Title: Visuospatial coding as ubiquitous scaffolding for human cognition

Author Names: Iris I. A. Groen[1], Tessa M. Dekker[2], Tomas Knapen[3,4] & Edward H. Silson[5*]

1. Institute for Informatics, University of Amsterdam, Amsterdam, The Netherlands.
2. Institute of Ophthalmology, University College London (UCL), London, United Kingdom
3. Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
4. Spinoza Centre for NeuroImaging, Royal Dutch Academy of Sciences, Amsterdam, the Netherlands.
5. Department of Psychology, School of Philosophy, Psychology & Language Sciences, University of Edinburgh, Edinburgh, United Kingdom, EH8 9JZ.

*Correspondence: ed.silson@ed.ac.uk (E.H. Silson).
* All authors contributed equally

Keywords: visuospatial coding, retinotopic maps, visual field biases, category-selectivity, cognition, scaffolding, cortical development, computational modelling.

Acknowledgements: We thank Dr. Chris Baker, Dr. Marty Sereno & Dr. Caroline Robertson for insightful comments on an earlier version. We also thank Dr. Michael Bonner, Dr. Michael Arcaro, Dr. Jesse Gomez, Prof. Kalanit Grill-Spector, Dr. Ben De Haas, & Dr. Silvia Rigato for permission to adapt figures, and Martin Szinte for providing novel figures.

Abstract

For over 100 years we have known that the visual field is mapped onto the surface of visual cortex, imposing an inherently spatial reference frame on visual information processing. Recent studies highlight visuospatial coding not only throughout visual cortex, but also brain areas not typically considered visual. Such widespread access to visuospatial coding raises important questions about its role in wider cognitive functioning. Here, we summarize these recent developments and propose that visuospatial coding scaffolds human cognition by providing a reference frame through which neural computations interface with environmental statistics and task demands via perception-action loops.

What is visuospatial coding?

As humans we act and perceive in a mostly visual world. We are not passive observers but participate actively in the world to achieve goals, such as recognising a friend, reading a menu, or picking up a coffee cup [1]. For many of these goals, it is vital that the brain accurately interprets the physical properties of the world, and a critical property of vision is the spatial arrangement of elements in the visual field [2]. Given the importance of spatial arrangement, it is likely unsurprising that the visual system has evolved to encode the spatial location of visual elements - **visuospatial coding (see Glossary)**. Here, we consider visuospatial coding to manifest within the brain in two forms: retinotopic maps and visual field biases***.***

Retinotopic maps constitute the strictest form of visuospatial coding***.*** Adjacent cells in the retina project to adjacent neurons in the brain, creating a contiguous mapping between visual field positions and cortical locations [2]. Retinotopy was originally identified by researchers correlating the location of visual cortex lesions with corresponding visual field deficits [3]. These revolutionary observations highlighted that primary visual cortex (V1) not only represents the **contralateral visual field** but also that the cortical image is inverted (just as it is on the retina) and that the centre of the visual field (where we fixate and have the highest acuity) is represented by a disproportionately large expanse compared to the peripheral visual field – **cortical magnification (Figure 1A)**. Later invasive animal studies confirmed this organisation in V1 [4], and identified additional retinotopic maps surrounding V1 **(Figure 1A)**. However, it was functional magnetic resonance imaging (fMRI) that provided a non-invasive way to exploit this organisation in-vivo [5–8], leading to many of the discoveries covered in this review. Contemporary estimates suggest each hemisphere comprises at least 25 cortical retinotopic maps [2,9,10].

A second form of visuospatial coding manifests as coarse-scale **visual field biases** of receptive fields tuned to certain locations [8]. A visually-biased region preferentially responds to a specific visual field position (e.g. contralateral visual field), but might otherwise lack evidence for a clear retinotopic map(s). For example, prior fMRI work demonstrated coarse-scale positional sensitivity within an area of ventral temporal cortex important for processing written words (visual word form area [VWFA]), even without clear evidence of an underlying retinotopic map [11]. Extensions of this work employing **population receptive field** (pRF) modelling [8] **(see Box 1)** demonstrate that pRFs in this region tend to be small, foveally centred, and when averaged extend their coverage of the visual field further horizontally than vertically [12,13] - mirroring the horizontal extent of words during reading (e.g. in English). Thus, visual field biases provide a coarse-level fingerprint for visuospatial coding **(Figure 1B)**.

