Distributed circuits, not circumscribed centers, mediate visual recognition

Marlene Behrmann and David C. Plaut

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Increasingly, the neural mechanisms that support visual cognition are being conceptualized as a distributed but integrated system, as opposed to a set of individual, specialized regions that each subserve a particular visual behavior. Consequently, there is an emerging emphasis on characterizing the functional, structural, and computational properties of these broad networks. We present a novel theoretical perspective, which elucidates the developmental emergence, computational properties, and vulnerabilities of integrated circuits using face and word recognition as model domains. Additionally, we suggest that, rather than being disparate and independent, these neural circuits are overlapping and subject to the same computational constraints. Specifically, we argue that both word and face recognition rely on fine-grained visual representations but, by virtue of pressure to couple visual and language areas and to keep connection length short, the left hemisphere becomes more finely tuned for word recognition and, consequently, the right hemisphere becomes more finely tuned for face recognition. Thus, both hemispheres ultimately participate in both forms of visual recognition, but their respective contributions are asymmetrically weighted.

What mechanisms support visual cognition?

In recent years, theorizing within cognitive neuroscience has increasingly moved away from a search for common, domain-general principles toward a view in which the brain mechanisms that support cognition are organized into discrete modules dedicated to specific, narrowly-defined functions. A clear case in point concerns visual recognition of faces and words, where neuroimaging observations of selective activation for faces in the fusiform face area (FFA; see Glossary) and for words in the visual word form area (VWFA) dovetail with classic neuropsychological findings of apparently selective deficits in face recognition (prosopagnosia) and in word recognition (pure alexia) following damage to these respective areas. In this opinion article, we examine and elaborate an alternative perspective – that cognitive behavior is supported not by dedicated modules, but by a highly distributed and interactive cortical network, whose organization is strongly shaped and modified by experience. On this view, the functional specialization of brain regions is graded, rather than absolute, and reflects the consequences of a set of general principles and constraints on neural computation that operate throughout cortex. In support of this view, we review evidence from behavioral and imaging studies of normal and brain-damaged individuals, from developmental investigations of face and word acquisition, and from detailed computational modeling. This evidence supports three specific principles as they apply to face and word processing: i) distributed representation and knowledge; ii) representational cooperation and competition; and iii) topography, proximity, and hemispheric organization. The integrated application of these principles reveals the commonalities in cortical organization and behavior in these two seemingly unrelated domains and accounts for a wide range of empirical findings, including the partial co-mingling of face and word processing, the association between the acquisition of word and face recognition skills over the course of development, and their related neural mechanisms.

Visual word and face recognition: underlying neural mechanisms

Several different theoretical perspectives have been offered to explain the manner by which biological structures, such as the ventral visual cortex, come to be functionally optimized in the service of visual pattern recognition. The

---

Glossary

Developmental dyslexia: refers to a learning disability that impairs a person's fluency or comprehension accuracy in being able to learn to read. The disorder is usually not attributable to a frank brain lesion and is evident even when the individual has had ample opportunity and instruction to acquire reading.

Diffusion tensor imaging: a magnetic resonance imaging-based technique that allows the mapping of the diffusion process of molecules, mainly water, in biological tissues (primarily white matter for the current purposes) *in vivo* and non-invasively.

Fusiform face area (FFA): a region of the inferior ventral cortex that shows substantial selectivity for faces compared with other visual classes.

Prosopagnosia: (Greek: 'prosopon' = 'face', 'agnosia' = 'not knowing') a disorder of face perception, in which the ability to recognize and perhaps even discriminate between faces is impaired, but sensory vision and intellectual function remain unaffected. Prosopagnosia can be acquired through brain damage or can be lifelong (presumably congenital) in nature and evident even in the absence of a frank lesion.

Pure alexia: a neuropsychological disorder in which a lesion to the left occipitotemporal cortex (usually in the vicinity of the VWFA) results in laborious sequential decoding of letters in a string, resulting in slow letter-by-letter reading.

Univariate versus multivariate analyses: in univariate approaches to the analysis of functional imaging data, each voxel is treated independently and, typically, within that voxel, the BOLD signal derived from one condition is compared with that derived from a second condition. Multivariate techniques (for example, multi-voxel pattern analysis or MVPA) take into account the activity of a larger number of voxels and can therefore have higher informational value and sensitivity than the univariate approach.

Visual word form area (VWFA): a region of the inferior ventral cortex that shows substantial selectivity for words compared with other visual classes.

first perspective proposes that there are distinct cortical modules or regions, which mediate behavioral processes, such as face, word, or object recognition, in a domain-specific manner [1,2] and are perhaps even genetically determined [3,4]. Consistent with this approach are the findings that different areas in ventral visual cortex respond selectively to particular categories of visual stimuli: for example, as evident from many functional MRI (fMRI) studies and as depicted in Figure 1, the FFA is selectively activated in response to faces [5,6], the parahippocampal place area (PPA) to scenes [7,8], the extrastriate body area (EBA) and fusiform body area to human bodies and body parts [9,10], the lateral occipital complex (LOC) to objects, and the VWFA to words [11–13]. Indeed, in each of these regions, the blood-oxygen-level-dependent (BOLD) response for the preferred visual category is approximately twice as strong as that for the non-preferred category. Moreover, these domain-selective responses are evident in most individuals and these patterns of selectivity are observed across many different studies conducted by many different investigators using many different paradigms. Finally, neuropsychological investigations have provided support for this apparent specificity of neural and behavioral function: a selective lesion to the right fusiform gyrus in the vicinity of the FFA or to the left occipitotemporal region in the vicinity of the VWFA results in a severe impairment in face ('prosopagnosia') or word ('pure alexia') recognition, respectively [14].

A second perspective (the 'one-to-many' view) acknowledges the apparent selectivity of circumscribed neural areas for mediating certain visual classes, but argues that this selectivity need not implicate modules that are domain-specific *per se*. On this account, there exists a one-to-many arrangement, with a single region representing

multiple object types, albeit perhaps with differing levels of specificity. Indeed, many recent investigations attest to the multiplexing of ventral cortical regions. For example, the so-called VWFA is not only highly responsive to orthographic forms, but also to objects [15], Braille reading, and even auditory soundscapes [16]. Moreover, the so-called FFA is highly responsive to multiple stimuli including Greebles, objects, scenes, musical notation, cars, and birds [17–19], especially as experience and the pressure for precise exemplar discrimination is increased [20].

