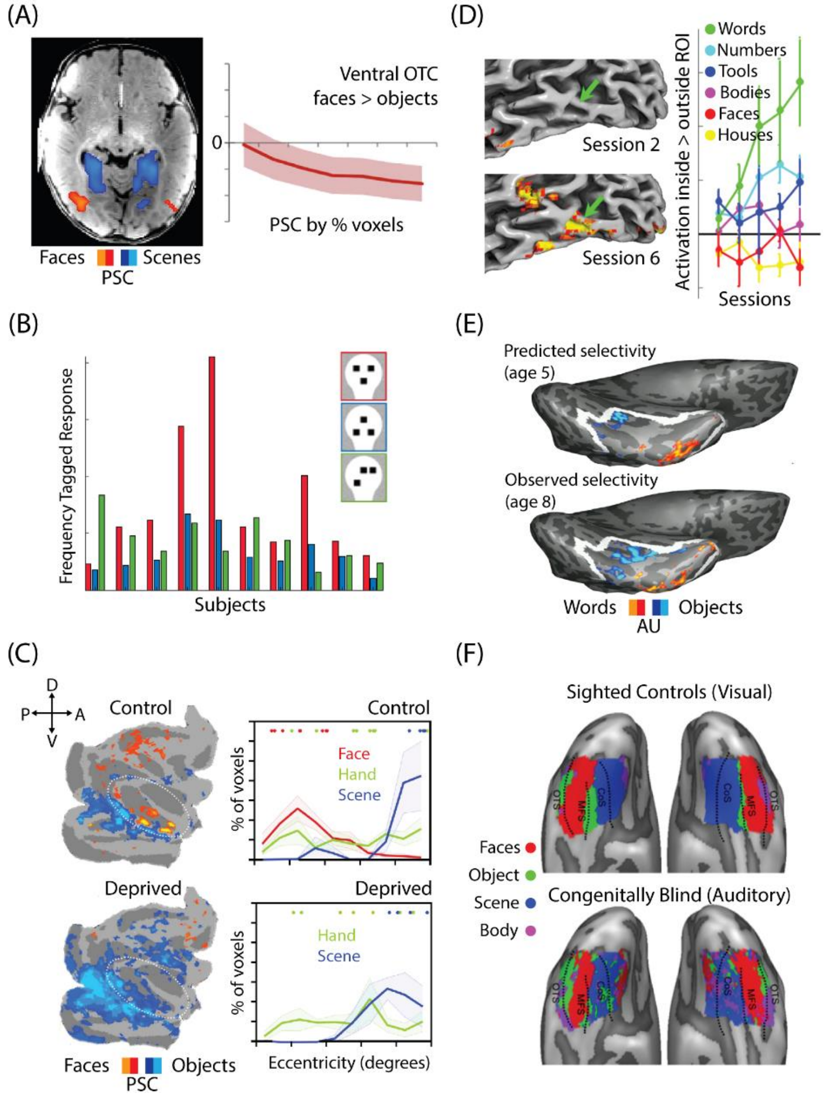

Figure 2: Recent findings on the development of category-selective areas. *(A) A Faces-scenes contrast shows typical face and scene selectivity (in fMRI percent signal change, PSC) in ventral OTC in 6 months old infants (left). However, voxels show no selectivity for the faces-objects contrast (right). Adapted from [21] . (B) Increase in EEG response to face templates at the repetition frequency as compared to inverted, but not scrambled face template stimuli, in newborns. Adapted from [19]. (C) Canonical face-selective patch areas fail to develop in faces-deprived monkeys relative to controls (left), while selectivity for other stimulus categories and its relationship to eccentricity bias is preserved (right). Adapted from [16]. (D) Top left: Lack of word form selectivity before schooling (Session 2). Bottom left: Clear selectivity at end of the school year (Session 6). Right: Activation to image categories in an ROI defined by responses from Sessions 6 and 7 (not depicted). Adapted from [18]. (E) Connectivity at age 5 predicts future location of word-form-selective voxels (in arbitrary units, AU) after children become literate (at age 8). Adapted from [17]. (F) Maps of preferred category in ventral OTC are strongly correlated between sighted and congenitally blind participants based on, respectively, visual and auditory category stimuli. Adapted from [20].*

Words: From weakly selective cortex to strong selectivity through experience and connectivity

Word form selectivity tends to emerge after other forms of selectivity, like for faces, have already developed [77,78]. In a longitudinal fMRI study selectivity for various categories was investigated in the part of left fusiform gyrus which was the future site of VWFA once children learned to read (Figure 2C) [18]. It was found that the VWFA appears in portions of cortex without obvious selectivity, at least not for the tested stimuli, except some weak selectivity for tools and the aforementioned foveal bias. The emergence of category selectivity can also be delayed substantially, or even not happen at all, depending on when an individual learns to read. For example, illiterate adults lack a VWFA [79].

The future sites of category-selective areas are strongly constrained by domain-specific connectivity. A longitudinal study was carried out on the VWFA both before and after literacy [17]. This research found that structural connectivity from left fusiform gyrus to language regions predicted the future location of the VWFA (Figure 2D). This result suggests that beyond visual feature selectivity, domain-specific connectivity may already anticipate where future areas will emerge. No similar data exist yet for the face domain, but similar constraints could be at play given that much of the connectivity with visual and nonvisual regions relevant for face perception (e.g., amygdala [80]) is already present in the infant brain [81,82].

Selectivity in the congenitally blind

The lack of category selectivity prior to visual experience with a particular category indicates that some sensory experience with a target category is necessary for areas to emerge. However, this seems at odds with the observation of selectivity in congenitally blind participants. Braille reading by the congenitally blind causes activity in approximately the same cortical location as VWFA in the sighted [83], and the emotional expression of voices induces activity in the typical location of FFA [84]. A study showed that the spatial distribution of face, body, object, and scene selectivity elicited in OTC of congenital blind participants, induced through category-related sounds (for faces e.g. laughing, whistling, chewing, blowing a kiss), is similar to the findings in normal control participants when viewing faces (Figure 2F) [20]. The blind participants have no retinotopic activation (indeed some of them even had no retina and thus no retinal waves), they cannot foveate faces, and yet the face selectivity tended to show a similar distribution as in normal control subjects. Exposure to the domain-specific function of a category without sensory experience is therefore sufficient to observe category selectivity with at least some of the properties of the selectivity seen in normal controls (also see Box 2).