While the historical view has been that retinotopic maps were the purview of early visual areas and did not overlap more anterior **category-selective** regions of visual cortex **(Figure 1C)**, advancing methodologies have generated evidence for both forms of visuospatial coding throughout the visual system **(Figure 1C)**, and in regions not typically considered visual, such as those involved in coordinating movements, mind wandering and memory. Indeed, evidence for visuospatial coding in unexpected places in the brain emerged in the early 2000s [14,15], and has increased as mapping techniques improved **(see Box 1)**. Notably, evidence of visual field biases is often superseded by evidence of retinotopic maps through methodological advances such as increased signal-to-noise or spatial resolution. Widespread access to visuospatial coding not only an efficient framework for communication between functionally different systems, but also a means to align cognitive processes internally and externally via perception-action loops. Below, we summarize these findings and raise new questions about the role of visuospatial coding in wider cognitive functioning.

Ubiquity of visuospatial coding in the brain

The last years have seen a burgeoning literature for visuospatial coding in unexpected places **(Figure 2)**. For example, retinotopic maps have been reported in frontal [16] and parietal [14,15] areas associated with high-level cognitive processes such as attention [16], working memory [15], eye movements [17], and fluid intelligence [18]. Recent pRF modelling highlights distinct clusters of retinotopic maps in superior and inferior parts of the precentral sulcus [19], and wide-field retinotopic mapping has revealed retinotopically organised regions (Prostriata and V2A) at the junction between early visual cortex and medial parietal cortex [20,21]. Visuospatial coding is also present in the subcortex. Both the pulvinar and reticular nuclei of the thalamus have been found to contain distinct retinotopic maps [22,23], as have the superior colliculus and substantia nigra in the midbrain [23] **(Figure 2A, 2B)**. These regions are involved in top-down attention [24], working memory [25], and eye movements [26], indicating that high-level yet subcortical modes of processing are facilitated by visuospatial coding.

Even in the cerebellum, far outside the scope of traditional visual neuroscience experiments, pRF modelling has demonstrated multiple retinotopic maps distributed across separate lobules [27,28] **(Figure 2C)**. The distinct properties of these retinotopic clusters in terms of visual field preferences, cognitive activations, and their overlap with cerebellar somatotopic maps hint at their diverse functions. These range from guiding eye movements (in the oculomotor vermis [28]), to spatial attention and visual working memory (in lobule VIIb [27]), to visuomotor integration in service of manual dexterity and locomotion (in lobule VIIIb [28]).

Two recent findings suggest that visuospatial coding is even present within the hippocampus, a structure long associated with episodic memory, navigation and **world-centered spatial processing**. First, pRF modelling demonstrated a coarse-scale bias for the contralateral visual field within the hippocampus [29]. Second, mutli-task ultra-high resolution fMRI demonstrated detailed **topographic connectivity (see Box 1)** between V1 and the hippocampus [30] **(Figure 2D)**. Whether these signatures of visuospatial coding reflect retinotopic inputs into the hippocampus or retinotopically tuned neurons within the hippocampus itself remains unclear but may indicate an unanticipated role for visuospatial coding in hippocampal function.

Recent work has also revealed unexpected visually responsive, but negative pRFs in the **default mode network (DMN)** - a series of regions associated with memory and mind-wandering [31]. Unlike pRFs in traditional visual regions, which respond positively to visual stimulation within their receptive fields, DMN pRFs deactivate in response to visual stimulation at specific visual field locations. Collectively, DMN pRFs exhibit coarse-scale biases for the contralateral visual field with some hints of map-like structure across the cortical surface, particularly the angular gyrus **(Figure 2E)**. Although the role of these negative DMN pRFs and how they relate to positive pRFs in antecedent visual areas is unclear, their presence alone hints at a linking mechanism between visual perception and visual-memory guided thought such as mind-wandering.

Collectively these data demonstrate both forms of visuospatial coding within functionally disparate systems: retinotopic maps in the parieto-frontal cortices and cerebellum, and coarse-scale contralateral biases in inferotemporal cortex, the DMN, and hippocampus. This visuospatial coding ubiquity raises an intriguing new question: what are the computational advantages of this organisation? A mechanistic account for why the brain initially recapitulates the retina across its surface is that it minimizes biologically expensive long-range connections (the wiring optimization principle; [32]). Continuing this reference frame throughout visual cortex may be advantageous because it reduces the need for transforming information, allowing distinct aspects of a visual image to be computed in different maps while preserving spatial arrangements [2] (but see [33] for an alternative explanation for repeating retinotopic maps). Having visuospatial codes in visuomotor regions responsible for generating eye movements (e.g. frontal eye fields [19]) or reaching behaviours may also facilitate coordination between vision and movement (i.e. visuomotor integration). However, the increasingly clear prevalence of visuospatial coding in and beyond visual and visuomotor brain regions suggests that its computational role is much broader than previously thought and may reflect a general coding principle that subserves broader cognition.

