Opinion

# Origins of the specialization for letters and numbers in ventral occipitotemporal cortex

**Thomas Hannagan**[1,2], **Amir Amedi**[3,4,5,6], **Laurent Cohen**[7,8,9,10],
**Ghislaine Dehaene-Lambertz**[1,2,11], **and Stanislas Dehaene**[1,2,11,12]

[1] Institut National de la Santé et de la Recherche Médicale (INSERM), Unité 992, Cognitive Neuroimaging Unit, Gif-sur-Yvette, France
[2] Commissariat à l'Energie Atomique (CEA), Division Sciences de la Vie (DSV), Institut d'Imagerie Biomédicale (I2BM), NeuroSpin Center, Paris, France
[3] Department of Medical Neurobiology, The Institute for Medical Research Israel–Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 91220, Israel
[4] The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem 91220, Israel
[5] The Cognitive Science Program, The Hebrew University of Jerusalem, Jerusalem 91220, Israel
[6] Sorbonne Universités, Université Pierre et Marie Curie (UPMC), Université Paris 06, Institut de la Vision, Unité Mixte de Recherche en Santé (UMRS) 968, Paris 75012, France
[7] INSERM, Centre de Recherches de l'Institut du Cerveau et de la Moelle Epinière, UMRS 975, Paris, France
[8] UPMC–Paris 6, Faculté de Médecine Pitié-Salpêtrière, 75013 Paris, France
[9] Assistance Publique-Hopitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière, Department of Neurology, 75651 Paris, France
[10] Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 7225, Paris, France
[11] Université Paris Sud, Orsay, France
[12] Collège de France, Paris, France

**Deep in the occipitotemporal cortex lie two functional regions, the visual word form area (VWFA) and the number form area (NFA), which are thought to play a special role in letter and number recognition, respectively. We review recent progress made in characterizing the origins of these symbol form areas in children or adults, sighted or blind subjects, and humans or monkeys. We propose two non-mutually-exclusive hypotheses on the origins of the VWFA and NFA: the presence of a connectivity bias, and a sensitivity to shape features. We assess the explanatory power of these hypotheses, describe their consequences, and offer several experimental tests.**

## The puzzle of symbol form areas in the brain

Written symbols such as letters or numbers are a late but far-reaching addition to the mental toolkit of humanity. It should therefore be no surprise that the brain dedicates significant resources to recognizing them. More surprising is that these resources, the VWFA and the NFA, are always localized and highly reproducible in the occipitotemporal cortex across subjects, fonts, and even sensory modalities. This is especially puzzling because letters and numbers are such recent cultural inventions that natural selection could not have shaped brain areas for the very purpose of processing them [1].

We review recent progress made in characterizing the origins of symbol form areas in children or adults, sighted or blind subjects, and humans or monkeys. Pieced together, these studies suggest a picture that is more interesting

### Glossary

**Deep convolutional networks:** neural network systems inspired from the primate brain, which classify patterns using a hierarchy of layers with restricted receptive fields, and interleaved pooling and sampling units.

**Diffusion tensor imaging (DTI)/tractography:** DTI is a brain-imaging technique whereby the properties of neural tissues are inferred from the pattern of diffusion of water molecules as measured by MRI. Tractography is the application of computerized algorithms to DTI images of white matter to reveal the pattern of structural connectivity between brain areas.

**EyeMusic/soundscapes:** sensory substitution devices whereby visual images can be turned into sounds using isomorphic mapping through time, frequency, and intensity modulations. Unlike the relation between an image and its spoken word label, this results in a systematic relation between visual and auditory dimensions across all stimuli.

**Fractional anisotropy (FA):** in DTI studies, a measure of the consistency of water diffusion in a given spatial direction. In white matter, FA provides an estimate of axonal tract organization and myelination.

**McGurk effect:** an illusion in which a heard syllable is subjectively changed by the simultaneous vision of a mouth speaking a different syllable.

**Metamodal theory:** the theoretical view that all brain regions, including those commonly considered unimodal, are essentially characterized by the representation or computation that they support (e.g., invariant visual recognition; sensori-motor transformation; number manipulation; etc.) rather than by their main input modality (e.g., visual, auditory, tactile, etc.).

**Predictive coding:** a set of principles, architectures and learning algorithms sharing the common stance that neural circuits are engaged in predicting their inputs.

**Prosopagnosia:** a selective deficit in visual face recognition that often leaves intact the recognition of other visual stimuli or of individuals through other modalities.

**Pure alexia:** a selective deficit in word recognition which severely impairs reading. Lesions affecting the visual word form area (VWFA), its afferents, or efferents may cause pure alexia.

**Split-brain:** a pathological condition in which callosal connections between the two brain hemispheres are absent, cut, or lesioned.

---

*Corresponding author:* Hannagan, T. (thomas.hannagan@cea.fr).

*Keywords:* VWFA; NFA; metamodal areas; cortical topology; structural connectivity; invariance.

1364-6613/

http://dx.doi.org/10.1016/j.tics.2015.05.006

than ever, possibly involving a reconsideration of organizing principles in the occipitotemporal cortex, and an acknowledgment of the role played by genes that were evolved for other purposes. On this basis we propose and evaluate two hypotheses to explain the origins of symbol form areas: the biased connectivity hypothesis and the shape hypothesis.

### *Symbol form areas in blind and sighted adults*

The existence of a number form area was long postulated on the basis of neuropsychological studies [2], but its discovery was impeded by severe fMRI signal dropout at and around its cortical location as a result of the pneumatized mastoids. Electrocorticography first revealed that the right- and, to a lesser extent, left-hemispheric sites in the ventral part of the inferior temporal gyrus (ITG) are significantly more activated during the identification of visual number symbols (i.e., Arabic numerals, e.g., 2) than for visually similar letters, false fonts, or even written and spoken number words such as 'two' (Figure 1C) [3]. Despite stronger ventral activations for numbers in the right hemisphere, there is evidence from split-brain (see Glossary) and pure alexic patients [4–6] that both the right and left hemispheres can identify Arabic numerals, suggesting that the NFA is actually bilateral. It is unclear whether the left NFA is merged with the VWFA or whether there are two distinct areas [7].

Crucially, the existence of a specialization for numerals at the site of the right NFA was recently extended to congenitally blind subjects (Figure 1D) [8]. In this fMRI study, signal dropout was circumvented by a novel signal thresholding method that ensured the exclusion of voxels with attenuated signal intensity [8]. Using a sensory substitution device (the EyeMusic SSD) [9] that transforms images into sounds ('soundscapes'), and an experimental design that rules out the possibility of confounds due to stimuli differences between conditions, researchers could show that essentially the same area (located at Talairach coordinates 53, −44, −12; slightly anterior to [3]) is preferentially activated in blind subjects during the processing of sound-encoded Roman numerals. Far from being isolated, this finding is only the most recent in a series showing that the 'visual' cortex continues to show category-selectivity in the blind (Box 1).