Selectivity for novel categories

Overall, studies with novel categories support the importance of all three factors, stimulus characteristics, functional domain, and computational constraints. The effect of stimulus characteristics can be most easily investigated with novel objects: the objects differ strongly in visual features, but the functional domain and the computational requirements are controlled. After relatively short training procedures, studies have found changes in distributed activity patterns that depend upon the visual features of the objects [45,85]. After applying eight months of training in young monkeys, areas were found with focal selectivity induced by training, and the location of selectivity seemed to bear some similarity across monkeys when trained with the same stimuli [86]. Stimulus characteristics can also be dissociated from function by using familiar stimuli in novel ways. For example, one study found that, when subjects were trained to recognize individual faces as placeholders for word forms, there was greater selectivity for these faces in left, rather than right, fusiform gyrus [87]. The reverse approach has also been taken: using novel stimuli in familiar domains. Such studies have shown an increase in the selectivity for novel objects that become landmarks in scene-selective areas and for novel objects that become tools in tool-selective cortex [88,89]. Finally, experimental manipulations of computational requirements can also influence the cortical distribution of category responses [90].

Why the three factors are essential

The three factors emphasized—stimulus, computation, and domain—help to frame, and compare and contrast, different proposals for why category-selective areas arise where they do (Figure 3). More specifically, they point to the commonalities between proposals of how different category-selective areas emerge and help to reveal how proposals that tend to emphasize a particular factor, should take into account the other factors. Furthermore, when taken together, they also have consequences for how to think about the specialization of category-selective areas more generally.

Figure 3. Framework on where category-selective areas develop. *(Left) Before experience-driven learning with categories such as faces and words, there are already several factors in place in OTC and in the brain at large. As an example, a large region within OTC (oval) is positioned at the high-level (oval has green outline) in the computational processing hierarchy with preceding regions (green circles) that are at the low- or mid-level respectively; already consists of several maps of stimulus features (pink gradient within oval e.g. retinotopic map showing foveal (upper part) to peripheral bias (lower part)); and has several domain-specific connections to other regions in the brain (orange circle and arrow e.g. social cognition domain, blue circle and arrow e.g. language domain). (Middle) Experience-driven learning takes place: e.g. experience with faces and words. (Right) The already present factors in the large region within OTC (oval) constrain where such category-selective areas will emerge: FFA (fusiform face area, red circle) develops in the high-level part of the hierarchy within the visual system (given its processing requires holistic computations) where there is already selectivity for stimulus features related to faces such as a foveal bias, and where there are connections to the specific domain relevant for faces (social cognition). VWFA (visual word form area, dark blue circle) also develops in the high-level part of the hierarchy within the visual system (given its processing requires recognizing font-invariant letters and incorporating multi-letter statistics) where there is already selectivity for stimulus features related to words such as a foveal bias, and where there are connections to the specific domain relevant for words (language).*

Unifying factors in localization

There are several accounts of the development of category-selective areas in specific domains, such as faces and word forms, which can be characterized in terms of the three-factor framework. It has been proposed that the location of VWFA is the result of a learning process that places it in portions of OTC that are relatively functionally uncommitted, but in a manner constrained by multiple factors [18,91]. Because of the visual properties of word forms, they may target portions of feature maps in OTC with high-resolution representation of foveal shape, high spatial frequencies, and line junctions, but location is further constrained by having the appropriate level in the computational hierarchy and by preexisting connectivity to downstream language regions. Similarly, the framework is broadly consistent with previous proposals about face selectivity [92]. The location of FFA may be impacted by the same topographic maps (for retinotopy, spatial frequency, and curvature) based on the features of face stimuli, a need for holistic computations, and further by connectivity to subcortical (e.g. amygdala) and social brain regions (e.g. medial prefrontal cortex), which may help to explain why infants preferentially look at faces from very early in development. Nevertheless, the amount of evidence for each of these factors is different across domains, with e.g. stronger direct evidence for a role of connectivity in the case of word forms. It is unclear whether this is due to a difference between domains in terms of the role of a specific factor such as domain-specific connectivity, or simply due to a temporary lack of studies on the matter.

The insufficiency of the stimulus and visual features

Some have argued that the major factor that explains the location of selectivity is the stimulus, based on the presence of feature maps that reflect an intrinsic connection between retinotopy, receptive field size, and curvature tuning in the visual system [75,76]. However, focusing solely on the stimulus cannot be the basis for a more general framework for explaining the location of different category-selective areas, as it cannot account for several findings that speak to the important role of the domain factor. First, a stimulus-only approach appears at odds with results regarding the constraints determining the location of the VWFA, which seems to show only modest prior feature selectivity and can be predicted by prior domain-specific connectivity [17,18]. Second, a focus on the visual properties of the stimulus alone cannot explain why Braille reading activates similar locations to VWFA, or category-related sounds reveal a similar map for categories in ventral OTC in the congenitally blind [20,83].

Strong evidence in favor of a crucial role of the stimulus comes from studies suggesting that experience with words [79] and faces [16] is a necessary factor for the emergence of category selectivity. However, it is important to understand that having experience with a category is much more than mere exposure. Illiteracy is not simply a failure of exposure to word forms, but rather to learn how to link them to (typically) sound patterns of spoken language. Thus, even if connections from the fusiform gyrus to language centers exist, in illiterate individuals they go unused. In other words, selectivity cannot exist in the absence of domain-specific use. The study showing that experience with faces is necessary for the development of face selectivity did not just deprive the monkeys of the typical exposure to faces, but also of the use of faces as a social stimulus [16]. Thus, for both faces and word forms it might be the combined experience with the stimulus and its domain-specific function that drives the emergence of selectivity. Exposure by itself is not enough.

The insufficiency of computational constraints

Since the first discovery of a category-selective area, the FFA, an important source of disagreement has been the extent to which category-selective areas are genuinely specialized to represent information about their area-defining category, or instead carry out category-general operations [93,94]. In the most notable example, the so-called "expertise" account has suggested that areas of OTC, and in particular FFA, are not specialized to represent particular categories like faces, but rather are recruited as people learn to perform fine-grained discrimination of individual exemplars and represent images of category members more holistically [95–97]. When similar discriminative and holistic capacities are developed for other categories then the same area will be recruited.

The holistic processing expertise hypothesis can be interpreted as a specific concretization of the aforementioned importance of considering the computational hierarchy of the visual system. In the later stages of this system, representations become more holistic, not only for face representations but as a general property of the system [49]. The holistic processing expertise hypothesis can be rephrased as a proposal for a shift towards more holistic representations in some domains of expertise, thus towards later stages of this hierarchical processing system. In our view, the other two factors, domain-specific function and related connectivity and stimulus-specific features, then further determine how exactly the category selectivity will be organized in these later stages of the visual system.