Below, we review recent evidence for this wider role from the domains of higher-order perception, cognitive development and action-perception integration. We then present a framework for visuospatial coding as a major neural motif that scaffolds the alignment of information processing.

Visuospatial coding for higher-order perception

One of the most striking discoveries of visual cortex organisation is that it contains category-selective regions (e.g. faces, scenes or objects) [34]. When considering high-order visual functions, such as face-, scene- or object-recognition, a prominent notion is that such representations are decoupled from visual field position (e.g. a dog is a dog, no matter where in the visual field it is). But electrophysiological and fMRI data are challenging this notion, demonstrating that visuospatial coding is not only measurable within category-selective regions, but may play a crucial role in the emergence of such selectivity.

Early fMRI work of lasting impact linked visuospatial coding with high-order visual functions by demonstrating that face- and scene-selective regions overlapped representations of the **fovea** and **periphery**, respectively [35,36]. This large-scale fovea-periphery organisation within high-level visual cortex was later refined through delineation of multiple (often pairs of) retinotopic maps overlapping human object- (LO1/LO2) [37], colour- (VO1/VO2) [38], motion- (TO1/TO2) [39] and scene-selective (PHC1/PHC2) [40] regions, respectively **(see bottom Figure 1C)** (but also scene-selective areas (OTS1/OTS2) in non-human primates [41]). Further, both fMRI and neurostimulation studies link processing of different visual features to individual retinotopic maps within larger areas. For instance, TO1/TO2, subdividing motion-selective V5/MT, play different roles in processing translational and rotational motion [42], whereas LO1/LO2, subdividing object-selective LO play different roles in orientation and shape processing, respectively [37,43]. Not all category-selective regions contain pairs of retinotopic maps. For example, scene-selective occipital place area (OPA) overlaps at least five [44,45], whereas face-selective occipital face area (OFA) shows little overlap [43].

Evidence for the influence of visuospatial coding in high-order visual perception also comes from work examining pRF characteristics of face- and scene-selective regions. These studies, which confirm and extend the earlier fovea-periphery work [35,36], consistently highlight non-uniform distributions of pRFs within these regions. For instance, pRFs in face-selective regions tend to be small and foveally centred [13,46,47], whereas those in scene-selective regions are larger and oeripherally biased [48]. This bias for foveal and peripheral visual input, does not mean that face-selective regions lack peripheral representation, or that scene-selective regions lack foveal representation [48]. Rather, it reflects a higher density of foveal and peripheral receptive fields within both regions. Electrophysiological recordings in non-human primates also suggest neural populations in high-level visual regions code visual field position. For instance, an objects visual field position can be decoded from inferotemporal cortex (IT), a region thought to be critical for invariant object recognition [49,50]. In humans, responses are larger in dorsal (i.e. LO) and ventral (i.e. pFS) object-selective regions for objects in the lower and upper visual fields [51]. Likewise, scene-selective PPA on the ventral surface is biased for the upper visual field, whereas OPA on the lateral surface is biased for the lower visual field [48] **(Figure 3A).**

These data highlight the link between higher-order category representations and visuospatial coding. Recent work is exploring the role that visuospatial coding plays in the emergence of category-selectivity during development.

Visuospatial coding for cortical development

The retinotopic structure of visual cortex is established in-utero prior to visual experience [52–54]. Accordingly, recent studies have reported high consistency in retinotopic map positioning in the visual cortex from birth onwards [47,55,56]. In contrast to such early visuospatial organisation, category-selectivity appears to emerge more gradually: work in infant humans shows that responses for faces and scenes are present in stereotypical locations by months 4-6, but adult-like selectivity is not [57]. Further, face-selective responses in infant macaques emerge slowly (within ~1.5 years), whereas visuospatial coding evidenced through **resting-state functional connectivity** is present early on (~18 days) [58] **(Figure 3B)**. Interestingly, cortical locations showing face-selectivity were predictable via pre-existing foveal representations (i.e. in line with eccentricity biases [13,35,36,48]). Instead, macaques that grow up face-deprived show no selectivity for faces within these same foveally biased regions [59], suggesting that visual experience is required for selectivity to emerge.