Resting-state functional connectivity analysis [8] further established that the rITG in both blind and sighted subjects is coactivated with the right intraparietal sulcus (rIPS), a brain region involved in the supramodal representation of numbers [10].

### *Visual word form area*

The VWFA (Figure 1A) is a small area downstream of the ventral visual system, in the left lateral occipitotemporal sulcus (OTS), which is consistently activated by visual letters and words [11].

The VWFA responds to strings of letters, irrespective of low-level visual properties such as letter size and case. It is reproducibly situated, around Talairach coordinates −42, −57, −12, across subjects and writing scripts. It is highly attuned to letter identity, letter order, and even the lexical status of short letter-strings [12–15]. It is much less responsive to visual objects matched to words for visual complexity [16], and usually remains inactive during passive exposure to spoken words [17], although top-down

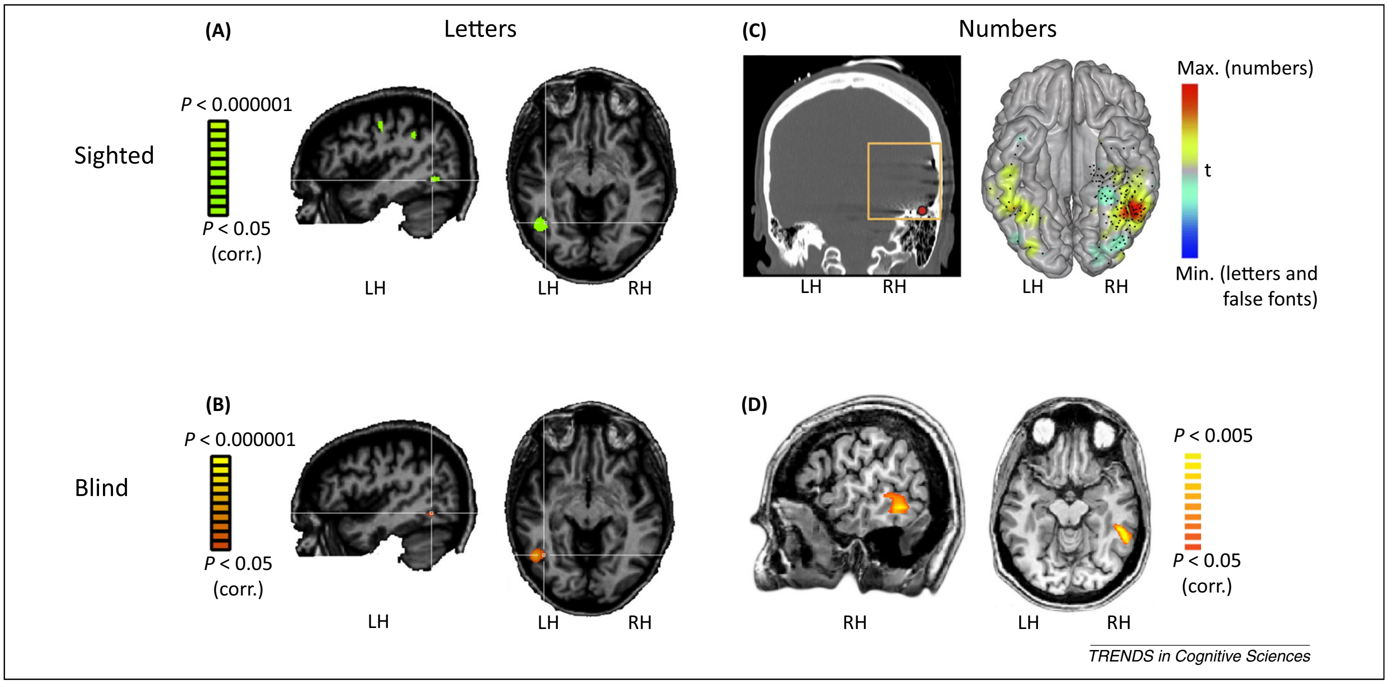


**Figure 1.** Letter and number form areas in sighted and congenitally blind subjects. (Left) Lateral and transverse views showing preferential activation in sighted adults for visual letters **(A)** and in blind adults for soundscape letters **(B)** in the left occipitotemporal sulcus (adapted and reprinted from [26], with permission, from Elsevier). The peak coordinates of −45, −58, and −5 are identical across groups, falling almost exactly at the canonical VWFA coordinates. (Right) **(C)** Preference for numbers in the sighted in the electrocorticography study of [3]. Coronal view (left): location of an intracranial electrode preferentially activated for numbers in the right inferior temporal gyrus. Ventral view (right): selectivity for numbers against scrambled numerals (peak activation MNI 51, −54, 24; adapted and reprinted from [3], with permission, from the Society for Neuroscience). For coherence with other figures, the original figure was flipped such that the right hemisphere appears on the right. **(D)** Preferential activation for roman numerals encoded by EyeMusic in congenitally blind subjects (peak activation 53, −44, −12; adapted and reprinted from [8] under a Creative Commons license CC-BY).

### Box 1. Brain areas as modality-independent task-operators

A widely shared approach in neuroscience distinguishes between unimodal regions and higher-order multimodal regions in the cortex, the latter integrating information from the former (the sensory division-of-labor principle [63]).

In the past decade, however, a wealth of evidence has cast serious doubts on this view. Studies of the visual system have revealed that many parts of the brain which were heretofore considered 'visual' according to this theory can be driven by any sensory modality and retain their functional selectivity even in the absence of visual experience [62,64–67]. Taken to the extreme, one could ask whether the entire brain can be driven and activated by any sensory modality [67].

A case in point is the lateral occipital tactile-visual area (LOtv), situated in the LO object recognition complex [62,64]. Because it is activated during visual and tactile object recognition, but not by auditory object labels, which lack geometry, it was suggested that specialization in LOtv was really shaped by the fundamental computation it engaged in – building a 3D geometric model of objects – rather than by its sensory input [62,68]. In the same year a different group coined the term 'metamodality' for this type of modality-independent computation [66]. More support for this view was brought by findings that both early blind and sighted individuals activate LOtv during tactile object recognition, irrespective of handedness or visual field [69], that the area is prone to crossmodal repetition-suppression effects (as revealed by a novel fMRI cross-modal adaptation approach) [70], and that it can be activated during soundscape presentation [62], but only when subjects actively learn to extract geometric shapes from these sounds.