Neuropsychological evidence is also relevant here, in that individuals with selective lesions, when examined closely, appear to be impaired not only in the expected domain (for example, faces after a lesion to right FFA or words after a lesion to left VWFA), but in the recognition of other stimuli, too. Thus, individuals with prosopagnosia are not only impaired at face recognition, but are also impaired at recognizing other visually similar items as well, such as Greebles [21] and words [22], although the impairment is not as severe as it is for face recognition. In a similar fashion, individuals with pure alexia after lesions to the VWFA region are not only impaired at word recognition, but also at recognizing numbers [23], objects (especially as visual complexity increases [24]) and even faces [22], although, again, the impairment for these other classes is not as severe as the dyslexia itself.

A third perspective, which we will call the 'many-to-many' view and which is the focus of this opinion piece, is that there exists a many-to-many arrangement between the brain and behavior. This view incorporates both the domain-specific and the one-to-many perspectives, but goes beyond both of them. This account recognizes that there are individual regions that are optimized for a particular type of representation and also acknowledges that these regions generally represent more than one stimulus type. The additional assumption here is that many regions are necessarily engaged in the representation of multiple visual stimulus classes. Thus, this multiplicity of regions forms distributed but integrated large-scale circuits. Critically, this perspective implies that a single region alone does not suffice for normal behavior and that these extended, multi-regional neural circuits are necessary to ensure accurate and rapid visual recognition [25,26]. In other words, functional specialization is not simply an intrinsic property of individual regions that compute specific representations and/or computations in isolation, but, rather, is an emergent property of the interactions between a set of spatially distributed nodes and their functional and structural connections (see also [27]).

To illustrate this many-to-many perspective, we first consider the domains of faces and words. These two stimulus classes are especially apt candidates, because domain-specific neural correlates for face recognition and for word recognition have been repeatedly proposed. Moreover, these two classes are, intuitively, diametrically opposed, differing obviously in their overt geometry and image statistics. Additionally, faces and words diverge substantially in their acquisition: face recognition develops incidentally, whereas, for most individuals, word recognition is acquired through specific instruction in a more formal schooling environment. Finally, the evolutionary status
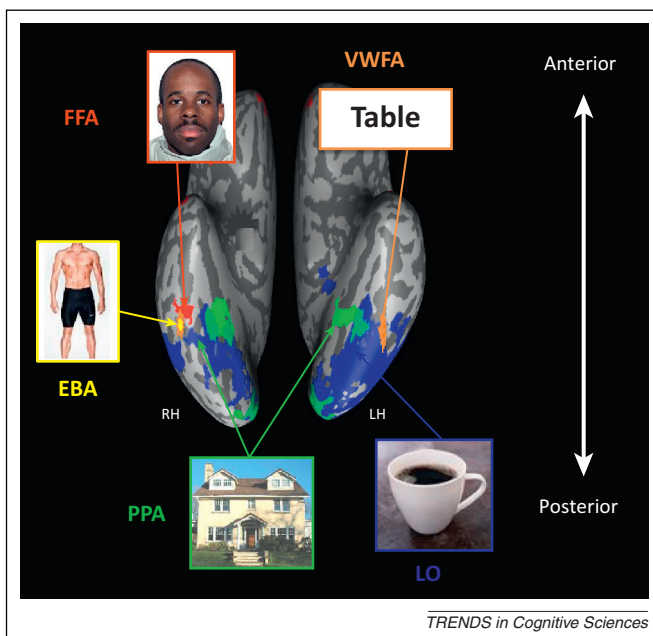

Figure 1. Ventral stream category-specific topography depicting domain-specific regions on a single representative inflated brain. As there is no single experiment that has examined the cortical activation associated with all these visual classes, this figure is a composite of the results of many different experiments and is partially a cartoon of the domain-specific activation of ventral cortex.

of words and faces is fundamentally different: reading is a relatively recent invention, introduced approximately 5,400 years ago [28] and, until roughly 150 years ago, its use was limited to a minority of the human population before basic education for the mass population was introduced (at least in the Western 'developed' nations). This relatively brief cultural time course is obviously not the case for face recognition. Taken together, these factors would seem to lead to different, specialized face and word modules – yet, as we review below, this is apparently not the case.

A distributed circuit mediates face recognition
Functional anatomy and connectivity

As noted above, much of the emphasis in the field of face recognition has been on the FFA as the pre-eminent neural correlate of face recognition. The particular focus on this cortical area might be a product of the widely-used methodological approach in which, using univariate analyses to contrast activation in response to two stimulus types (e.g., faces versus houses), a threshold is established for differentiating the activation between the two visual classes. Additionally, methodological simplifications including spatial smoothing, lower-resolution fMRI and fMRI artifacts might contribute to the derivation of a single circumscribed region that shows selectivity, whereas more nuanced approaches, including multivariate approaches and electrocorticography in humans [29], uncover a more complex picture [20,30,31].

There is, then, a growing consensus that, in addition to the FFA [32], multiple cortical regions evince face-selectivity [33–35], including the occipital face area (OFA), the lateral occipital sulcus (LOS) [36,37], and the posterior superior temporal sulcus [38,39]. In addition to this 'core' network, faces selectively activate extended regions, such as the anterior temporal lobe [40–42] (see also [43,44]), the amygdala [35], and even regions of inferior frontal and orbitofrontal cortex [34,45,46] (Figure 2). Consistent with the multiplicity of face-selective regions in humans, recent fMRI studies with non-human primates have also uncovered multiple, disparate face-selective temporal and frontal patches, many of which are functionally co-activated and exhibit correlated activity [47,48]. This substantial commonality in the organization of these face-selective patches in humans and non-human primates indicates wide-scale homology in cortical topographies [41,49].