The computation factor makes specific predictions. For example, the more two domains share computational constraints, the more they compete for the same part of the visual system. Words appear to compete for the foveal cortical territory occupied by faces when people learn to read [79], and in experts in other domains the objects of expertise also compete with faces [98]. However, due to the working of the other two factors, there will still be further variability in the location where selectivity emerges. For example, even when the expertise is associated with an increase in holistic processing, the resulting selectivity does not overlap with face selectivity [90], or at least shows only a partially overlapping spatial distribution of selectivity [99,100], reflecting the fact that the computational factor only partially explains the nature and location of category selectivity.

The insufficiency of domain-specific connectivity

Proposals have been made that emphasize the role of domain-specific and potentially **innate** connectivity for understanding the functional organization of category selectivity [101]. This is consistent with the observation that prior domain-specific connectivity with nonvisual regions predicts later category selectivity and with finding similar selectivity in the congenitally blind—two results that cannot easily be explained by the other factors. However, the strength and content of this selectivity in the blind relative to selectivity in normal controls is not yet clear (Box 2), and may be weaker [20]. Furthermore, even though category and object selectivity in the blind has now been demonstrated for a range of categories [102–106], the number of participants in studies in this domain is often relatively low which might cause an overestimation of effect size—especially when compared to the numerous participants and studies in which visual category selectivity has been demonstrated at the single-subject level (as happens in each study that uses a functional localizer to define category-selective areas in each individual participant). The strength of selectivity in the blind might also interact with the domain, being smaller for animals and faces [107].

In addition, there are several findings which seem to invoke the other factors. For example, it remains to be seen how an exclusive connectivity account would explain the partial overlap between pre-existing retinotopic maps (and possibly other feature maps) and emerging category selectivity for categories such as faces [16] (see Box 3). Also, an exclusive focus upon domain-specific connectivity is at odds with a distinct spatial distribution of category selectivity after visual exposure to novel object categories with different visual properties but under similar task conditions – as has been observed in young monkeys and to a lesser extent in human adults [85,86].

Concluding remarks

This discussion capitalized upon the large amount of results available for different category domains, and focused on determining the factors that contribute to the emergence of category-selective areas. Instead of considering how selectivity emerges for individual domains, e.g. words [91] or faces[92], this article explicitly searched for convergence between domains. Unlike previous proposals that focus strongly upon single factors, including stimulus properties [75], computational constraints [108] and domain-specific function and related connectivity [101], reviewing recent findings points to a framework in which all three factors are crucial for a full understanding of how functional selectivity emerges in the human brain.

Note that there is limited direct experimental evidence for some of the factors, even for often-cited determinants such as eccentricity biases (see Box 3). As far as studies exist, such as for the importance of experience with faces for the development of face selectivity [16], the conclusions are sometimes based upon one experiment with few subjects and the experimental design does not experimentally dissociate the role of stimulus (exposure) and function. Even though there is sufficient evidence to retain each of the three factors, there is a clear lack of knowledge about further details, such as the relative weight of the factors and their potential interactions. The Outstanding Questions provide several other examples of specific questions that future studies should investigate in order to better understand the specific contribution of each of the three factors and their combination to the emergence of category-selective areas.

Box 1: Feature-based coding for visual categories

The presence of both feature maps and category selectivity in OTC has helped to motivate the revisionary hypothesis that OTC does not in fact exhibit true category selectivity, but rather is responsive to visual features that happen to be confounded when comparing exemplars from different categories [109]. Empirical support for this hypothesis, that apparent category selectivity can be explained by low-level image properties, comes from studies showing that at least in some cases the neural selectivity for faces, scenes, and other categories, can be accounted for by orientation, size, and spatial frequency properties of the images used [110–112]. While this low-level feature coding hypothesis presents an important challenge to the orthodoxy regarding the representational function of OTC, it is poorly supported on both theoretical and empirical grounds (for more extensive discussion, [113]).

First, the feature maps in OTC are likely partially constitutive of the manner in which visual categories are represented in areas of OTC; that is, areas likely reflect a feature-based categorical code because visual features cannot be fully divorced from categories. Therefore, the mere fact that responses in (by hypothesis) category-selective areas partially depend on visual features of the stimuli does not provide evidence that these areas are only representing correlated visual features. Second, the studies in support of the feature-confound hypothesis have typically used stimuli in which low-level features such as shape and texture are not simply correlated, but indeed confounded, with category information [110,112]. When studies have explicitly aimed to make category information orthogonal to target visual features, such as shape, both category and shape information is still present in portions of OTC [42,114,115]. Such a result is not predicted if category-selective OTC is in fact coding for low-level visual features. Instead, the evidence is consistent with the hypothesis that OTC codes for object category using a representational format that is sensitive to the features of objects, a feature-based categorical code [113].

Box 2: What is represented in the visual cortex of the blind?

Box 1 proposed that categories are coded in a feature-based manner. These features seem mostly visual in nature, and the complete information processing hierarchy of the visual system is implicated in the reconstruction of these features from the incoming retinal input. Given reports of category selectivity in the blind, should this notion of feature-based categorical coding be abandoned?

A first remark to make is that, in sighted people, category-selective areas of OTC can be activated by other senses, such as touch and audition [20,116]. The structure of these responses can also be remarkably similar to those induced by visual simulation [117]. While it could be argued that such findings can be explained by visual imagery, a more parsimonious explanation would be to argue in favor of a multisensory feature-based categorical code, given the robust category selectivity in the OTC of congenitally blind participants.

The results of blind studies could also be used to go a step further and argue in favor of a more abstract code for categories in OTC. In particular, it has been proposed that so-called “sensory-independent” task specialization is the main organizational principle explaining category selectivity in OTC, and in cortex as a whole [118,119]. The task does indeed matter, also in the three-factor framework of this review, as it is the main reason why domain-specific connectivity comes into play. However, in sighted participants, the role of visual feature maps and of computational constraints points towards a significant contribution of sensory-specific processes.

From this perspective, it is natural to propose that congenitally blind participants also represent object categories through a feature-based categorical code. In this situation, the features cannot be visual and are related to other senses. To provide a specific example, to the extent that an area like VWFA subserves connecting symbols with sounds and/or meaning, the prediction can be made that in the blind it represents those sensory features that are helpful for the same functions in sighted individuals. Nevertheless, given that these nonvisual features will not be associated with topographic maps in visual cortex, and cannot be computed in the visual system, it can be expected that domain-specific connectivity will dominate the emergence of this selectivity to a larger extent than in sighted participants.