Based on these findings, Arcaro and colleagues proposed that earlier-developing visuospatial biases in higher-order cortex act as a 'protomap' that anchor where category-selectivity will emerge based on local tuning to category-relevant viewing eccentricities, spatial frequencies, and curvature (which are largely co-localised in cortex) [58,60]. Further, empirical evidence for shape- and eccentricity-driven category-selectivity comes from work in adult macaques, where shape-selective patches for novel symbols emerged in stereotyped locations independent of expertise [61], and from a recent study in human expert pokemon-players who displayed pokemon-selective patches in higher-order cortex biased for small viewing eccentricities where these stimuli were experienced [62]. Both findings provide additional evidence for experience-dependent category-level representations linked to visuospatial coding.

However, visuospatial coding is not the only mechanism thought to underpin the emergence of category-selectivity [63]. For example, its role has been questioned based on findings that areas selective for faces, scenes, and objects emerge in similar locations in congenitally blind [64–67]. This shows that visual experience may not be *the* essential driver for category-selectivity in higher-order cortex, but does not necessarily mean that visuospatial coding plays no role: indeed, retinotopic organisation is retained in congenitally blind [68] and topographic maps across the brain share pre-determined developmental mechanisms and common organisational principles, such as expansion of cortex for high-acuity processing (i.e., for foveation and manipulation). It is conceivable that such inherent neural specialisation constrains where category-selectivity emerges, even when visual cortex takes on different functions in absence of visual input. In line with an important role for early brain-wide connectivity [69], regions that develop face- and scene-selectivity show higher within-category patterns of connectivity (i.e. face-face > face-scene) in infancy [69], and the position of the VWFA was found to be predictable on the basis of connectivity with anterior language areas [70,71]. These accounts do not invalidate a role for visuospatial coding in the emergence of category-selectivity, but rather suggest that it is one of several mechanisms at play [72].

While visuospatial coding may not be the only mechanism shaping category-selectivity, it appears to play a functional role in perception that continues to be refined throughout development. For instance, a recent study [47] showed that neuronal populations in face-selective cortex are well-tuned for covering a whole face at typical viewing distances in adults, but only select features in children **(Figure 3C)**. This developmental increase in foveal coverage also coincided with changes in viewing patterns for faces, suggesting a link between developmentally initiated changes in viewing behaviour and changes in pRF properties within high-order visual cortex. Further, pRFs in face-selective regions appear to shift their coverage when processing inverted (vs upright) faces, implying that visual processing strategy is linked with visuospatial sampling [73]. So, the development of adult-level object perception skill may in part reflect more efficient indexing of category-relevant features in visuospatial formats.

Collectively, these data showcase that while visuospatial coding might not be the only mechanism shaping cortical development, it is important to acknowledge both the early onset and pervasiveness of visuospatial coding throughout the brain. Such coding is not ‘thrown-away’ during development, instead how the brain uses it is continually refined.

Visuospatial coding for action-perception integration

The utility of visuospatial coding in broader cognition is also evident when considering how visual information may be used to guide actions. A key question when considering how visual information is used to guide motor actions is how to align visuospatial representations across reference frames (e.g. somatosensory representation of the body). Sereno and Huang found a somatosensory map of the face in parietal cortex, which also contained a map of visual space [74]. These overlapping maps were systematically aligned, such that tuning to visual and somatosensory directions co-localised in cortex. For example, lower visual field preferences coincided with tuning to touches on the lower part of the face. A similar alignment of multisensory sensitivities occurs in lateral visual cortex [75] where representations of visual space and face part positions overlap: regions tuned to the upper visual field are sensitive to upper face parts at fixation **(Figure 3D)** [75]. Cortical areas responsible for reaching and grasping also appear to benefit from visuospatial coding, with the superior parietal-occipital cortex showing a preference for hand-actions in the lower visual field - where graspable objects typically occur [76,77].