A further and more general example of modality-independent task-selectivity can be found in the ventral/dorsal dichotomy within the visual cortex: the same differentiation of the basic dorsal and ventral visual streams occurs in congenitally blind subjects [71]. Since then, many other visual areas have shown a similar recruitment by other senses in the blind (see [68] for review). These include the VWFA for reading Braille [25] or reading via soundscapes [26], the motion-selective middle temporal areas (MT) for non-visual motion [72–77], the extrastriate body area for body image [78], the mirror network for auditory perception of action [77], the middle occipital gyrus for sound localization [79,80], and parieto-occipital regions in reaching tasks [81]. Listening to sound echoes can also activate the visual (calcarine) cortex rather than the auditory cortex in blind echolocation experts [82].

activation can be demonstrated in some tasks (e.g., lexical decision or rhyming tasks [18,19]). The maturation of a functional response to letter strings in the VWFA closely tracks the acquisition of reading [20,21] and colateralizes to the same hemisphere as spoken language [22,23].

As in the case of numbers, a key insight into the nature of the VWFA is given by its presence, at exactly the same location, in the congenitally blind. Braille reading in blind subjects activates the VWFA [24], and significantly more than during control tasks such as nonsensical tactile stimuli [25]. Recent experiments have also shown that, in congenitally blind subjects, the VWFA becomes selectively activated after a few hours of auditory exposure to soundscapes of words, relative to soundscapes of textures or of complex objects (Figure 1B) [26].

These studies demonstrate that the VWFA and the NFA are reproducible across subjects and that their locations are actually not determined by features of the stimuli in the visual modality: the areas appear to be metamodal.

### *The maturation of symbol form areas*

Symbol form areas have reproducible locations not only in absolute coordinates but also relative to other form areas: the VWFA, for instance, is consistently located immediately lateral to the face-selective left fusiform face area (FFA) [27], while the NFA is consistently more lateral, as is the fusiform body area that js selective for body parts (FBA) [28]. It is therefore likely that understanding the computations performed by symbol form areas will require consideration of how the larger cortical map of which they are part emerges (see [29] for review). Although computational models currently fail to account for the emergence of specialized areas and for human-level recognition performance (Box 2), several recent experimental studies have shed light on how symbol form areas can emerge.

*Maturation of the VWFA in children.* One way to investigate the maturation of symbol form areas is to contrast their activation by letters and numbers with activation by

### Box 2. Deep learning networks and symbol form areas

Convolutional network architectures once inspired by the primate visual system now achieve human-level performance (or better) on classification, recognition, and localization of real-world visual stimuli [83,84]. However, although convolutional network representations can resemble neural firing patterns in IT cortex [85], such models systematically fail to produce localized category-specific areas in their spatially organized maps, let alone a reproducible organization of these areas (although see [86] for an example of emergent topography in related network models).

Conversely, self-organizing models with shallow architectures have long been known to produce topographical and topological maps under a variety of activity-dependent learning rules [87–92]. One such model [90] simulates the emergence of separate areas for letters and digits by spontaneous symmetry-breaking, but fails to explain the systematic location of these cortical territories. Self-organizing models also display sub-par performance in real-life classification and recognition tasks compared to both humans and convolutional networks.

The dissociation in our best computational models between these two criteria (performance and topological organization) shows that topological organization is not mandatory for efficiently solving the computational task of invariant visual recognition. This suggests that topological organization may only be an accident of our biology, a by-product of parochial constraints such as a finite energy budget or a cortical sheet squeezed into a limited skull volume. For similar reasons of parsimony, the metamodality of symbol form areas may simply reflect a general implementation strategy in this part of the cortex to avoid reduplication of circuits devoted to the same task (recognizing shapes) whenever the shape is encoded through different modalities (e.g., visual vs somatosensory). Alternatively, metamodality could have a definite computational purpose. Within a predictive coding framework, for instance [93–95], a single area predicting the input it gets from several modalities may, assuming systematically correlated inputs, discover more robust and efficient representations than those within reach of two distinct unimodal areas.

An important task for future computational neuroscience is to develop a unified model of the emergence of symbol form areas that can: (i) operate at a human level of performance on letter and number recognition, (ii) account for the reproducible location of symbol form areas in absolute Talairach coordinates and relative to other areas, (iii) account for the metamodality of symbol form areas, and (iv) display the developmental trajectory observed in humans. Such a model may require combining multi-sensory deep convolutional networks with self-organizing models, in a way that is informed by the cortical folds in the vOTC.

other visual categories (e.g., faces and shoes) in preschoolers and in adults. In this way, it was established that children aged 5 years who could easily name letters and numbers, but who had not yet learned to read and operate with numbers at school, already exhibited a dissociation between faces and letters at the canonical site of the VWFA in adults [30]. However, at age 5 this site still showed identical responses for letters and numbers, suggesting that automatic and fluent association with their meaning and thus connections to specific target circuits may be needed before these symbols segregate at specific cortical locations.

*Longitudinal studies of the reading circuit.* Three independent longitudinal studies have recently examined how the reading circuit of children develops. In a joint electroencephalogram (EEG)/fMRI study before, during, and after children become literate, researchers have shown that the onset of VWFA selectivity for visual words coincides with the acquisition of grapheme-to-phoneme correspondences [31]. Moreover, the fractional anisotropy of water diffusion – a measure of how compact and/or myelinated fiber tracts are – along the inferior longitudinal fasciculus which connects the VWFA to the anterior and medial temporal lobe was found to increase with both age and reading proficiency in a longitudinal study of children aged 7–12 years [32]. In an independent voxel-based morphometric study, evidence was found that, by age 5–6, several regions involved in language processing are connected to the VWFA, the volume of which increases as reading skills improve [33].

*Literacy acquisition in adults.* Comparison of literate and illiterate adults has established that learning to recognize letter strings produced a modest but significant competition with face and checkerboard stimuli at the canonical site of the VWFA, resulting in shrinking of face-selective responses to within 10 mm of the VWFA and a shift to right-hemisphere FFA ([20], see also [34,35] for a similar observation reported in the comparison of the responses of dyslexic and normal readers at 9 years of age). In a follow-up study with the same adult subjects, the posterior arcuate fasciculus – which connects the VWFA to posterior superior temporal and inferior parietal regions involved in phonological representations (among others) – was found to be reinforced by the acquisition of literacy [36].