Box 3: The case study of eccentricity bias

As the present review makes clear, multiple factors influence where category selectivity areas emerge, and further studies are needed to pinpoint the exact contribution of the individual factors, their respective weights, and how they interact. This can be illustrated with one specific stimulus feature that has been studies extensively: the retinal position at which the stimuli from a particular category are typically seen, and more specifically its eccentricity. The importance of eccentricity to predicting the location of category-selective areas was originally proposed because it could elegantly explain why areas selective for faces and words, two stimuli that are typically foveated, fall in parts of OTC that have a preference for foveal stimulation, whereas areas selective for scenes, which are often seen in periphery, occupy parts of OTC that prefer eccentric input [41].

This is an interesting proposal, but there are several problems associated with it. For example, it cannot explain that the selectivity for such eccentricity-associated categories is found in similar areas of OTC in the congenitally blind. This motivates the search to find other factors that help explain the location of category-selective areas. So why, then, would there seem to be a correlation with eccentricity maps in OTC in the sighted? One obvious candidate is domain-specific connectivity. Possibly the visual system is already set up in such a way that domain-specific connectivity is correlated with eccentricity preferences. For example, monkeys already foveate faces before category selectivity emerges in the visual system. If evolution has built in this bias to foveate stimuli from certain domains, then why would it not also have a bias to connect foveally-preferring regions with brain networks for these domains? If this bias is present, then the observation that face-selective areas emerge in regions of OTC that already have a foveal bias [76] might present equivocal evidence for a direct role of the eccentricity bias in determining where selectivity would land. Instead, the primary factor might still be domain-specific connectivity, which, as explained in the main text, is also activated when subjects obtain experience with faces in a natural context.

At the heart of this discussion is the lack of experiments that cleanly dissociate and manipulate the different factors. No study has ever experimentally manipulated the eccentricity at which a stimulus is shown and demonstrated that this manipulation completely changes the position at which selectivity emerges with respect to the known eccentricity map. It would be particularly relevant to test this for domains such as faces and words, where domain-specific connectivity has an important contribution. Recently, a study suggested a role of eccentricity bias after observing Pokémon selectivity in Pokémon experts centered in regions with a foveal bias that fits with the foveal presentation of these stimuli during game play [120]. However, another interpretation is that the Pokémon selectivity might reflect the recruitment of already existing distributed selectivity for animals and cartoons, with which Pokémon stimuli share visual features and which happen to be in foveally preferring portions of OTC. Furthermore, the study did not experimentally manipulate the eccentricity of Pokémon stimuli during training.

In sum, even for an eccentricity bias the experimental evidence is far from complete, notwithstanding its status as one of the most frequently cited determinants of how and where category selectivity emerges in the visual system.

Outstanding questions

Does evidence support a role for all three factors in the localization of other category-selective areas, such as for bodies or tools?

What is the relative importance of the three factors in explaining the development of different category-selective areas? Can they be thought of as having influence chronologically, or perhaps iteratively throughout the course of development?

How strong and uniform is the role of connectivity as a constraint for emerging selectivity? Can it be proven more directly in the case of face selectivity? Can the effect of the visual stimulus and the domain-specific function that together characterize domain-specific visual experience be dissociated?

Can independent evidence be found for a role of visual feature maps, retinotopy and other maps alike, which cannot be explained by pre-existing connectivity?

What would happen if faces or words would be experienced mostly in eccentric positions, yet still be used with the same goal (e.g., for identity and emotion recognition in case of faces)?

How does the strength of selectivity and the represented features differ between sighted and congenitally blind participants? Is domain-specific connectivity more important for the emergence of selectivity in blind participants?

What is the upper bound on the number of areas and categories that can be reliably differentiated in OTC? Can this be pushed further by advances in spatial resolution? Does the current three-factor framework explain the development of the detailed properties of this functional organization?

Acknowledgments

This work was supported by the European Research Council (ERC-2011-StG-284101), a Hercules grant (ZW11_10), a KU Leuven research council grant (C14/16/031), and an Excellence Of Science (EOS) grant (G0E8718N) to HO; and by funding from the FWO and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501, via a FWO [PEGASUS]^2 Marie Skłodowska-Curie fellowship (12T9217N) to J.B.R.

Glossary

Area: a portion of neocortex that exhibits preferential selectivity different from surrounding cortex. Typically operationalized in fMRI using a contrast in selectivity between experimental conditions. In comparison to this functional definition, neuroanatomical studies use two additional defining criteria, specific anatomical connectivity and a histological fingerprint.

Computation: the processing of information in line with some sort of procedure or rule (e.g., local or global processing; part-based or holistic; taking the maximum of inputs or averaging). Sensory and cognitive processing typically involves a hierarchy of computations.

Category: a natural class of entities in the environment, which can be taxonomic (e.g. biological, artefactual or scene kinds) or functional (e.g. manually manipulable, food source, possible threat), and might be defined at multiple hierarchical levels (e.g., animals, mammals, cats).

Connectivity: the extent to which two regions of cortex communicate between each other.

Domain-specific: the specialization of psychological capacities for addressing a class of problems relating to how we understand or interact with the world (e.g. social interaction, non-verbal communication).

Fusiform face area (FFA): an area (or collection of areas at higher spatial resolution) in the fusiform gyrus exhibiting selectivity for faces over other categories such as scenes and objects.

Innate: a psychological capacity or functional unit in the brain that is not the result of learning processes in psychological or neural development.

Occipitotemporal cortex (OTC): a broad region of cortex centered at the intersection between the temporal and occipital lobes, and including portions of lateral occipital and ventral temporal cortex.

Selectivity: the extent to which a neural signal differs between experimental conditions.

Topographic map: a distribution of differential selectivity for visual features in a portion of neocortex such that a gradual change in feature value results in a gradual change in the anatomical position of the most responsive neurons/voxels.

Visual feature: A visual characteristic that can be of varying complexity, from simple (e.g., orientation, spatial frequency, retinotopic shape or size, chromatic properties) to complex (e.g., perceived shape).

Visual word form area (VWFA): an area (or collection of areas at higher spatial resolution) exhibiting selectivity for word forms over other categories such as faces and objects.

References

[1] N. Kanwisher, J. McDermott, and M. M. Chun, “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” *J. Neurosci.*, vol. 17, no. 11, pp. 4302–4311, 1997.