In particular, the application of novel analytic techniques to functional imaging data has been of help in uncovering the relative contributions of the multiple face-selective regions. For example, in one recent study designed to investigate the neural code of facial identity perception [42], dynamic multivariate mapping employing information-based brain mapping and dynamic discrimination analysis was used to locate spatiotemporal patterns that support face classification at the individual level (for example, using a searchlight and a multivariate classifier to uncover those neural regions that demonstrate above chance discrimination of the identity of faces of specific individuals, independent of variation in emotional expression). These analyses revealed a network of bilateral fusiform and anterior temporal areas that carry information about facial identity.
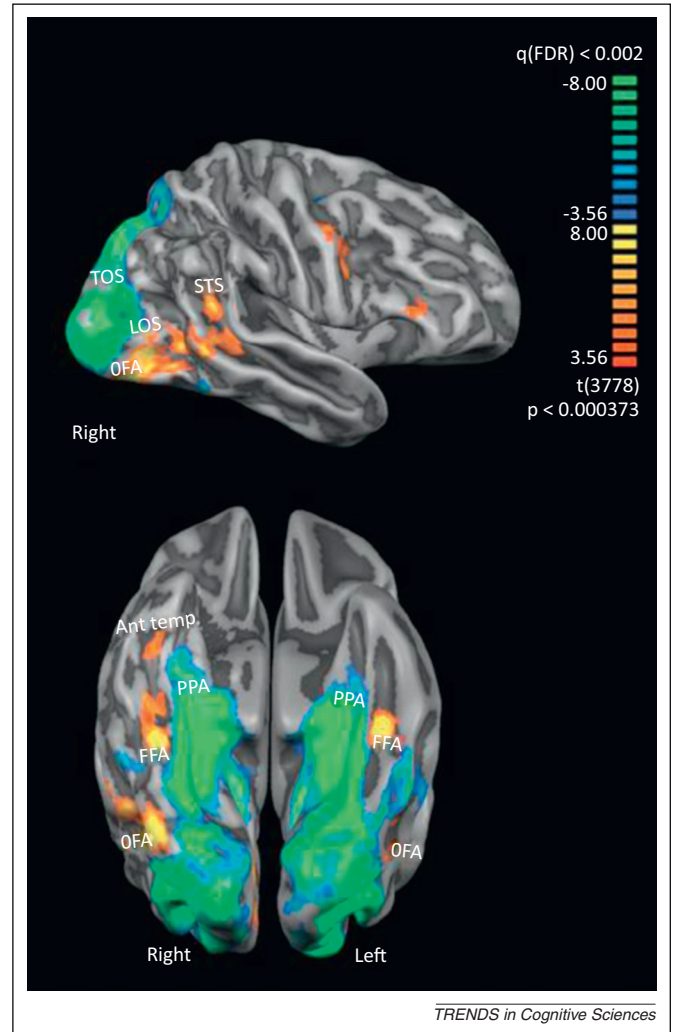

Figure 2. Activation maps, obtained by the contrast all faces > buildings (red to yellow colors), as overlaid on a group-averaged folded cortical mesh. The map is shown in a lateral view (top row) and a ventral view (bottom row). Note the large number of face-selective patches in the core face network, which includes bilateral OFA, LOS, FFA, and pSTS, as well as in the extended network, which includes the anterior temporal cortex (Ant. temp) in the right hemisphere and some patches in frontal cortex. Note that in this projection it is not possible to see all the regions in the distributed face network (for example, the amygdala). Also shown is the building-selective activation obtained from the contrast buildings > all faces (blue to green colors) in the PPA and TOS. Adapted, with permission, from [59].

Moreover, diagnostic information about the individual facial identities was distributed evenly among the regions of the mapped network. Pairwise correlations between the regions and mutual information analyses revealed that, perhaps unsurprisingly, an anterior region of the right fusiform gyrus plays a central (hub-like) role within the information network that mediates face individuation, but that the other regions play key roles, too.

Consistent with the many-to-many account, a single activated region is unlikely to suffice for normal face recognition. It is the case that a lesion to a node of the core network gives rise to prosopagnosia: patients with lesions to the vicinity of the FFA evince prosopagnosia and this is also true after a lesion to the OFA [50] or anterior temporal lobe [51–53]. Evidence for the need for an integrated face circuit comes from the study of prosopagnosic individuals in whom the key nodes of the face network

function normally, but the connectivity between these nodes and other more remote parts of the face network is compromised. These individuals, who have normal cognition and normal vision, suffer from 'congenital prosopagnosia' (CP), a lifelong impairment in face recognition. CP occurs in a small percentage of the population (approximately 2–2.5%) [54,55] and has a hereditary component [56,57]. Of particular interest, activation of the posterior or 'core' face regions, including FFA, LOF, OFA, and STS, is apparently normal in CP individuals and this is true in magnitude and extent of selectivity, adaptation profile, number of voxels, and site of activation (as determined by peak of coordinates) [43,58,59] (although, for some diverging results, see [60–62]). Particularly surprising and counterintuitive is that this normal pattern of activation is demonstrated at the very same time that the individuals perform abnormally on tasks that require face recognition [58]. Relatedly, in a recent magnetoencephalography study, the face-selective M170 responses within the right lateral occipital and right fusiform regions did not differ in magnitude between individuals with CP and controls [63]. The normal activation profile in posterior regions, however, as shown by Avidan and colleagues, was not accompanied by normal activation in more 'extended' regions, leading to the hypothesis that the posterior cortical regions might well be computing normally, but that the impairment in face recognition might arise from a failure to propagate the signals to more anterior regions [43,58,59].

Indeed, evidence for this disconnection hypothesis came initially from a structural imaging study, which showed, through anatomical decomposition of the volume of the temporal lobe, that the anterior temporal region was smaller in CP individuals than in control participants [64]. The advent of diffusion tensor imaging (DTI) enabled a clearer characterization of the white matter tracts, which interconnect the more posterior and core regions with the more anterior extended system, and has revealed that the inferior-longitudinal fasciculus (ILF; projects to the anterior temporal lobe) and the inferior fronto-occipital fasciculus (IFOF, which has same posterior trajectory and then bends past the uncinate fasciculus and projects to the frontal lobe) were both microstructurally and macrostructurally reduced in CP individuals, relative to control subjects. Moreover, there was a statistically significant correlation between the reduction of the ILF in the right hemisphere and the severity of prososopagnosia in the CP group. No other tracts sampled revealed a difference between the CP and control group.

As might be expected from a disconnection view, individuals with CP revealed reduced functional connectivity between the posterior regions and the anterior temporal lobe, the terminal point of the ILF. The dissociation within the network was rather circumscribed, given that the functional connectivity between the posterior regions and the amygdala (also part of the extended network) was normal [59], consistent with the findings that individuals with prosopagnosia can evaluate social characteristics normally, including judgments of aggression, attractiveness, confidence, intelligence, sociability, trustworthiness [65], and that their recognition of facial expression appears

to be well preserved (although this may not be true in all cases). This reduction in posterior–anterior temporal lobe connectivity was not only evident under task-related conditions, but also under resting state conditions, consistent with the notion that regions of networks established in the service of a particular function are correlated even when not under task demand [59].