At first glance, the notion of widespread visuospatial coding seems suboptimal in a world in which we incessantly move our eyes, thereby shifting incoming visual information on the retina. In most studies visual field mapping is performed during fixation, which means that the retinotopic and so-called **spatiotopic**, or **world-centered**, frames of reference coincide. Yet, fMRI studies directly pitting retinotopy against spatiotopy overwhelmingly highlight the predominance of the retinotopic reference frame [78–80]. This means that when our eyes move and shift a stimulus on the retina, this shifts the imprint of the stimulus throughout visual cortex. Gaze position does impact visual responses but does so by primarily modulating response amplitudes encoded retinotopically [81,82]. Indeed, it is thought that non-linear interactions between vision, eye movements, somatosensation, and action are subserved by so-called gain-field mechanisms [83], whereby the gain of responses in one frame of reference is modulated by the spatial activations in another frame of reference. These gain-field mechanisms are thought to subserve attention’s influence on visual-spatial processing [83–85], providing a parsimonious mechanism for implementing goal-directed actions in the world while remaining fundamentally grounded in the visusospatial reference frame.

Whilst cortical maps that encode either retinal and somatosensory positions are present at birth [58,87], the computations needed to track multiple spatial codes during body movement develop later: Only after 6 months do infants start orienting visually and manually to the correct location of a touch on hands crossed along the midline, suggesting online alignment of multimodal spatial codes is still developing. In adults, crossing the limbs across the midline impairs the detection of tactile stimuli on the hand, attributed to spontaneous re-coding of the body-centered touch estimates into other reference frames [88]. This re-coding is also reflected in enhanced EEG signal amplitude evoked by a touch on the crossed versus uncrossed hand. Recent studies show that poorer touch localisation and EEG modulation for crossed hands are not present in pre-reaching infants [89,90] (**Figure 3E**) or congenitally blind adults [91,92]. Collectively, this suggests a driving role of visual experience with the moving limb in the development of action-perception integration. A key role of vision in touch encoding, is in line with the notion of gain-field mechanisms for sensorimotor integration that widely incorporate the visuospatial reference frame [82]; Because much of human action is visually guided this coding scheme may allow the developing system to improve its sensory predictions across modalities [93].

Frames of reference encoded in primary sensorimotor cortices form an early-developing scaffold for spatial coding. However, the translation of different frames and spatial codes into integrated representations requires learning the statistics of the world and our bodies as they act in it. The visuospatial reference frame may emerge as an efficient ubiquitous coding scheme because much of human action is visually guided.

Visuospatial coding as scaffold for human cognition

Over 100 years ago visuospatial coding was identified as an organizational principle of early visual cortex, and this hallmark finding has yielded profound insights into the computations underlying visual perception [2]. In this review, we summarized contemporary findings demonstrating its pervasiveness throughout the brain. From different vantage points, a convergent picture emerges that visuospatial information is not *‘thrown away’* by the brain, but instead is retained to not only facilitate efficient visual processing and the emergence of category-selectivity, but also the exchange of visual information with other functional systems to benefit broader cognition. What, then, may be the reason visuospatial coding is retained so prominently, and how might it aid cognition? While the underlying computational mechanisms remain largely unknown, we highlight several important aspects that may play a role.

First, it is important to consider where specific types of information typically appear in the visual field during natural viewing **(Figure 4A)**. We tend to fixate faces directly, thereby positioning them in the high-resolution fovea [13,94]. Objects that we interact with tend to be located at specific positions in the environment [95,96] (e.g. coffee cup in the lower visual field). Further, as we move through the world, navigable routes invariably occur in the lower visual field [97], while large-scale structures used to orient ourselves will be in the upper periphery [98]. Such regularities in ‘input statistics’ may help explain why higher-order visual regions exhibit the spatial biases they do **(Figure 4B)**.

For instance, the dominance of small foveally centred pRFs in face-selective regions potentially facilitates high-resolution processing required to distinguish identities [13,94]; whereas the larger more peripheral pRFs in scene-selective areas may enable efficient landmark recognition [1,48,99]. Objects are decoded from object-selective regions more accurately from the lower than upper visual field, reflecting their typical locations in visual space [96], and viewing face and body parts in typical locations evokes more distinct fMRI responses and can actively facilitate recognition [100–102]. Finally, the lower visual field bias in scene-selective OPA [48] **(see Figure 3A)** is consistent with its putative role in **navigational affordance** coding [97,103]. The idea that visuospatial coding is important for capturing functionally relevant input statistics is supported by studies on convolutional neural networks (CNNs), which can be thought of as representational systems unconstrained by wiring or energetic cost limitations, but whose inputs (i.e. photographs approximating human viewing behaviours such as a face at the centre) contain similar input statistics (**see Box 2**).