[2] D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell, and M. S. Livingstone, “A cortical region consisting entirely of face-selective cells,” *Science (80-.).*, vol. 311, no. 5761, pp. 670–674, 2006.

[3] N. Kanwisher and G. Yovel, “The fusiform face area: a cortical region specialized for the perception of faces,” *Philos Trans R Soc L. B Biol Sci*, vol. 361, no. 1476, pp. 2109–2128, 2006.

[4] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, “Face-specific processing in the human fusiform gyrus,” *J Cogn Neurosci*, vol. 9, pp. 605–610, 1997.

[5] L. Cohen *et al.*, “The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients,” *Brain*, vol. 123, no. 2, pp. 291–307, 2000.

[6] C. I. Baker, J. Liu, L. L. Wald, K. K. Kwong, T. Benner, and N. Kanwisher, “Visual word processing and experiential origins of functional selectivity in human extrastriate cortex,” *Proc Natl Acad Sci U S A*, vol. 104, no. 21, pp. 9087–9092, 2007.

[7] J. Bai, J. Shi, Y. Jiang, S. He, and X. Weng, “Chinese and Korean characters engage the same visual word form area in proficient early Chinese-Korean bilinguals,” *PLoS One*, vol. 6, no. 7, p. e22765, 2011.

[8] P. E. Downing, Y. Jiang, M. Shuman, and N. Kanwisher, “A cortical area selective for visual processing of the human body,” *Science (80-.).*, vol. 293, no. 5539, pp. 2470–2473, 2001.

[9] M. V Peelen and P. E. Downing, “Selectivity for the human body in the fusiform gyrus,” *J. Neurophysiol.*, vol. 93, no. 1, pp. 603–608, 2005.

[10] S. Bracci, M. Ietswaart, M. V Peelen, and C. Cavina-Pratesi, “Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex,” *J. Neurophysiol.*, vol. 103, no. 6, pp. 3389–3397, 2010.

[11] R. Epstein and N. Kanwisher, “A cortical representation of the local visual environment,” *Nature*, vol. 392, no. 6676, p. 598, 1998.

[12] S. Kornblith, X. Cheng, S. Ohayon, and D. Y. Tsao, “Article A Network for Scene Processing in the Macaque Temporal Lobe,” *Neuron*, vol. 79, no. 4, pp. 766–781, 2013.

[13] G. K. Aguirre, E. Zarahn, and M. D’Esposito, “An area within human ventral cortex sensitive to ‘building’ stimuli: evidence and implications,” *Neuron*, vol. 21, no. 2, pp. 373–383, 1998.

[14] A. Martin, C. L. Wiggs, L. G. Ungerleider, and J. V Haxby, “Neural correlates of category-specific knowledge,” *Nature*, vol. 379, no. 6566, p. 649, 1996.

[15] J. Shum *et al.*, “A brain area for visual numerals,” *J. Neurosci.*, vol. 33, no. 16, pp. 6709–6715, 2013.

[16] M. J. Arcaro, P. F. Schade, J. L. Vincent, C. R. Ponce, and M. S. Livingstone, “Seeing faces is necessary for face-domain formation,” *Nat Neurosci*, vol. 20, no. 10, pp. 1404–1412, 2017.

[17] Z. M. Saygin *et al.*, “Connectivity precedes function in the development of the visual word form area,” *Nat Neurosci*, vol. 19, no. 9, pp. 1250–1255, 2016.

[18] G. Dehaene-Lambertz, K. Monzalvo, and S. Dehaene, “The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition,” *PLoS Biol*, vol. 16, no. 3, p. e2004103, 2018.

[19] M. Buiatti *et al.*, “Cortical route for facelike pattern processing in human newborns,” *Proc. Natl. Acad. Sci.*, vol. 116, no. 10, pp. 4625–4630, 2019.

[20] J. van den Hurk, M. Van Baelen, and H. P. Op de Beeck, “Development of visual category selectivity in ventral visual cortex does not require visual experience,” *Proc Natl Acad Sci U S A*, vol. 114, no. 22, pp. E4501–E4510, 2017.

[21] B. Deen *et al.*, “Organization of high-level visual cortex in human infants,” *Nat Commun*, vol. 8, p. 13995, 2017.

[22] E. Di Giorgio, C. Turati, G. Altoè, and F. Simion, “Face detection in complex visual displays: An eye-tracking study with 3-and 6-month-old infants and adults,” *J. Exp. Child Psychol.*, vol. 113, no. 1, pp. 66–77, 2012.

[23] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, “How does the brain solve visual object recognition?,” *Neuron*, vol. 73, no. 3, pp. 415–434, 2012.

[24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[25] J. Kubilius, S. Bracci, and H. P. Op de Beeck, “Deep Neural Networks as a Computational Model for Human Shape Sensitivity,” *PLoS Comput Biol*, vol. 12, no. 4, p. e1004896, 2016.

[26] H. P. Op de Beeck, J. Haushofer, and N. G. Kanwisher, “Interpreting fMRI data: maps, modules and dimensions,” *Nat. Rev. Neurosci.*, vol. 9, no. 2, p. 123, 2008.

[27] D. D. Cox and R. L. Savoy, “Functional magnetic resonance imaging (fMRI) ‘brain reading’: detecting and classifying distributed patterns of fMRI activity in human visual cortex,” *Neuroimage*, vol. 19, no. 2, pp. 261–270, 2003.

[28] J. V Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, “Distributed and overlapping representations of faces and objects in ventral temporal cortex,” *Science (80-.).*, vol. 293, no. 5539, pp. 2425–2430, 2001.

[29] K. Grill-Spector and K. S. Weiner, “The functional architecture of the ventral temporal cortex and its role in categorization,” *Nat. Rev. Neurosci.*, vol. 15, no. 8, pp. 536–548, 2014.

[30] T. Konkle and A. Oliva, “A real-world size organization of object responses in occipitotemporal cortex,” *Neuron*, vol. 74, no. 6, pp. 1114–1124, 2012.

[31] R. Lafer-Sousa, B. R. Conway, and N. G. Kanwisher, “Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans , as in Macaques,” *J. Neurosci.*, vol. 36, no. 5, pp. 1682–1697, 2016.

[32] T. Orlov, T. R. Makin, and E. Zohary, “Topographic representation of the human body in the occipitotemporal cortex,” *Neuron*, vol. 68, no. 3, pp. 586–600, 2010.

[33] P. E. Downing, A. W. Chan, M. V Peelen, C. M. Dodds, and N. Kanwisher, “Domain specificity in visual cortex,” *Cereb Cortex*, vol. 16, no. 10, pp. 1453–1461, 2006.