Compatible with the hypothesis that a widespread circuit is required for face recognition, a reduction in macroand microstructural integrity of the ILF and IFOF ventral tracts, associated with an increase in aging (which presumably results from a reduction of myelin), leads to a reduction in competence in face recognition [66] (see also [67,68]). Finally, disproportionately large age-related changes in the volume, fractional anisotropy (FA), and mean and radial, but not axial, diffusivities of the ILF, which reflect increasing myelination, have been observed in individuals aged 6–23 years [69]. Of interest, these structural changes are tightly and specifically linked with the increasing size of the FFA. Taken together, the solidifying of the white matter tracts over the course of development, their reduction with aging, and their compromise in CP, all point to a potential mechanism in which the complex face circuit is structurally instantiated and, over the course of experience and maturation, becomes organized and optimized in human cortex or subsequently degenerates and is compromised.

Computational properties
Although much remains to be determined regarding the computational properties of the network, to the extent that there are temporal signatures of information flow, it is consistent with the idea of both feedforward and feedback projections. For example, in intracranial event-related potential (ERP) recordings performed in the late 1990s, multiple peaks of face-selective activation were noted in the lateral temporal cortex of humans, including peaks in ventral temporal cortex at approximately N170-200, at P290, and at N700, and at AP350 (these are all ERPs that reflect a response to a visual stimulus, with N and P indexing the negative or positive deflection of the electrophysiological response and the three-digit number indicating the approximate time post-stimulus onset at which the deflection peaks; AP350 refers to a positive deflection at approximately 350 ms), with this last component in the vicinity of the anterior temporal cortex. Indeed, during the course of exploring these electrophysiological potentials, Puce *et al.* [70] suggested a forward path of information flow from posterior potentials to the anterior temporal cortex, with the last peak indicative of the feedback to posterior regions. A similar temporal dynamic has been observed in single-unit recordings in macaque temporal cortex [71]; more global information – categorizing stimuli as monkey faces, human faces, or shapes – was conveyed in the earliest part of the neurophysiological responses, followed, 51 ms later, by fine information about identity or expression.

A distributed circuit mediates word recognition
As with face recognition, much progress has been made in going beyond a circumscribed, modular approach and in

uncovering the neural circuit that subserves the representation of orthographic information. For example, there has been growing recognition that a large swath of posterior-to-anterior left ventral cortex computes progressively more complex orthographic representations, from letters through bigrams to words [72–74]. In addition to this pathway, it appears that an alternative pathway, from occipital cortex to left motor and premotor regions, via activity in a central part of the left superior temporal gyrus, may also be engaged in the recognition of short, familiar words and may be disproportionately engaged in individuals in whom the left ventral VWFA is weak [75].

Although there appear to be several regions, rather than just one, that support word recognition, the focus of investigation has been largely restricted to the left hemisphere. This is somewhat surprising, because close scrutiny of existing studies reveals bilateral BOLD activation for words compared with stimuli such as objects (albeit with greater magnitude in the left hemisphere) (e.g., [76]). More recent pattern-based analyses and multivariate mapping of functional magnetic resonance imaging data, similar to that adopted for identifying the face circuit as noted above, have, in fact, revealed fairly extensive sensitivity in the ventral cortex for letter strings and this includes regions in the right hemisphere, as well [77]. Additionally, in this same study, multivariate analyses uncovered a number of cortical areas, including the portions of the right ventral lobe, that showed above chance discrimination of different pseudowords, irrespective of differences in font.

Analogous to the connectivity that is observed among the diverse nodes of the face network, several studies have mapped out the detailed white matter connectivity between different regions engaged in word recognition [78] and recent investigations have focused on correlations between white matter properties and reading skills in adults and children [79]. Consistent with the idea that word recognition involves the synthesis of bottom-up sensory input and top-down predictions that are generated automatically from prior experience [15], VWFA has been shown to be linked to the occipital cortex through the ILF and to perisylvian language areas (supramarginal gyrus) through the arcuate fasciculus. Moreover, in a patient who developed pure alexia following a small surgical lesion in the vicinity of the VWFA, progressive and selective degeneration of the ILF ensued, whereas the VWFA remained anatomically intact [80]. This disconnection is analogous to the dissociation of posterior from more anterior regions of ventral cortex in individuals with CP.

Finally, just as in cases of congenital prosopagnosia, a deficit in white matter circuitry has been proposed to account for developmental dyslexia [81]. Several investigations using DTI have reported fractional anisotropy differences between individuals with and without dyslexia (e.g., [78]), with notable changes in the bundle of perisylvian white matter referred to as the superior longitudinal fasciculus [82]. Although most of these studies focused their efforts on the changes in the perisylvian language network, some studies show that the white matter perturbations extend beyond this network. For example, adults showed a greater overall anterior-to-posterior orientation of white matter in control participants compared to those with dyslexia; and children with dyslexia showed differences in mean FA values for the IFOF and for the ILF, in addition to differences in the posterior limb of the internal capsule [83].

Face and word recognition: overlapping circuits

Thus far, we have described widespread cortical circuits that are functionally engaged in face or in word recognition. We have reported similar architectural arrangements, in which multiple, diverse regions are engaged in differentiating individual faces or individual letter strings. We also noted that white matter tracts that propagate signals between these regions are critical for normal perception and that dissociation between regions that are themselves intact can give rise to disorders in face or word recognition. Thus far, however, we have considered the circuits for each of the visual classes to be independent.

Determining whether the neural circuits for face and word recognition overlap or are independent has been difficult, because there have been rather few investigations that have simultaneously explored, within-subjects, the neural correlates of face and word recognition. As mentioned above, one recent study using multi-voxel pattern analysis (and spatiotemporal searchlight) showed substantial interdigitation of those voxels that are diagnostic for individual pseudoword and individual face identification [77]. That abutting and even overlapping voxels perform above-chance discrimination of both individual face patterns and individual pseudowords indicates that both stimulus types exploit common neural resources within the ventral cortical network.

This overlap in activation revealed by this simultaneous multivariate investigation is compatible with many fMRI and ERP studies that show bilateral activation for words and for faces, typically with greater activation for words on the left and faces on the right side of the brain [5,11,42,76,84–86], although the multivariate analyses go beyond this to show that some of the very same voxels may be dual-tasking. Importantly, the presence of voxels diagnostic for faces in the left hemisphere and for words in the right hemisphere is not epiphenomenal: prosopagnosia has been reported in a right-hander subsequent to a left hemisphere lesion [87] and pure alexia has been reported in a right-hander after a unilateral right occipitotemporal lesion [88]. Moreover, prosopagnosia is generally more severe following a bilateral than unilateral right lesion. It is perhaps unsurprising, then, that a lesion to the left ventral area in the vicinity of the VWFA results not only in pure alexia, but also in a mild version of prosopagnosia; and that a lesion to the right ventral area in the vicinity of the FFA results not only in prosopagnosia, but also in a mild form of pure alexia [22], given the bilateral representation, albeit with asymmetric weighting, in the two hemispheres.