Although most data on the maturation of symbol form areas concern the VWFA, the available evidence in this case suggests that the key to the reproducible location of this area may reside in a yet to be understood preference for the visual shapes of faces and symbols, and/or in privileged connectivity with perisylvian language areas.

### *Monkey models of symbol acquisition*

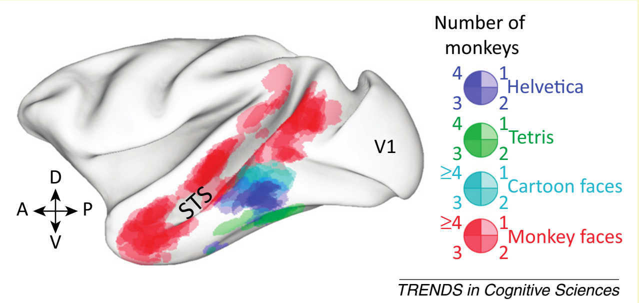
Whether form areas could have evolved specifically for written symbols is not a subject of debate: written symbols are too recent an invention, and until one century ago concerned too marginal a fraction of the population, for genetic evolution to have had an impact. However, two distinct lines of evidence suggest that these symbol form systems are shaped by general constraints shared by Old World monkeys and humans. These findings are detailed in Box 3. First, a proto-symbol form area that is arguably the precursor of the NFA was recently shown to emerge in the IT cortex of juvenile macaques trained to map symbols onto quantities [37,38]. Second, baboons were trained to process strings of letters and to distinguish between English words and non-words [39] in a

### Box 3. Monkey models for the acquisition of visual symbols

Compelling evidence for a proto-area similar to the VWFA or NFA can be found in macaques. Juvenile macaques trained to recognize Helvetica characters will develop, at a reproducible location in IT cortex, a specialized form area [37]. A follow-up experiment extended training to three types of stimuli: cartoon faces, Tetris blocks, and Helvetica characters, and could establish that areas selective for each of these stimuli had appeared by the end of training in the IT cortex, with the Helvetica site lying close to face regions (Figure I) [38].

The training scheme in this study involved associating each symbol with the number of drops of juice that the monkey would receive as a reward for choosing that character. Because of this association to numerical magnitude, anything the proto-area should be considered to be a homolog of the NFA rather than of the VWFA. However, establishing homologies between specialized areas in the IT cortex of monkeys and the human OTC is a delicate task [96]. The location of monkey and human face patches suggests a major migration of lateral inferotemporal visual recognition circuits towards the ventral side in humans [97], congruent with the emergence of language circuits in middle temporal cortex. Because the monkey proto-number area is ventral to face patches, such a ventral shift would predict that the human NFA should be medial to the FFA, whereas it actually lies lateral to it, in the ventral inferior temporal gyrus. It is possible that the establishment of homology will require both a shift and a mirror-image flip of cortical sites [98].

Independently, behavioral studies have shown that baboons, after training on tens of thousands of trials, could successfully tell written English words and non-words apart [39]. Baboon performance remained above chance even for novel stimuli that had never been seen during training. Moreover, the type of error patterns made by baboons (e.g., transposition effects, confusing TIME with TMIE more often than with TUNE) mimicked that of human adults engaged in the same task [40]. In the absence of brain-imaging evidence, the cortical mechanisms underlying such behavior are unknown. However, deep learning convolutional networks trained on the same sequence of stimuli as baboons can accommodate most of the experimentally observed effects [41], suggesting that baboons were engaged in orthographic coding and statistical learning of letter pairs in a manner similar to that held to be subserved by the VWFA in humans [14,99].



TRENDS in Cognitive Sciences

**Figure I.** Proto-symbol form areas in seven juvenile macaques trained to classify Helvetica characters (dark blue, the light-to-dark color gradient codes for overlap across macaques), Tetris bricks (green), and cartoon faces (turquoise). Patches of significant activation are projected onto a semi-inflated macaque brain, collapsed across the two hemispheres. The main Helvetica area is consistently located in the IT cortex, adjacent to monkey face patches (red), especially the medial face patch on the STS (superior temporal sulcus) (adapted and reprinted from [38], with permission, from Nature Publishing Group).

demonstratively human-like manner [40] suggestive of a proto-VWFA [41].

The evidence in monkeys thus seems to run counter to the data in blind human subjects, and invites the inference that symbol form areas have their origins in a visual cortex architecture shared across primates. Nevertheless, the visual features of letters and digits are virtually identical, suggesting that form alone cannot explain the existence of two distinct areas in the ventral occipitotemporal cortex (vOTC). Furthermore, symbol form areas activate whenever letter and digit stimuli are presented in non-visual modalities. How can these findings be reconciled?

## Two hypotheses for symbol form areas in vOTC

We spell out two hypotheses on the initial conditions that lead to the maturation of the letter and number form areas depicted in Figure 1. These hypotheses are not mutually exclusive but have distinct consequences.

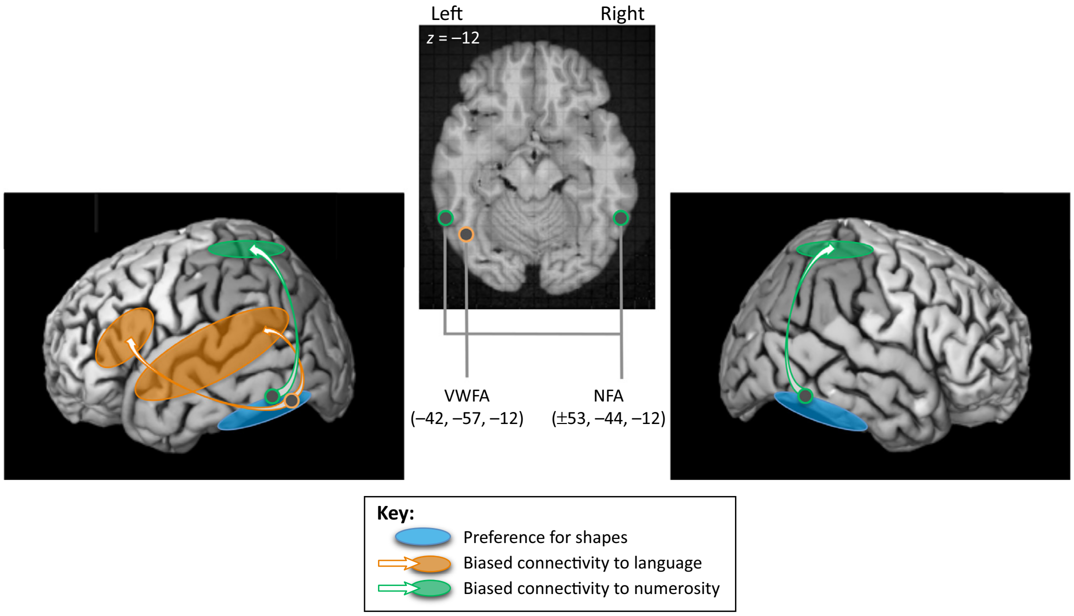
### *The biased connectivity hypothesis*