[34] R. F. Schwarzlose, J. D. Swisher, S. Dang, and N. Kanwisher, “The distribution of category and location information across object-selective regions in human visual cortex,” *Proc Natl Acad Sci U S A*, vol. 105, no. 11, pp. 4447–4452, 2008.

[35] E. H. Silson, A. W.-Y. Chan, R. C. Reynolds, D. J. Kravitz, and C. I. Baker, “A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex,” *J. Neurosci.*, vol. 35, no. 34, pp. 11921–11935, 2015.

[36] J. Kubilius, A. Baeck, J. Wagemans, and H. P. Op de Beeck, “Brain-decoding fMRI reveals how wholes relate to the sum of parts,” *Cortex*, vol. 72, pp. 5–14, 2015.

[37] I. Levy, U. Hasson, G. Avidan, T. Hendler, and R. Malach, “Center-periphery organization of human object areas,” *Nat Neurosci*, vol. 4, no. 5, pp. 533–539, 2001.

[38] J. Larsson and D. J. Heeger, “Two retinotopic visual areas in human lateral occipital cortex,” *J Neurosci*, vol. 26, no. 51, pp. 13128–13142, 2006.

[39] C. C. Hemond, N. G. Kanwisher, and H. P. Op de Beeck, “A preference for contralateral stimuli in human object- and face-selective cortex,” *PLoS One*, vol. 2, no. 6, p. e574, 2007.

[40] U. Hasson, I. Levy, M. Behrmann, T. Hendler, and R. Malach, “Eccentricity bias as an organizing principle for human high-order object areas,” *Neuron*, vol. 34, no. 3, pp. 479–490, 2002.

[41] R. Malach, I. Levy, and U. Hasson, “The topography of high-order human object areas,” *Trends Cogn Sci*, vol. 6, no. 4, pp. 176–184, 2002.

[42] S. Bracci and H. Op de Beeck, “Dissociations and associations between shape and category representations in the two visual pathways,” *J. Neurosci.*, vol. 36, no. 2, pp. 432–444, 2016.

[43] H. Hong, D. L. K. Yamins, N. J. Majaj, and J. J. Dicarlo, “Explicit information for category-orthogonal object properties increases along the ventral stream,” *Nat. Neurosci.*, vol. 19, no. 4, pp. 613–622, 2016.

[44] H. P. Op de Beeck, K. Torfs, and J. Wagemans, “Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway,” *J Neurosci*, vol. 28, pp. 10111–10123, 2008.

[45] H. P. Op de Beeck, C. I. Baker, J. J. DiCarlo, and N. G. Kanwisher, “Discrimination training alters object representations in human extrastriate cortex,” *J. Neurosci.*, vol. 26, no. 50, pp. 13025–13036, 2006.

[46] H. P. Op De Beeck, J. A. Deutsch, W. Vanduffel, N. G. Kanwisher, and J. J. DiCarlo, “A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex,” *Cereb. Cortex*, vol. 18, no. 7, pp. 1676–1694, 2008.

[47] R. Rajimehr, K. J. Devaney, N. Y. Bilenko, J. C. Young, and R. B. Tootell, “The ‘parahippocampal place area’ responds preferentially to high spatial frequencies in humans and monkeys,” *PLoS Biol*, vol. 9, no. 4, p. e1000608, 2011.

[48] X. Yue, I. S. Pourladian, R. B. H. Tootell, and L. G. Ungerleider, “Curvature-processing network in macaque visual cortex,” *Proc. Natl. Acad. Sci.*, vol. 111, no. 33, pp. E3467–E3475, 2014.

[49] J. C. Taylor and P. E. Downing, “Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects,” *J. Cogn. Neurosci.*, vol. 23, no. 12, pp. 4122–4137, 2011.

[50] K. S. Weiner and K. Grill-Spector, “The improbable simplicity of the fusiform face area,” *Trends Cogn. Sci.*, vol. 16, no. 5, pp. 251–254, 2012.

[51] K. S. Weiner and K. Grill-Spector, “Sparsely-distributed organization of face and limb activations in human ventral temporal cortex,” *Neuroimage*, vol. 52, no. 4, pp. 1559–1573, 2010.

[52] K. S. Weiner and K. Grill-Spector, “Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex,” *Neuroimage*, vol. 56, no. 4, pp. 2183–2199, 2011.

[53] D. Y. Tsao, S. Moeller, and W. A. Freiwald, “Comparing face patch systems in macaques and humans,” *Proc Natl Acad Sci U S A*, vol. 105, no. 49, pp. 19514–19519, 2008.

[54] N. Caspari *et al.*, “Fine-grained stimulus representations in body selective areas of human occipito-temporal cortex,” *Neuroimage*, vol. 102 Pt 2, pp. 484–497, 2014.

[55] I. D. Popivanov, J. Jastorff, W. Vanduffel, and R. Vogels, “Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI,” *Neuroimage*, vol. 63, no. 2, pp. 723–741, 2012.

[56] D. Y. Tsao, “The Macaque Face Patch System: A Window into Object Representation,” *Cold Spring Harb Symp Quant Biol*, vol. 79, pp. 109–114, 2014.

[57] J. Haushofer, M. S. Livingstone, and N. Kanwisher, “Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity,” *PLoS Biol*, vol. 6, no. 7, p. e187, 2008.

[58] T. C. Kietzmann, S. Poltoratski, P. König, R. Blake, F. Tong, and S. Ling, “The occipital face area is causally involved in facial viewpoint perception,” *J. Neurosci.*, vol. 35, no. 50, pp. 16398–16403, 2015.

[59] F. Vinckier, S. Dehaene, A. Jobert, J. P. Dubus, M. Sigman, and L. Cohen, “Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system,” *Neuron*, vol. 55, no. 1, pp. 143–156, 2007.

[60] S. Dehaene, L. Cohen, M. Sigman, and F. Vinckier, “The neural code for written words: a proposal,” *Trends Cogn Sci*, vol. 9, no. 7, pp. 335–341, 2005.

[61] K. Grill-Spector, K. S. Weiner, K. Kay, and J. Gomez, “The Functional Neuroanatomy of Human Face Perception,” *Annu Rev Vis Sci*, vol. 3, pp. 167–196, 2017.

[62] D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, and M. Mishkin, “The ventral visual pathway: an expanded neural framework for the processing of object quality,” *Trends Cogn. Sci.*, vol. 17, no. 1, pp. 26–49, 2013.