Common principles: face and word recognition

The empirical findings reviewed offer support for a perspective in which many brain regions subserve the recognition of faces and many brain regions subserve the recognition of words. Moreover, there is overlap between these distributed circuits and there is some overlap or joint
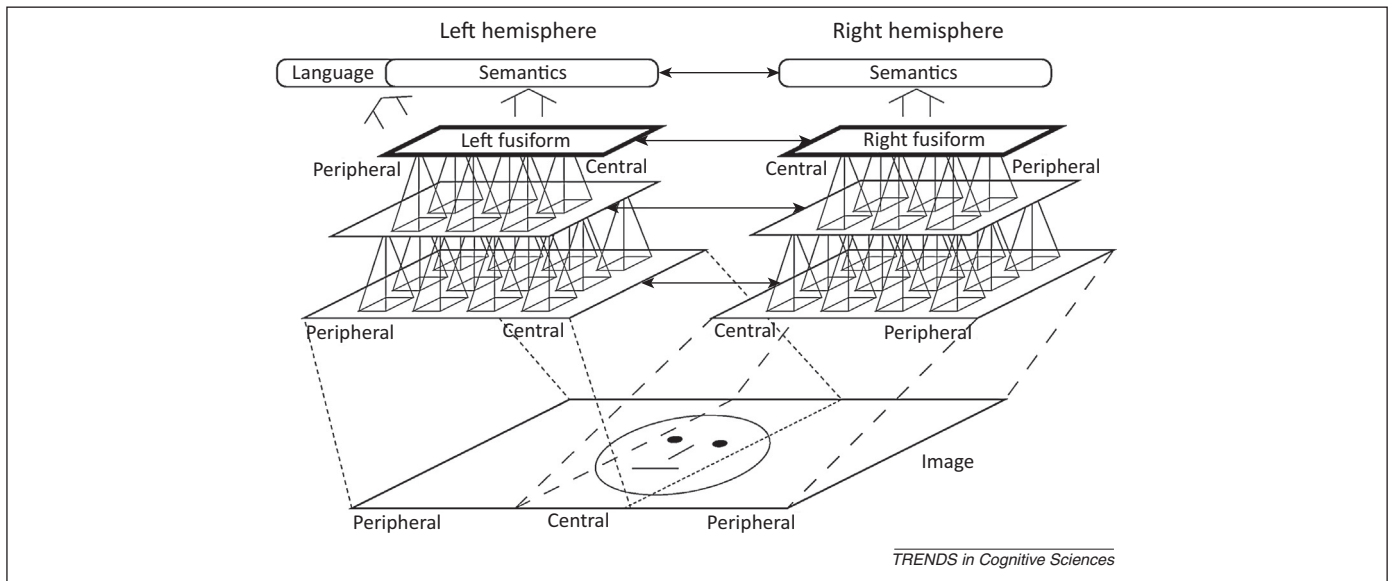
Figure 3. The architecture of a model for simulating the hemispheric organization of visual recognition. The model maps each visual field onto the contralateral hemisphere. The model has two sets of hierarchically organized layers that correspond to the two hemispheres. The two input layers (shown above the 'image' layer) encode visual information from the contralateral visual field in polar coordinates, using a Gaussian blurring function that is very narrow for central visual information, but increases quickly in moving towards the periphery, reflecting the sharp decrease in visual acuity away from the fovea. Connectivity within the retinotopic processing hierarchy is highly local, restricted largely to neighboring units within a layer and to small neighborhoods around units that code the same location in adjacent layers and in the other hemisphere. This retinotopic hierarchy allows the model to develop complex features at the highest level that are useful for object recognition and that can be derived from input across the entire visual field. The next layers in the model are non-retinotopic, but have a spatial extent that spans from central to peripheral visual information in each hemisphere. These layers correspond to fusiform cortex, containing the FFA and VWFA. They receive topographically constrained inputs from the highest retinotopic region, from each other, and from homologous units in the other hemisphere. They are also interconnected with a final set of layers that code the output of the model, primarily in terms of information about the identity and meaning of stimuli, but also including language-related outputs in the left hemisphere.

representation of multiple visual classes in several regions. Plaut and Behrmann [89] have articulated a theory, supported by explicit computational simulation, for how such an organization may result from learning within a neural architecture that is subject to a set of general principles and constraints on neural computation. At its core, this theory espouses interplay between cooperation and competition among representations. The theory assumes that representations are hierarchically organized, such that the representation of information at each level cooperates with (i.e., mutually activates and strengthens) the representations of consistent information at lower and higher levels (see Figure 3 for a depiction of such a network). Within a given level, the representations of different (inconsistent) information compete with each other, both to become active and to become stronger through learning (e.g., by recruiting more neurons to encode them). Critically, representational cooperation and competition depend on available connectivity, which is strongly constrained to minimize total white-matter volume (by minimizing axon length). In particular, connectivity within each hemisphere is highly local and often topographically organized, whereas interhemispheric connectivity is largely restricted to homologous areas. As a result, as internal representations become organized over the course of development, there is a strong pressure to locate representations spatially close to the information that they need to cooperate with and to keep competing representations spatially distant.

To be clear, these principles of representational learning and organization are not intended to be novel; in fact, we take them to be largely non-controversial. Moreover, they are not domain-specific; we assume that they constrain and

influence representations throughout neocortex. Even so, they have to be instantiated in detail in specific domains in order to fully understand their consequences.

As applied to faces and words, the theory has interesting and otherwise unexpected implications concerning the co-mingling of these two seemingly unrelated domains. Specifically, because both word and face recognition are highly overlearned and place extensive demands on high-acuity vision, they each need to cooperate with (i.e., be connected to and, hence, be adjacent to) representations of central visual information. Thus, in both hemispheres, words and faces compete for representational space in the specific cortical subarea adjacent to regions of retinotopic cortex that encode information from central vision [90], notably the VWFA and the FFA. Orthographic representations are further constrained to be as close as possible to language-related (especially phonological) information, which is left-lateralized in most individuals. As a result, words are co-localized in the LH and gradually come to rely more strongly – although not exclusively – on the left ventral region (VWFA) in order to mediate between early vision and language, integrating top-down predictions with bottom-up visual inputs [12,15]. Because of the competition between faces and words, face representations subsequently become more strongly – although, again, not entirely – lateralized to the right ventral region (FFA). Thus, the theory makes specific predictions concerning the co-mingling of face and word representations within and across hemispheres, over the course of development, in the brain organization of mature, neurologically intact individuals, and in patterns of impaired performance following brain damage.