The 'biased connectivity' hypothesis proposes that form areas emerge at cortical sites that exhibit a higher density of white-matter fiber tracts to and from the cortical circuits that are crucial for the target task (Figure 2, white arrows). In the case of the VWFA, these target circuits are the perisylvian spoken language areas that antedate reading – because they are demonstrably present even in 3-month-old infants [42]. In the case of the NFA, the primary target circuits would be the bilateral intraparietal sites that encode non-symbolic numerical quantities, a competence that is again demonstrably present in infancy [43]. The biased connectivity hypothesis builds upon several previous studies that link connectivity patterns to functional specialization for faces and other categories in the vOTC [25,44–46].

The biased connectivity hypothesis can explain why form areas are conserved at the same coordinates in normal sighted readers and in blind subjects. According to this view, where a symbol form area should mature in the cortex is not so much determined by the visual properties of the stimulus as by some pre-existing (possibly innate) structural connections at this site with the regions that are targeted by the symbol. In both blind and sighted subjects, lateral inferotemporal regions (including NFA) would show a biased connectivity with the IPS, and slightly more mesial regions (including VWFA) to the perisylvian language cortex.

The hypothesis is consistent with what is known of the adult functional and structural connectivities of symbol form areas. First and foremost, the two resting-state functional connectivity networks obtained with seed locations at the NFA and VWFA, respectively, have recently been contrasted [8]. This analysis revealed that the former network is significantly more connected to regions that are involved in representing quantities (such as the rIPS), whereas the latter is significantly more connected to the left temporal cortex and the left inferior frontal gyrus (lIFG), which are notably recruited in language processing. Importantly, this contrast in functional connectivities was shown to be similar in sighted and blind subjects. Another study did, however, report one important discrepancy in the connectivity of the blind relative to the sighted [26]: only in the former did resting-state analyses with a VWFA seed provide evidence for a functional connection between the VWFA and the AWFA (auditory word form area, an area selectively activated by spoken words [47]).

Tractographic studies starting from the VWFA further support these functional connectivity findings. A review of the cortical circuits involved in reading has identified three main underlying tracts: the arcuate fasciculus, the



TRENDS in Cognitive Sciences

**Figure 2.** Two converging hypotheses for the origins of symbol form areas: biased connectivity (white arrows) and a preference for shapes (blue areas). (Left) Left lateral view. (Middle) Axial cut ($z = -12$). (Right) Right lateral view. Arrows and areas are schematic and only indicative. All brain views were generated using the Allen Human Brain Atlas software [100]. Abbreviations: NFA, number form area; VWFA, visual word form area.

posterior corpus callosum, and the inferior longitudinal fasciculus, the latter connecting the VWFA to the anterior and medial temporal lobe [48]. Crucially, the structural connectivity of the VWFA has recently been directly compared to that of the FFA [49], revealing that the former is significantly more connected than the latter to several left-lateralized perisylvian language areas engaged in mapping grapheme to phonemes (posterior MTG), or in access to meaning (anterior basal temporal lobe). At present, no such structural connectivity data are available for the NFA.

It is important to note that such connectivity studies in adult subjects, although compatible with the biased connectivity hypothesis, do not constitute definitive evidence for it either, and for two reasons. First, they are performed in skilled readers, after the acquisition of reading, and the observed biased connectivity might be a consequence rather than a cause of reading acquisition at this cortical site [36]. Second, it is not known whether other adjacent areas might initially possess the same connectivity, but ultimately fulfill another function. Positive evidence (as opposed to absence of negative evidence) could only be obtained by systematically mapping the connectivity of the inferior temporal cortex to the rest of the brain before subjects learn how to read and to count with visual symbols. Although they fall short of the above criterion, the abovementioned DTI [32] and morphometric studies [33] have established early connectivity patterns between the VWFA and language networks. Both studies therefore lend support to the biased connectivity hypothesis.

If the reproducibility of form areas comes from gradients of connectivity that are possibly genetically encoded, one may ask which genes could be involved and for what purpose, if any, they evolved. The available data, in the case of the VWFA, point to a genetically determined colateralization between this area and the left superior temporal sulcus (STS) [50], which is activated for speech. One interpretation, then, would be that such genes are crucial for building neural circuits that can associate shapes to sounds. In this view, preferential connections between the pre-VWFA and the left STS would essentially have evolved for naming visual objects, and were later pre-empted by visual letters. Alternatively, the initial function of these repurposed connections could have had more to do with lip-reading (Box 4), an idea which has the advantage of explaining why word and face regions always lie next to each other in occipitotemporal cortex.

The biased connectivity hypothesis predicts that children should already exhibit white-matter tracts from the future VWFA to language regions, and from the NFA to numerosity regions, before they learn what letters and numbers refer to. The hypothesis can be directly tested by tractography analyses. Another prediction derived from this hypothesis is that, if one were to adapt the experiment in [38] so that two different systems of symbols were respectively associated to monkey communication calls and to quantities, then juvenile monkeys should develop two well-differentiated number and letter form areas after training.

### *The shape hypothesis*

The shape hypothesis assumes no special initial connectivity pattern between form areas and target circuits in higher associative cortex, but holds that the intrinsic circuitry of the inferotemporal cortex makes neurons in the VWFA and the NFA particularly apt at recognizing the shapes of objects (Figure 2, blue regions). In the following we define a shape as a representation of the adjacency of the component parts of an object [51,52] that is at least partially invariant to translation, reflection, rotation, distance, and other variations in the stimulus. Note that this definition does not imply a visual input: equivalent shape information can be extracted from auditory soundscapes. The hypothesis, made mathematically precise in [53], is that the circuitry of the ventral occipitotemporal pathway is particularly apt at extracting from any sensory input, regardless of modality, a representation of its invariant shape features (whereas the dorsal occipitotemporal pathway would have the 'dual' task of extracting the distance

#### Box 4. From lip-reading to word-reading

The biased-connectivity hypothesis proposes that the VWFA and NFA owe their location to a pre-existing pattern of connections from these sites, respectively, to speech- and number-related areas. However, why would such connections exist in the first place?