[63] D. E. Osher, R. R. Saxe, K. Koldewyn, J. D. Gabrieli, N. Kanwisher, and Z. M. Saygin, “Structural Connectivity Fingerprints Predict Cortical Selectivity for Multiple Visual Categories across Cortex,” *Cereb Cortex*, vol. 26, no. 4, pp. 1668–1683, 2016.

[64] J. Gomez *et al.*, “Functionally defined white matter reveals segregated pathways in human ventral temporal cortex associated with category-specific processing,” *Neuron*, vol. 85, no. 1, pp. 216–

227, 2015.

[65] E. Premereur, J. Taubert, P. Janssen, R. Vogels, and W. Vanduffel, “Effective connectivity reveals largely independent parallel networks of face and body patches,” *Curr. Biol.*, vol. 26, no. 24, pp. 3269–3279, 2016.

[66] S. Moeller, W. A. Freiwald, and D. Y. Tsao, “Patches with links: a unified system for processing faces in the macaque temporal lobe,” *Science (80-.).*, vol. 320, no. 5881, pp. 1355–1359, 2008.

[67] M. Gschwind, G. Pourtois, S. Schwartz, D. Van De Ville, and P. Vuilleumier, “White-matter connectivity between face-responsive regions in the human brain,” *Cereb Cortex*, vol. 22, no. 7, pp. 1564–1576, 2012.

[68] J. Davies-Thompson and T. J. Andrews, “Intra- and interhemispheric connectivity between face-selective regions in the human brain,” *J Neurophysiol*, vol. 108, no. 11, pp. 3087–3095, 2012.

[69] Z. M. Saygin, D. E. Osher, K. Koldewyn, G. Reynolds, J. D. Gabrieli, and R. R. Saxe, “Anatomical connectivity patterns predict face selectivity in the fusiform gyrus,” *Nat Neurosci*, vol. 15, no. 2, pp. 321–327, 2011.

[70] P. Grimaldi, K. S. Saleem, and D. Y. Tsao, “Anatomical Connections of the Functionally Defined ‘Face Patches’ in the Macaque Monkey,” *Neuron*, vol. 90, no. 6, pp. 1325–1342, 2016.

[71] W. D. Stevens, D. J. Kravitz, C. S. Peng, M. H. Tessler, and A. Martin, “Privileged functional connectivity between the visual word form area and the language system,” *J. Neurosci.*, vol. 37, no. 21, pp. 5288–5297, 2017.

[72] F. Bouhali *et al.*, “Anatomical connections of the visual word form area,” *J Neurosci*, vol. 34, no. 46, pp. 15402–15414, 2014.

[73] M. Behrmann and D. C. Plaut, “Distributed circuits, not circumscribed centers, mediate visual recognition,” *Trends Cogn. Sci.*, vol. 17, no. 5, pp. 210–219, 2013.

[74] L. Cohen and S. Dehaene, “Specialization within the ventral stream: the case for the visual word form area,” *Neuroimage*, vol. 22, no. 1, pp. 466–476, 2004.

[75] M. S. Livingstone, M. J. Arcaro, and P. F. Schade, “Cortex Is Cortex: Ubiquitous Principles Drive Face-Domain Development,” *Trends Cogn. Sci.*, vol. 23, no. 1, pp. 3–4, 2019.

[76] M. J. Arcaro and M. S. Livingstone, “A hierarchical, retinotopic proto-organization of the primate visual system at birth,” *Elife*, vol. 6, 2017.

[77] S. Brem *et al.*, “Brain sensitivity to print emerges when children learn letter-speech sound correspondences,” *Proc Natl Acad Sci U S A*, vol. 107, no. 17, pp. 7939–7944, 2010.

[78] M. Ben-Shachar, R. F. Dougherty, G. K. Deutsch, and B. A. Wandell, “The development of cortical sensitivity to visual word forms,” *J Cogn Neurosci*, vol. 23, no. 9, pp. 2387–2399, 2011.

[79] S. Dehaene *et al.*, “How learning to read changes the cortical networks for vision and language,” *Science (80-.).*, vol. 330, no. 6009, pp. 1359–1364, 2010.

[80] J. Taubert *et al.*, “Amygdala lesions eliminate viewing preferences for faces in rhesus monkeys,” *Proc Natl Acad Sci U S A*, vol. 115, no. 31, pp. 8043–8048, 2018.

[81] H. R. Rodman, “Development of inferior temporal cortex in the monkey,” *Cereb Cortex*, vol. 4, no.

5, pp. 484–498, 1994.

[82] M. J. Webster, L. G. Ungerleider, and J. Bachevalier, “Connections of inferior temporal areas TE and TEO with medial temporal-lobe structures in infant and adult monkeys,” *J Neurosci*, vol. 11, no. 4, pp. 1095–1116, 1991.

[83] L. Reich, M. Szwed, L. Cohen, and A. Amedi, “A ventral visual stream reading center independent of visual experience,” *Curr Biol*, vol. 21, no. 5, pp. 363–368, 2011.

[84] S. L. Fairhall, K. B. Porter, C. Bellucci, M. Mazzetti, C. Cipolli, and M. I. Gobbini, “Plastic reorganization of neural systems for perception of others in the congenitally blind,” *Neuroimage*, vol. 158, pp. 126–135, 2017.

[85] M. Brants, J. Bulthe, N. Daniels, J. Wagemans, and H. P. Op de Beeck, “How learning might strengthen existing visual object representations in human object-selective cortex,” *Neuroimage*, vol. 127, pp. 74–85, 2016.

[86] K. Srihasam, J. L. Vincent, and M. S. Livingstone, “Novel domain formation reveals proto-architecture in inferotemporal cortex,” *Nat Neurosci*, vol. 17, no. 12, pp. 1776–1783, 2014.

[87] M. W. Moore, C. Durisko, C. A. Perfetti, and J. A. Fiez, “Learning to read an alphabet of human faces produces left-lateralized training effects in the fusiform gyrus,” *J. Cogn. Neurosci.*, vol. 26, no. 4, pp. 896–913, 2014.

[88] G. Janzen and M. van Turennout, “Selective neural representation of objects relevant for navigation,” *Nat Neurosci*, vol. 7, no. 6, pp. 673–677, 2004.

[89] J. Weisberg, M. van Turennout, and A. Martin, “A Neural System for Learning about Object Function,” *Cereb Cortex*, vol. 17, pp. 513–521, 2007.