Box 1. Prolonged developmental trajectory for face and word processing

Given the difference in the ways in which face and word recognition are acquired, with face representations honed incidentally, but word processing usually acquired by direct instruction, it is surprising that both stimulus types evince a similarly protracted developmental trajectory. Adult levels of identity face-matching performance are not yet evident in 10-year-olds when the faces differ in the spacing between the features [95] and face perception continues to show large improvements in recognition until approximately 12 years of age, in contrast to recognition of houses. In fact, peak performance with faces may not be evident until approximately 30 years of age [96]. Consistent with the slow maturation, selective activation of the FFA for faces is three times smaller in children than in adults [97] and emerges gradually through childhood and adolescence [98]. Although some face selectivity may be apparent in the right, but not left fusiform gyrus as early as 4–5 years of age (faces vs. shoes; [94]), the laterality pattern is still far from adult-like in 5- to 8-year-olds [99] and is not stable until early adolescence (12–14 years; [100]).

Selective activation of the VWFA is also not evident in young children and, as with face recognition, an adult pattern of VWFA lateralization emerges at or around 10 years of age [101]. This maturational pattern is also evident in ERP studies that show that the differential neural response to words emerges roughly after a year and a half of reading instruction [102] and, although lateralization is observed between 7 and 10 years of age, it is still not apparent to the same degree as in the adult pattern.

As we propose in the text, despite being phylogenetically much older, the emerging lateralization of face processing in later childhood arises as a consequence of earlier word lateralization (which, itself, is driven by left-lateralized language processes). Evidence favoring this view comes from the empirical findings by Dundas et al. [91], which show that adults, adolescents, and even young children with some reading experience show lateralized superiority for word recognition, but that only adults show lateralized superiority for faces.

Box 2. Outstanding questions

- Do the neural networks for face and word recognition overlap in the same way for left-handers (especially those with right-hemisphere language dominance) as they do for right-handed individuals?
- Do the neural networks for face and word recognition overlap in the same way for readers of a non-alphabetic script (where right hemisphere involvement in decoding a logographic script is greater) as they do for English readers?
- Do the neural networks for deaf individuals differ from those of hearing individuals not only for word reading, but for face recognition, as well?
- Do the neural networks for dyslexic individuals differ from those of hearing individuals, not only for word reading but also for face recognition?
- Despite the apparent co-localization of function, it remains possible that the actual computations engaged for face and word recognition may differ. Is this so? And, if yes, in what way? What is the nature of these underlying computations?
- What is the differential contribution of the multiplicity of nodes of the distributed network? Although they likely contribute somewhat different information (and hence are not redundant), the nature of the informational contribution remains unknown.

For example, although the system for face processing may be undergoing refinement with development, the hypothesis is that the pressure for lateralization is triggered by competition from word representations in the left hemisphere. This account makes the specific prediction that lateralization only emerges once word recognition is acquired and that competition for resources results in the subsequent shift of face representations into the right hemisphere (see also Box 1).

A recent empirical study provides support for these predictions [91]. This study examined the hemispheric superiority for faces and words in children (aged 7–9 years), young adolescents (aged 11–13 years), and adults in a half-field discrimination task. Whereas adults showed the expected left and right visual field superiority for face and word discrimination, respectively, the young adolescents demonstrated only the right field superiority for words and no field superiority for faces. Although the children's overall accuracy was lower than that of the older groups, like the young adolescents, they exhibited right visual field superiority for words, but no visual field superiority for faces. Interestingly, the emergence of face lateralization correlated with reading competence, measured on an independent standardized test, after partialing out age, quantitative reasoning scores, and face discrimination accuracy. Taken together, these findings suggest that the hemispheric organization of face and word recognition do not develop independently and that word lateralization, which emerges earlier, may drive later face lateralization.

A related prediction of the account is that individuals who have less strongly left-lateralized word representations should also have less strongly right-lateralized face representations. Two populations are relevant here and are the subject of ongoing research. First, many left-handed individuals have a more bilateral language organization, which would be expected to lead to more bilateral orthographic word representations. Second, non-alphabetic orthographic scripts, such as Chinese, are less closely tied to phonology and hence have a more bilateral organization. As a result, we predict that both left-handed individuals and native readers of Chinese should have less strongly right-lateralized face processing (as compared to right-handed English readers) and that this will have an impact on the organization of face and word processing (for relevant work, see [92]; Box 2).

The non-independence of the neural substrates for face and word recognition has been uncovered in other studies, as well. For example, adults with no formal education in reading have heightened left hemisphere activation to faces, compared with literate controls, and formal instruction in reading subsequently decreases the left fusiform activation to faces [93]. Similarly, young children show decreasing responses to faces in the left fusiform (VWFA) with increasing letter knowledge [94].

Finally, the aforementioned principles have important implications for interpreting the empirical evidence, reviewed earlier, for a many-to-many organization of the neural mechanisms that support the visual recognition of different classes of stimuli. In particular, evidence of overlap in face and word processing, even at the level of individual voxels [77], might nonetheless be interpreted as consistent with a one-to-one (or one-to-many) view, in which faces and words are processed by independent mechanisms that just happen to be interdigitated within the same regions of cortex. This view, however, violates the principle that unrelated information should be represented as separately as possible to avoid interference. If the principle is abandoned, the empirically observed

mixing of face and word processing remains unexplained. On the other hand, if, as we would espouse, this principle is embraced, it carries with it an important implication. If face and word processing were truly independent, but co-localized simply to be near central visual information, one would expect clustering of face-only and word-only regions at a scale much larger than individual voxels (much as left- and right-eye information gives rise to ocular dominance columns in primary visual cortex). If, however, face and word mechanisms are actually mixed within individual voxels, as the evidence suggests, it implies that processing within these two domains is not as unrelated as first thought and that the system has developed some types of representations and mechanisms that contribute to both domains.

Concluding remarks

Conventional wisdom holds that faces and words are independent domains of high-level vision, subserved by independent neural mechanisms located in opposite hemispheres. Recent research has reformulated this view, showing that regions that subserve each of these visual classes are not domain-specific – instead, multiple visual classes can be represented simultaneously within a cortical region. Here, we go beyond this and propose that visual classes of stimuli, such as faces and words, are mediated by distributed neural networks, in which multiple nodes are tightly coupled through structural connections. Furthermore, the very same networks may be engaged in both face and word recognition, albeit with differential weighting for faces in the right and words in the left hemisphere. Empirical data from functional and structural imaging studies, as well as from investigations of lesioned individuals, support this account. Findings that track the emergence of lateralization of regions that are optimized for face or word recognition are also consistent with the account.