One possibility is that such connections were required, during our evolution, to efficiently link (i) visual recognition of faces with temporal lobe speech-recognition mechanisms and (ii) visual recognition of bodies with parietal lobe action-recognition mechanisms. Because letters require mapping onto speech sounds, their acquisition would encroach onto the face-to-speech circuit. Because numbers require mapping onto quantity representations located amidst other parietal lobe circuits for spatial representation and action, they would encroach onto the body-to-action circuit. The relative location of these projections would then explain the lateral location of the NFA and the mesial location of the VWFA.

This account implies that those circuits pre-date reading and number acquisition in both development and evolution. Although no study has yet assessed this specific possibility, resting-state functional imaging in newborns has shown that the left temporal cortex, the right parietal lobule, and the right posterior occipitotemporal cortex constitute three cortical hubs in the infant brain [101]. Furthermore, human infants aged 10 weeks are known to associate lip movements to sounds and are sensitive to their congruence [102] – they already exhibit a McGurk effect [103]. Interestingly, whereas fMRI studies on children show that the left pSTS is the most important site for AV integration (lip-reading), bilateral activity in the fusiform gyrus also correlates with the size of the McGurk effect [104], suggesting early functional connections between left pSTS and the left and right fusiform gyri. Maturational asymmetries in the posterior STS are also known to be associated with asymmetries in the arcuate fasciculus [105,106], which later in childhood connects the VWFA to language circuits [32]. Finally, a close relationship between letter recognition and lip-reading in the left VWFA-STS circuit is implied by a case study of one prosopagnosic and one alexic patient, showing that impaired lip-reading was observed only in the latter [107].

Taken together, these studies suggest that the connections between the future site of the VWFA and the left STS plausibly pre-date the acquisition of reading, and could have evolved for the purpose of lip-reading, thus providing an efficient connection from visual face recognition to spoken language early in child development. This system would later be repurposed for literacy acquisition, thus explaining why the VWFA is systematically located next to face-responsive cortex and competes with it during reading acquisition [18,20].

and angle of objects, regardless of their identity). In this view, initial retinotopic visual processing stages abstract away from the geometric transformations common to all rigid objects (e.g., size and translation invariance), whereas the next cortical stages specialize for invariance over transformations which are unique to a class of objects (e.g., the different expressions and lightings of a given face; the different cases and fonts for a given letter) [53]. Each of these classes would map onto a distinct region of inferotemporal cortex because recognizing them implies abstracting away from a distinct group of transformations.

The shape hypothesis can explain why images, soundscapes, and Braille preferentially activate the VWFA and the NFA, whereas spoken words selectively activate the AWFA [47] but not the VWFA [54]. According to this view, the differences in selectivity between these areas are not due to an innate lack of inputs from other sensory channels, but to the specific tuning of neurons in the VWFA and NFA to shape features. Because soundscapes neither have the same frequency spectra nor the same invariants as speech, they do not elicit preferential activation of speech regions. On the other hand, as blind subjects become expert at reverse mapping from soundscapes to images, ventral occipitotemporal areas become able to process the recovered shape information.

A few lines of evidence support the shape hypothesis. First, children initially make mirror errors when they learn to read and write, a phenomenon which has been suggested to reflect a pre-reading competence for mirror invariance in the VWFA [55–58]. In addition, although word representations retain a measure of location-sensitivity in the VWFA, they are much more location-invariant than V1 representations and allow the recognition of words anywhere on the fovea [11,59]. Approximate location and mirror invariance in the VWFA are consistent with the shape hypothesis.

Second, the shape hypothesis predicts that expert Morse readers should not preferentially activate the VWFA when reading in Morse, a purely temporal code that represents letters using sequences differing only in their time intervals. Indeed, although expert Morse readers recruit similar left posterior temporoparietal areas when reading words in print or in Morse, the VWFA does not stand out as a cluster of peak activation [60]. More recently, a combined functional and morphometric MRI study was conducted before and after subjects were taught Morse [61]. Again in support of the shape hypothesis, no preferential activation was found at the usual locus of the VWFA. However, the density of white-matter fibers afferent to the VWFA increased after training, suggesting that the region is still somehow involved in Morse reading.

Finally, in one study, a group of sighted subjects were taught to merely associate soundscapes with their names, without being told about the underlying SSD image-to-sound mapping algorithm [62]. In this condition, soundscapes failed to activate the lateral occipital tactile-visual area (LOtv) region, suggesting that an explicit reverse mapping from sounds to shapes is needed before the region can be recruited.

The shape hypothesis also makes several new predictions. If congenitally blind subjects were trained to recognize letters and numbers through a shape-less code such as Morse, they should not show preferential activation in the NFA or in the VWFA. On the other hand, if the hypothesis holds for both blind and sighted subjects, the latter should also show preferential activation after training for soundscape stimuli in the VWFA (respectively the NFA). Similarly, sighted subjects presented with spoken numbers should not show preferential activation in the NFA, and sighted Morse experts should not show preferential activation for Morse numbers in the NFA.

**Box 5. Outstanding questions**

- Can the future cortical sites of the visual number and word form areas already be predicted during infancy on the basis of white-matter connectivity? Do these sites exhibit any selectivity before learning?
- How can the architecture of cortical columns and horizontal connections in vOTC be biased to detect invariant shapes in its inputs, and continue to do so even when the shapes are conveyed through the auditory modality?
- How can a simulated network be designed to systematically remove the superficial variance due to geometric transformations such as translation, rotation, or size, and only remain sensitive to core shape invariants?
- What type of computational model of letter and number recognition could simultaneously perform at a level comparable to humans, and account for the emergence of a reproducible cortical topology for those shapes?

## Concluding remarks

Symbol form areas are highly reproducible in vOTC across subjects of different cultures, and localized proto-symbol form areas also consistently emerge in IT cortex after training juvenile macaques. Nonetheless, these areas could not have specifically evolved for processing symbols, and their locations, which are conserved in congenitally blind humans, cannot proceed from purely visual constraints. Our analysis of the available data suggests that two hypotheses, the biased connectivity and the shape hypotheses, are equally needed to reconcile the puzzling characteristics of symbol form areas. Some sectors of the cortex, such as the vOTC, may be tuned to invariant shapes. Independently, some cortical sectors may exhibit a stronger connectivity to language- or number-related regions. Our proposal is that symbol form areas always emerge at the intersection of these two constraints. This proposal is coherent with the available data, but it also raises several new questions (Box 5) and makes falsifiable predictions. Perhaps the most provocative of these, which remains untested, is that connectivity analyses in infants should already reveal white-matter tracts that predict the precise cortical location where visual number and word form areas will later preferentially develop.