Developmental studies suggest these spatial biases capturing visual input statistics may scaffold onto pre-existing visuospatial coding schemes, thereby endowing certain cortical networks with the ability to efficiently sample visual field positions most relevant to the functions they subserve. For example, face-selective regions in infant macaques were predictable via pre-existing biases for foveal positions prior to selectivity emerging [58,59]. It is likely that the developing face-processing network co-opts these pre-existing foveal preferences to serve recognition [13,72,94]. Likewise, pRFs in the ventral occipitotemporal reading circuitry appear to overrepresent the location that words appear in the visual field during reading (i.e. horizontal meridian, **see Figure 2C**) [11,12]. It is possible that the VWFA emerges where it does in cortex through a combination of anatomical connections with anterior language regions [69,70], plus the prerequisite neural circuitry [13]. These data suggest that high-level visual areas not only contain visuospatial coding, but use this information in order to interface effectively with environmental demands [13,73]. Importantly, however, the representation in these areas continues to be shaped throughout development, by a persistent reinforcing interaction between input statistics and viewing behaviours.

Extrapolating this principle beyond high-level visual cortex, the common visuospatial coding scheme across multiple brain regions may facilitate *internal alignment* for information processing **(Figure 4C)**. That is, by continually referencing information processing to the visual reference frame, the brain can effectively integrate its distributed processing of items in the external world. The contralateral biases in the hippocampus [29,30] and visually-sensitive negative pRFs in the DMN [31] suggest that such a reference frame might even subserve a translation between stimulus-oriented perception and memory/mental-imagery. A recent proposal - the spatiotemporal similarity hypothesis [104] - posits that the hippocampus represents stimuli that co-occur in space and time and that these joint representations are used to generate visual predictions about what might be out there in the world and guide eye-movements to test those predictions. The contralateral biases suggest that the left hippocampus might encode potential sequences of eye movements related to the right visual field, and vice versa. Further, visuospatial coding also appears to support top-down mental imagery [105], with regions active during mental navigation showing clear contralateral visual field representations. Access to a common reference frame in perceptual and memory systems may also contribute to the successful retrieval of memory information through **gaze-reinstatement**, whereby the pattern of eye-movements during perception is recapitulated during cued memory retrieval [106]. The visually-selective negative pRFs within the DMN demonstrate how access to a common representation of visual space could serve to store perceptual signals before they are reused for other forms of cognition, such as mind-wandering. That positive pRFs in perceptual regions and negative pRFs in the DMN exhibit the same visual field biases, may also increase processing efficiency in cognitive computations. The visually tuned push-pull interaction between high-level processing in DMN and perceptual regions may serve to reduce redundancies in high-level cognitive processing [31].

Finally, a common reference frame could also help guide actions to external targets: completing the perception-action loop through *external alignment* of input and output spaces (**Figure 4D)**. For example, shared lower visual field biases between object-selective regions of visual cortex [37,43] the cerebellum [28], and the superior occipito-parietal cortex [76,77] could facilitate our ability to efficiently reach out and grasp objects in the lower visual field. Similarly, access to the same visuospatial code in face-selective regions [13,45,46] and frontal parietal areas governing eye movements [19] could combine when moving our eyes to fixate a friend's face. The aligned visual representation of space and somatosensory representation of the face in parietal cortex [74] and the correlated maps of visual and face-part space in occipital cortex [75] provide further illustrations of how shared access to visuospatial coding can align input and output spaces. We continuously monitor, react to and form expectations about elements in our environment. Given that visuospatial coding inherently reflects our window on the world, aligning different sensory inputs and modes of processing to this reference frame offers a plausible mechanism for achieving this efficiently.

The ubiquitous presence of visuospatial coding may facilitate interactions between brain regions through sensorimotor coding schemes that are tuned to the statistics of the environment. This perspective recasts visuospatial coding as a dimension orthogonally intersecting across more traditional distinctions between different cortical functions/systems (e.g. vision vs. memory; object identity vs. location), offering a means to not only internally align communication between different brain areas **(Figure 4C)**, but also externally align the perception-action loop **(Figure 4D)**.

Concluding remarks

Recent research has revealed the truly pervasive nature of visuospatial coding within a broad array of cortical and subcortical systems whose functions range from visual perception, to coordinating movements, mind-wandering and even memory. The early importance of visuospatial coding in development is also becoming evident, as is its potential to inform computational models of human brain function **(see Box 2)**. The ubiquity of visuospatial coding suggests that the brain makes use of this common reference frame to scaffold human cognition around environmental statistics and task demands. Continued investigations of *how* visuospatial codes interface with other sensory reference frames and its role in linking perception with memory will be key goals for the future (**see Outstanding Questions)**.