In the absence of any apparent relationship between face and word processing (and perhaps the pressure for segregation, given their overt differences), a natural assumption is that the observed similarities in the emergence of their neural organization and performance are coincidental and that the two domains are independent. The apparent co-mingling of face and word mechanisms is unexpected from a domain-specific perspective, but follows naturally as a consequence of a many-to-many, interactive, learning-based account, in which neural processing for both faces and words is the result of an optimization procedure that embodies specific computational principles and constraints.

References

1 Kanwisher, N. (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11163–11170

2 McKone, E. and Robbins, R. (2011) Are faces special? In *The Oxford Handbook of Face Perception* (Calder, A.J. *et al.*, eds), pp. 147–176, Oxford University Press

3 Zhu, Q. *et al.* (2010) Heritability of the specific cognitive ability of face perception. *Curr. Biol.* 20, 137–142

4 Wilmer, J.B. *et al.* (2010) Human face recognition ability is specific and highly heritable. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5238–5241

5 Kanwisher, N. *et al.* (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311

6 Puce, A. *et al.* (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J. Neurophysiol.* 74, 1192–1199

7 Epstein, R.A. (2011) Cognitive neuroscience: scene layout from vision and touch. *Curr. Biol.* 21, R437–R438

8 Sewards, T.V. (2011) Neural structures and mechanisms involved in scene recognition: a review and interpretation. *Neuropsychologia* 49, 277–298

9 Peelen, M.V. and Downing, P.E. (2005) Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* 93, 603–608

10 Willems, R.M. *et al.* (2010) Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. *Cereb. Cortex* 20, 1719–1725

11 Price, C.J. and Mechelli, A. (2005) Reading and reading disturbance. *Curr. Opin. Neurobiol.* 15, 231–238

12 Devlin, J.T. *et al.* (2006) The role of the posterior fusiform gyrus in reading. *J. Cogn. Neurosci.* 18, 911–922

13 Dehaene, S. *et al.* (2005) The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341

14 Barton, J.J.S. (2011) Disorder of higher visual function. *Curr. Opin. Neurol.* 24, 1–5

15 Price, C.J. and Devlin, J.T. (2011) The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253

16 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

17 Grill-Spector, K. *et al.* (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. *Nat. Neurosci.* 9, 1177–1185

18 Epstein, R.A. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nature* 392, 598–601

19 Baker, C.I. *et al.* (2007) Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9087–9092

20 Williams McGugin, R. *et al.* (2012) High-resolution imaging of expertise reveals reliable object selectivity in the FFA related to perceptual performance. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17063–17068

21 Gauthier, I. *et al.* (1999) Can face recognition really be dissociated from object recognition? *J. Cogn. Neurosci.* 11, 349–370

22 Behrmann, M. and Plaut, D.C. (2012) Bilateral hemispheric representation of words and faces: evidence from word impairments in prosopagnosia and face impairments in pure alexia. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhs390

23 Starrfelt, R. and Behrmann, M. (2011) Number reading in pure alexia – a review. *Neuropsychologia* 49, 2283–2298

24 Behrmann, M. *et al.* (1998) Visual complexity in letter-by-letter reading: 'pure' alexia is not so pure. *Neuropsychologia* 36, 1115–1132

25 Catani, M. (2007) From hodology to function. *Brain* 130, 602–605

26 Mesulam, M.M. (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* 28, 597–613

27 Gschwind, M. *et al.* (2012) White-matter connectivity between face-responsive regions in the human brain. *Cereb. Cortex* 22, 1564–1576

28 Dehaene, S. and Cohen, L. (2007) Cultural recycling of cortical maps. *Neuron* 56, 384–398

29 Parvizi, J. *et al.* (2012) Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920

30 Weiner, K.S. and Grill-Spector, K. (2012) The improbable simplicity of the fusiform face area. *Trends Cogn. Sci.* 16, 251–254

31 Xu, G. *et al.* (2012) Similar spatial patterns of neural coding of category selectivity in FFA and VWFA under different attention conditions. *Neuropsychologia* 50, 862–868

32 Sergent, J. (1992) Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* 115, 5–36

33 Fairhall, S.L. and Ishai, A. (2007) Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex* 17, 2400–2406

34 Ishai, A. (2008) Let's face it: it's a cortical network. *Neuroimage* 40, 415–419

35 Gobbini, M.I. and Haxby, J.V. (2007) Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41

36 Gauthier, I. *et al.* (2000) Does visual subordinate-level categorisation engage the functionally defined fusiform face area? *Cogn. Neuropsychol.* 17, 143–164

37 Rossion, B. *et al.* (2003) A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain* 126, 2381–2395

38 Haxby, J.V. (2000) The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233

39 Puce, A. *et al.* (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J. Neurosci.* 18, 2188–2199

40 Kriegeskorte, N. *et al.* (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20600–20605

41 Rajimehr, R. *et al.* (2009) An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1995–2000

42 Nestor, A. *et al.* (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9998–10003

43 Avidan, G. and Behrmann, M. (2009) Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Curr. Biol.* 19, 1146–1150

44 Haxby, J.V. and Ida Gobbini, M. (2007) The perception of emotion and social cues in faces. *Neuropsychologia* 45, 1

45 Henson, R.N. *et al.* (2003) Electrophysiological and haemodynamic correlates of face perception, recognition and priming. *Cereb. Cortex* 13, 793–805

46 Collins, H.R. *et al.* (2012) Process and domain specificity in regions engaged for face processing: an fMRI study of perceptual differentiation. *J. Cogn. Neurosci.* http://dx.doi.org/10.1162/jocn_a_00273

47 Tsao, D.Y. *et al.* (2008) Patches of face-selective cortex in the macaque frontal lobe. *Nat. Neurosci.* 11, 877–879

48 Moeller, S. (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320, 1355–1359

49 Pinsk, M.A. *et al.* (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J. Neurophysiol.* 101, 2581–2600

50 Rossion, B. *et al.* (2003) The functionally defined right occipital and fusiform 'face areas' discriminate novel from visually familiar faces. *Neuroimage* 19, 877–883

51 Williams, M.A. *et al.* (2006) Abnormal configural face perception in a patient with right anterior temporal lobe atrophy. *Neurocase* 12, 286–291

52 Fox, C.J. *et al.* (2011) Perceptual and anatomic patterns of selective deficits in facial identity and expression processing. *Neuropsychologia* 49, 3188–3200