## Acknowledgments

The research leading to these results has received funding from the European Commission Seventh Framework Programme (FP7/2007-2013) under grant agreement 604102 (Human Brain Project), from the program ‘Investissements d’Avenir’ ANR-10-IAIHU-06, and was supported by INSERM, the CEA, the Collège de France, and the Bettencourt-Schueller foundation. A.A. is a European Research Council (ERC) fellow and is supported by ERC-ITG grant (310809) as well as the Gatsby Charitable Foundation and a James S. McDonnell Foundation scholar award for understanding human cognition (grant 220020284).

## References

1 Dehaene, S. (2009) *Reading in the Brain*, Penguin Viking
2 Dehaene, S. and Cohen, L. (1995) Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120
3 Shum, J. *et al.* (2013) A brain area for visual numerals. *J. Neurosci.* 33, 6709–6715
4 Dejerine, J.J. (1892) Contribution à l'étude anatomo-pathologique et clinique des différentes variétés de cécité verbale. *Mem. Soc. Biol.* 4, 61–90
5 Cohen, L. and Dehaene, S. (1995) Number processing in pure alexia: the effect of hemispheric asymmetries and task demands. *Neurocase* 1, 121–137
6 Cohen, L. and Dehaene, S. (1996) Cerebral networks for number processing: evidence from a case of posterior callosal lesion. *Neurocase* 2, 155–174
7 Starrfelt, R. and Behrmann, M. (2011) Number reading in pure alexia – a review. *Neuropsychologia* 49, 9 2283–2298
8 Abboud, S. *et al.* (2015) A number-form area in the blind. *Nat. Commun.* 6, 6026
9 Abboud, S. *et al.* (2013) EyeMusic: introducing a 'visual' colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* 32, 247–257
10 Nieder, A. and Dehaene, S. (2009) Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208
11 Cohen, L. *et al.* (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307
12 Dehaene, S. *et al.* (2004) Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. *Psychol. Sci.* 15, 307–313
13 Dehaene, S. *et al.* (2005) The neural code for written words: a proposal. *Trends Cogn. Neurosci.* 9, 335–341
14 Vinckier, F. *et al.* (2007) Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron* 55, 143–156
15 Glezer, L.S. *et al.* (2015) Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *J. Neurosci.* 35, 4965–4972
16 Szwed, M. *et al.* (2011) Specialization for written words over objects in the visual cortex. *Neuroimage* 56, 330–344
17 Dehaene, S. *et al.* (2002) The visual word form area. A prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13, 321–325
18 Dehaene, S. and Cohen, L. (2011) The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262
19 Yoncheva, Y.N. *et al.* (2010) Auditory selective attention to speech modulates activity in the visual word form area. *Cereb. Cortex* 20, 622–632
20 Dehaene, S. *et al.* (2010) How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364
21 Ben-Shachar, M. *et al.* (2011) The development of cortical sensitivity to visual word forms. *J. Cogn. Neurosci.* 23, 2387–2399
22 Pinel, P. and Dehaene, S. (2010) Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *J. Cogn. Neurosci.* 22, 48–66
23 Cai, Q. *et al.* (2010) The left ventral occipito-temporal response to words depends on language lateralization but not on visual familiarity. *Cereb. Cortex* 20, 1153–1163
24 Büchel, C. *et al.* (1998) Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain* 121, 409–419
25 Reich, L. *et al.* (2011) A ventral visual stream reading center independent of visual experience. *Curr. Biol.* 21, 1–6
26 Striem-Amit, E. *et al.* (2012) Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 640–652
27 Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for the perception of faces. *J. Neurosci.* 17, 4302–4311
28 Schwarzlose, R.F. *et al.* (2005) Separate face and body selectivity on the fusiform gyrus. *J. Neurosci.* 25, 11055–11059
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 Cantlon, J.F. *et al.* (2011) Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb. Cortex* 21, 191–199
31 Brem, S. *et al.* (2010) Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7939–7944
32 Yeatman, J.D. *et al.* (2012) The development of white matter and reading skills. *Proc. Natl. Acad. Sci. U.S.A.* 109, E3045–E3053
33 Simon, G. *et al.* (2013) Dynamics of the anatomical changes that occur in the brains of schoolchildren as they learn to read. *PLoS ONE* 8, e81789
34 Pegado, F. *et al.* (2014) Timing the impact of literacy on visual processing. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5233–E5242
35 Monzalvo, K. *et al.* (2012) Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status. *Neuroimage* 61, 258–274
36 Thiebaut de Schotten, M. *et al.* (2012) Learning to read improves the structure of the arcuate fasciculus. *Cereb. Cortex* 24, 989–995
37 Srihasam, K. *et al.* (2012) Behavioral and anatomical consequences of early versus late symbol training in macaques. *Neuron* 73, 608–619
38 Srihasam, K. *et al.* (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 12, 1776–1783
39 Grainger, J. *et al.* (2012) Orthographic processing in baboons (*Papio papio*). *Science* 336, 249–255
40 Ziegler, J.C. *et al.* (2013) Transposed-letter effects reveal orthographic processing in baboons. *Psychol. Sci.* 24, 1609–1611
41 Hannagan, T. *et al.* (2014) Deep learning of orthographic representations in baboons. *PLoS ONE* 9, e84843
42 Dehaene-Lambertz, G. *et al.* (2002) Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015
43 Izard, V. *et al.* (2008) Distinct cerebral pathways for object identity and number in human infants. *PLoS Biol.* 6, e11
44 Mahon, B.Z. and Caramazza, A. (2011) What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* 15, 97–103
45 Saygin, Z.M. *et al.* (2012) Anatomical connectivity patterns predict face-selectivity in the fusiform gyrus. *Nat. Neurosci.* 15, 321–327
46 Osher, D.E. *et al.* (2015) Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cereb. Cortex* Published online January 26, 2015. http://dx.doi.org/10.1093/cercor/bhu303
47 DeWitt, I. and Rauschecker, J.P. (2012) Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 109, 505–514
48 Wandell, B.A. *et al.* (2012) Learning to see words. *Annu. Rev. Psychol.* 63, 31–53
49 Bouhali, F. *et al.* (2014) Anatomical connections of the visual word form area. *J. Neurosci.* 34, 15402–15414
50 Pinel, P. *et al.* (2014) Genetic and environmental influences on the visual word form and fusiform face areas. *Cereb. Cortex* Published online May 13, 2014. http://dx.doi.org/10.1093/cercor/bhu048
51 Tsunoda, K. (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* 4, 832–838
52 Brincat, S.L. and Connor, C.E. (2006) Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49, 17–24
53 Anselmi, F. and Poggio, T. (2014) Representation learning in sensory cortex: a theory. *CBMM Memo* 26
54 Cohen, L. *et al.* (2004) Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage* 23, 1256–1270
55 Dehaene, S. (2010) Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. *Neuroimage* 49, 1837–1848
56 Pegado, F. *et al.* (2011) Breaking the symmetry: mirror discrimination for single letters but not for pictures in the visual word form area. *Neuroimage* 55, 742–749
57 Logothetis, N. *et al.* (1995) Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563
58 Freiwald, W.A. and Tsao, D.Y. (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851
59 Rauschecker, A.M. *et al.* (2013) Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1568–E1577