[90] A. C. Wong, T. J. Palmeri, B. P. Rogers, J. C. Gore, and I. Gauthier, “Beyond shape: how you learn about objects affects how they are represented in visual cortex,” *PLoS One*, vol. 4, no. 12, p. e8405, 2009.

[91] T. Hannagan, A. Amedi, L. Cohen, G. Dehaene-Lambertz, and S. Dehaene, “Origins of the specialization for letters and numbers in ventral occipitotemporal cortex,” *Trends Cogn Sci*, vol. 19, no. 7, pp. 374–382, 2015.

[92] L. J. Powell, H. L. Kosakowski, and R. Saxe, “Social Origins of Cortical Face Areas,” *Trends Cogn Sci*, vol. 22, no. 9, pp. 752–763, 2018.

[93] I. I. Gauthier, “What constrains the organization of the ventral temporal cortex?,” *Trends Cogn Sci*, vol. 4, no. 1, pp. 1–2, 2000.

[94] N. Kanwisher, “Domain specificity in face perception,” *Nat. Neurosci.*, vol. 3, no. 8, p. 759, 2000.

[95] I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore, “Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects,” *Nat Neurosci*, vol. 2, no. 6, pp. 568–573, 1999.

[96] C. M. Bukach, I. Gauthier, and M. J. Tarr, “Beyond faces and modularity: the power of an expertise framework,” *Trends Cogn. Sci.*, vol. 10, no. 4, pp. 159–166, 2006.

[97] M. Bilalić, “Revisiting the role of the fusiform face area in expertise,” *J. Cogn. Neurosci.*, vol. 28,

no. 9, pp. 1345–1357, 2016.

[98] S. Hagen and J. W. Tanaka, "Examining the neural correlates of within-category discrimination in face and non-face expert recognition," *Neuropsychologia*, vol. 124, pp. 44–54, 2019.

[99] A. Harel, S. Gilaie-Dotan, R. Malach, and S. Bentin, "Top-down engagement modulates the neural expressions of visual expertise," *Cereb. Cortex*, vol. 20, no. 10, pp. 2304–2318, 2010.

[100] R. W. McGugin, J. C. Gatenby, J. C. Gore, and I. Gauthier, "High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance," *Proc Natl Acad Sci U S A*, vol. 109, no. 42, pp. 17063–17068, 2012.

[101] B. Z. Mahon and A. Caramazza, "What drives the organization of object knowledge in the brain?," *Trends Cogn Sci*, vol. 15, no. 3, pp. 97–103, 2011.

[102] M. V Peelen, C. He, Z. Han, A. Caramazza, and Y. Bi, "Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults," *J Neurosci*, vol. 34, no. 1, pp. 163–170, 2014.

[103] E. Striem-Amit, L. Cohen, S. Dehaene, and A. Amedi, "Reading with sounds: sensory substitution selectively activates the visual word form area in the blind," *Neuron*, vol. 76, no. 3, pp. 640–652, 2012.

[104] E. Striem-Amit, S. Ovadia-Caro, A. Caramazza, D. S. Margulies, A. Villringer, and A. Amedi, "Functional connectivity of visual cortex in the blind follows retinotopic organization principles," *Brain*, vol. 138, no. Pt 6, pp. 1679–1695, 2015.

[105] E. Striem-Amit and A. Amedi, "Visual cortex extrastriate body-selective area activation in congenitally blind people 'seeing' by using sounds," *Curr Biol*, vol. 24, no. 6, pp. 687–692, 2014.

[106] C. He, M. V Peelen, Z. Han, N. Lin, A. Caramazza, and Y. Bi, "Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience," *Neuroimage*, vol. 79, pp. 1–9, 2013.

[107] X. Wang, M. V Peelen, Z. Han, C. He, A. Caramazza, and Y. Bi, "How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals," *J. Neurosci.*, vol. 35, no. 36, pp. 12545–12559, 2015.

[108] I. Gauthier and C. Bukach, "Should we reject the expertise hypothesis?," *Cognition*, vol. 103, pp. 322–330, 2007.

[109] T. J. Andrews, D. M. Watson, G. E. Rice, and T. Hartley, "Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway visual pathway," *J. Vis.*, vol. 15, no. 2015, pp. 1–12, 2015.

[110] G. E. Rice, D. M. Watson, T. Hartley, and T. J. Andrews, "Low-Level Image Properties of Visual Objects Predict Patterns of Neural Response across Category-Selective Regions of the Ventral Visual Pathway," *J. Neurosci.*, vol. 34, no. 26, pp. 8837–8844, 2014.

[111] D. M. Watson, T. Hartley, and T. J. Andrews, "Patterns of response to visual scenes are linked to the low-level properties of the image," *Neuroimage*, vol. 99, pp. 402–410, 2014.

[112] D. D. Coggan, W. Liu, D. H. Baker, and T. J. Andrews, "Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information," *Neuroimage*,

vol. 135, pp. 107–114, 2016.

[113] S. Bracci, J. B. Ritchie, and H. Op de Beeck, “On the partnership between neural representations of object categories and visual features in the ventral visual pathway,” *Neuropsychologia*, vol. 105, pp. 153–164, 2017.

[114] D. Kaiser, D. C. Azzalini, and M. V. Peelen, “Shape-independent object category responses revealed by MEG and fMRI decoding,” *J. Neurophysiol.*, vol. 115, no. 4, pp. 2246–2250, Apr. 2016.

[115] D. Proklova, D. Kaiser, and M. V. Peelen, “Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction,” *J. Cogn. Neurosci.*, vol. 28, no. 5, pp. 680–692, May 2016.

[116] M. S. Beauchamp, “See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex,” *Curr. Opin. Neurobiol.*, vol. 15, no. 2, pp. 145–153, 2005.

[117] H. Lee Masson, J. Bulthé, H. P. Op de Beeck, and C. Wallraven, “Visual and haptic shape processing in the human brain: unisensory processing, multisensory convergence, and top-down influences,” *Cereb. Cortex*, vol. 26, no. 8, pp. 3402–3412, 2016.

[118] A. Amedi, S. Hofstetter, S. Maidenbaum, and B. Heimler, “Task selectivity as a comprehensive principle for brain organization,” *Trends Cogn. Sci.*, vol. 21, no. 5, pp. 307–310, 2017.

[119] L. Reich, S. Maidenbaum, and A. Amedi, “The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches,” *Curr. Opin. Neurol.*, vol. 25, no. 1, pp. 86–95, 2012.

[120] J. Gomez, M. Barnett, and K. Grill-Spector, “Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex,” *Nat. Hum. Behav.*, p. 1, 2019.