53 Barton, J.J.S. (2008) Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *J. Neuropsychol.* 2, 197–225

54 Bowles, D.C. *et al.* (2010) Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cogn. Neuropsychol.* 26, 423–435

55 Kennerknecht, I. *et al.* (2006) First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *Am. J. Med. Genet. A* 140, 1617–1622

56 Duchaine, B. (2007) Family resemblance: ten family members with prosopagnosia and within-class object agnosia. *Cogn. Neuropsychol.* 24, 419–430

57 Schmalzl, L. *et al.* (2008) Cognitive heterogeneity in genetically based prosopagnosia: a family study. *J. Neuropsychol.* 2, 99–117

58 Avidan, G. *et al.* (2005) Detailed exploration of face-related processing in congenital prosopagnosia: 2. Functional neuroimaging findings. *J. Cogn. Neurosci.* 17, 1150–1167

59 Avidan, G. *et al.* (2013) Selective dissociation between core and extended regions in the face processing network in congenital prosopagnosia. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bht007

60 Furl, N. *et al.* (2011) Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *J. Cogn. Neurosci.* 23, 1723–1740

61 Bentin, S. *et al.* (2007) Too many trees to see the forest: performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. *J. Cogn. Neurosci.* 19, 132–146

62 Minnebusch, D.A. *et al.* (2009) A bilateral occipitotemporal network mediates face perception. *Behav. Brain Res.* 198, 179–185

63 Rivolta, D. *et al.* (2012) Investigating the features of the m170 in congenital prosopagnosia. *Front. Hum. Neurosci.* 6, 45

64 Behrmann, M. *et al.* (2007) Structural imaging reveals anatomical alterations in inferotemporal cortex in congenital prosopagnosia. *Cereb. Cortex* 17, 2354–2363

65 Quadflieg, S. *et al.* (2012) Normal face-based judgements of social characteristics despite severely impaired holistic face processing. *Vis. Cogn.* http://dx.doi.org/10.1080/13506285.2012.707155

66 Thomas, C. *et al.* (2008) Reduction in white matter connectivity, revealed by diffusion tensor imaging, may account for age-related changes in face perception. *J. Cogn. Neurosci.* 20, 268–284

67 Grossi, D. *et al.* (2012) Structural connectivity in a single case of progressive prosopagnosia: The role of the right inferior longitudinal fasciculus. *Cortex* http://dx.doi.org/10.1016/j.cortex.2012.09.010

68 Migliaccio, R. *et al.* (2012) Brain networks in posterior cortical atrophy: a single case tractography study and literature review. *Cortex* 48, 1298–1309

69 Scherf, K.S. *et al.* Emerging structure-function relations in the developing face processing system. *Cereb. Cortex* (in press)

70 Puce, A. *et al.* (1999) Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458

71 Sugase, Y. *et al.* (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873

72 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

73 Szwed, M. *et al.* (2012) Towards a universal neurobiological architecture for learning to read. *Behav. Brain Sci.* 35, 46–47

74 Dehaene, S. and Cohen, L. (2011) The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262

75 Seghier, M.L. *et al.* (2012) Reading without the left ventral occipito-temporal cortex. *Neuropsychologia* 50, 3621–3635

76 Hasson, U. *et al.* (2002) Center-biased representation for characters in the human ventral visual stream. *Neuron* 34, 479–490

77 Nestor, A. *et al.* (2012) The neural basis of visual word form processing: a multivariate investigation. *Cereb. Cortex* http://dx.doi.org/10.1093/cercor/bhs158

78 Dougherty, R.F. *et al.* (2007) Temporal-callosal pathway diffusivity predicts phonological skills in children. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8556–8561

79 Ben-Shachar, M. (2007) White matter pathways in reading. *Curr. Opin. Neurobiol.* 17, 258–270

80 Epelbaum, S. *et al.* (2008) Pure alexia as a disconnection syndrome: new diffusion imaging evidence for an old concept. *Cortex* 44, 962–974

81 Rimrodt, S.L. *et al.* (2010) White matter microstructural differences linked to left perisylvian language network in children with dyslexia. *Cortex* 46, 739–749

82 Carter, J.C. *et al.* (2009) A dual DTI approach to analyzing white matter in children with dyslexia. *Psychiatry Res.* 172, 215–219

83 Rollins, N.K. *et al.* (2009) Simple developmental dyslexia in children: alterations in diffusion-tensor metrics of white matter tracts at 3 T. *Radiology* 251, 882–891

84 Kronbichler, M. *et al.* (2004) The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage* 21, 946–953

85 Puce, A. *et al.* (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16, 5205–5215

86 Sergent, J. *et al.* (1992) Functional neuroanatomy of face and object processing. *Brain* 115, 15–36

87 Anaki, D. *et al.* (2007) Associative (prosop)agnosia without (apparent) perceptual deficits: a case-study. *Neuropsychologia* 45, 1658–1671

88 Davous, P. and Boller, F. (1994) Transcortical alexia with agraphia following a right temporo-occipital hematoma in a right-handed patient. *Neuropsychologia* 32, 1263–1272

89 Plaut, D.C. and Behrmann, M. (2011) Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* 28, 251–275

90 Levy, I. *et al.* (2001) Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539

91 Dundas, E. *et al.* (2012) The joint development of hemispheric lateralization for words and face. *J. Exp. Psychol. Gen.* http://dx.doi.org/10.1037/a0029503

92 Li, S. *et al.* (2013) Neural competition as a developmental process: early hemispheric specialization for word processing delays specialization for face processing. *Neuropsychologia* 51, 950–959

93 Dehaene, S. *et al.* (2010) How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364

94 Cantlon, J.F. *et al.* (2011) Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cereb. Cortex* 21, 191–199

95 Mondloch, C.J. (2010) Discrimination of facial features by adults, 10-year-olds, and cataract-reversal patients. *Perception* 39, 184–194

96 Germine, L.T. *et al.* (2011) Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* 118, 201–210

97 Golarai, G. *et al.* (2007) Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nat. Neurosci.* 10, 512–522

98 Cohen Kadosh, K. *et al.* (2012) Differential face-network adaptation in children, adolescents and adults. *Neuroimage* 69, 11–20

99 Scherf, K.S. *et al.* (2007) Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev. Sci.* 10, F15–F30

100 Aylward, E.H. *et al.* (2005) Brain activation during face perception: evidence of a developmental change. *J. Cogn. Neurosci.* 17, 308–319

101 Turkeltaub, P.E. *et al.* (2008) Development of ventral stream representations for single letters. *Ann. N. Y. Acad. Sci.* 1145, 13–29

102 Maurer, U. *et al.* (2005) Emerging neurophysiological specialization for letter strings. *J. Cogn. Neurosci.* 17, 1532–1552