60 Maier, J. *et al.* (2004) Reading with the ears. *Neurosci. Lett.* 364, 185–188

61 Schmidt-Wilcke, T. *et al.* (2010) Distinct patterns of functional and structural neuroplasticity associated with learning Morse code. *Neuroimage* 51, 1234–1241

62 Amedi, A. *et al.* (2007) Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.* 10, 687–689

63 Zeki, S. (1978) Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J. Physiol.* 277, 273–290

64 Amedi, A. *et al.* (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat. Neurosci.* 4, 324–330

65 Pascual-Leone, A. *et al.* (2005) The plastic human brain cortex. *Annu. Rev. Neurosci.* 28, 377–401

66 Pascual-Leone, A. and Hamilton, R. (2001) The metamodal organization of the brain. *Prog. Brain. Res.* 134, 427–445

67 Kayser, C. and Logothetis, N. (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–231

68 Reich, L. *et al.* (2012) The brain as a flexible task machine: implications for visual rehabilitation using noninvasive vs. invasive approaches. *Curr. Opin. Neurol.* 25, 86–95

69 Amedi, A. *et al.* (2010) Cortical activity during tactile exploration of objects in blind and sighted humans. *Restor. Neurol. Neurosci.* 28, 143–156

70 Tal, N. and Amedi, A. (2009) Multisensory visual-tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. *Exp. Brain Res.* 198, 165–182

71 Striem-Amit, E. *et al.* (2012) The large-scale organization of 'visual' streams emerges without visual experience. *Cereb. Cortex* 22, 1698–1709

72 Matteau, I. *et al.* (2010) Beyond visual, aural and haptic movement perception: hMTþ is activated by electrotactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain Res. Bull.* 82, 264–270

73 Poirier, C. *et al.* (2005) Specific activation of the V5 brain area by auditory motion processing: an fMRI study. *Cogn. Brain Res.* 25, 650–658

74 Poirier, C. *et al.* (2005) Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage* 31, 279–285

75 Ptito, M. *et al.* (2009) Recruitment of the middle temporal area by tactile motion in congenital blindness. *Neuroreport* 20, 543–547

76 Summers, I.R. *et al.* (2009) A functional-magnetic-resonance imaging investigation of cortical activation from moving vibrotactile stimuli on the fingertip. *J. Acoust. Soc. Am.* 125, 1033–1039

77 Ricciardi, E. *et al.* (2009) Do we really need vision? How blind people 'see' the actions of others. *J. Neurosci.* 29, 9719–9724

78 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

79 Collignon, O. (2011) Sensory rehabilitation in the plastic brain. *Prog. Brain Res.* 191, 211–231

80 Renier, L. and De Volder, A.G. (2010) Vision substitution and depth perception: early blind subjects experience visual perspective through their ears. *Disabil. Rehabil. Assist. Technol.* 5, 175–183

81 Lingnau, A. *et al.* (2014) Cross-modal plasticity preserves functional specialization in posterior parietal cortex. *Cereb. Cortex* 24, 541–549

82 Thaler, L. *et al.* (2011) Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE* 6, e20162

83 Sermanet, P. *et al.* (2014) OverFeat: integrated recognition, localization and detection using convolutional networks. *arXiv* 1312.6229

84 Szegedy, C. *et al.* (2015) Going deeper with convolutions. In *IEEE Computer Vision and Pattern Recognition*, pp. 1–9, IEEE

85 Yamins, D.L.K. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8619–8624

86 Gregor, K. *et al.* (2011) Structured sparse coding via lateral inhibition. *Adv. Neural Inform. Process. Syst.* 24, 4271

87 Linsker, R. (1988) Self-organization in a perceptual network. *IEEE Comput.* 21, 105–117

88 Kohonen, T. (1982) Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69

89 Kohonen, T. (2013) Essentials of the self-organizing map. *Neural Netw.* 37, 52–65

90 Polk, T.A. and Farah, M. (1998) The neural development and organization of letter recognition; evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proc. Natl. Acad. Sci. U.S.A.* 95, 847–852

91 Detorakis, G.I. and Rougier, N.P. (2012) A neural field Model of the somatosensory cortex: formation, maintenance and reorganization of ordered topographic maps. *PLoS ONE* 7, e40257

92 Stevens, J.L. *et al.* (2013) Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *J. Neurosci.* 33, 15747–15766

93 Rao, R.P.N. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87

94 Friston, K.J. (2005) A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 360, 815–836

95 Spratling, M.W. (2013) Predictive coding. In *Encyclopedia of Computational Neuroscience* (Jaeger, D. and Jung, R., eds), pp. 2491–2494, Springer-Verlag

96 Mantini, D. *et al.* (2012) Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nat. Methods* 9, 277–282

97 Tsao, D.Y. *et al.* (2008) Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* 49, 19514–19519

98 Hasson, U. *et al.* (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37, 1027–1041

99 Binder, J.R. *et al.* (2006) Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* 33, 739–748

100 Hawrylycz, M.J. *et al.* (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391–399

101 Fransson, P. *et al.* (2013) Early development of spatial patterns of power-law frequency scaling in FMRI resting-state and EEG data in the newborn brain. *Cereb. Cortex* 23, 638–646

102 Bristow, D. *et al.* (2009) Hearing faces: how the infant brain matches the face it sees with the speech it hears. *J. Cogn. Neurosci.* 21, 905–921

103 Kushnerenko, E. *et al.* (2008) Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11442–11445

104 Nath, A.R. *et al.* (2011) Neural correlates of interindividual differences in children's audiovisual speech perception. *J. Neurosci.* 31, 13963–13971

105 Dubois, J. *et al.* (2009) Structural asymmetries in the infant language and sensori-motor networks. *Cereb. Cortex* 19, 414–423

106 Leroy, F. *et al.* (2011) Early maturation of the linguistic dorsal pathway in human infants. *J. Neurosci.* 31, 1500–1506

107 Campbell, R. *et al.* (1986) Face recognition and lipreading: a neurological dissociation. *Brain* 109, 509–521