References

1 Malcolm, G.L. *et al.* (2016) Making Sense of Real-World Scenes. *Trends Cogn. Sci.* 20, 843–856
2 Wandell, B.A. *et al.* (2007) Visual Field Maps in Human Cortex. *Neuron* 56, 366–383
3 Holmes, G. (1918) DISTURBANCES OF VISION BY CEREBRAL LESIONS. *Br. J. Ophthalmol.* 2, 353–384

4 Tootell, R. *et al.* (1988) Functional anatomy of macaque striate cortex. II. Retinotopic organization. *J. Neurosci.* 8, 1531–1568
5 Engel, S.A. *et al.* (1994) fMRI of human visual cortex. *Nature* 369, 525–525
6 Sereno, M. *et al.* (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893
7 DeYoe, E.A. *et al.* (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl. Acad. Sci.* 93, 2382–2386
8 Dumoulin, S.O. and Wandell, B.A. (2008) Population receptive field estimates in human visual cortex. *NeuroImage* 39, 647–660
9 Wang, L. *et al.* (2015) Probabilistic Maps of Visual Topography in Human Cortex. *Cereb. Cortex* 25, 3911–3931
10 Benson, N.C. *et al.* (2018) The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *J. Vis.*
11 Rauschecker, A.M. *et al.* (2012) Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci.* 109, E1568–E1577
12 Le, R. *et al.* (2017) The field of view available to the ventral occipito-temporal reading circuitry. *J. Vis.* 17, 6
13 Gomez, J. *et al.* (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* 9, 788
14 Sereno, M.I. (2001) Mapping of Contralateral Space in Retinotopic Coordinates by a Parietal Cortical Area in Humans. *Science* 294, 1350–1354
15 Hagler, D.J. and Sereno, M.I. (2006) Spatial maps in frontal and prefrontal cortex. *NeuroImage* 29, 567–577
16 Silver, M.A. and Kastner, S. (2009) Topographic maps in human frontal and parietal cortex. *Trends Cogn. Sci.* 13, 488–495
17 Corbetta, M. *et al.* (1998) A Common Network of Functional Areas for Attention and Eye Movements. *Neuron* 21, 761–773
18 Duncan, J. (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–179
19 Mackey, W.E. *et al.* (2017) Visual field map clusters in human frontoparietal cortex. *eLife* 6, e22974
20 Elshout, J.A. *et al.* (2018) Human V2A: A map of the peripheral visual hemifield with functional connections to scene-selective cortex. *J. Vis.* 18, 22
21 Mikellidou, K. *et al.* (2017) Area Prostriata in the Human Brain. *Curr. Biol.* 27, 3056-3060.e3
22 Arcaro, M.J. *et al.* (2015) The Anatomical and Functional Organization of the Human Visual Pulvinar. *J. Neurosci.* 35, 9848–9871
23 DeSimone, K. *et al.* (2015) Population Receptive Field Estimation Reveals New Retinotopic Maps in Human Subcortex. *J. Neurosci.* 35, 9836–9847
24 Schneider, K.A. (2011) Subcortical Mechanisms of Feature-Based Attention. *J. Neurosci.* 31, 8643–8653
25 Rahmati, M. *et al.* Spatially-specific working memory activity in the human superior colliculus. *Journal of Neuroscience*, *40*(49), 9487-9495.
26 Savjani, R.R. *et al.* (2018) Polar-angle representation of saccadic eye movements in human superior colliculus. *NeuroImage* 171, 199–208
27 Brissenden, J.A. *et al.* (2018) Topographic Cortico-cerebellar Networks Revealed by Visual Attention and Working Memory. *Curr. Biol.* 28, 3364-3372.e5
28 van Es, D.M. *et al.* (2019) Topographic Maps of Visual Space in the Human Cerebellum. *Curr. Biol.* 29, 1689-1694.e3
29 Silson, E.H. *et al.* (2021) Representation of contralateral visual space in the human

hippocampus. *J. Neurosci.* DOI: 10.1523/JNEUROSCI.1990-20.2020
30 Knapen, T. (2021) Topographic connectivity reveals task-dependent retinotopic processing throughout the human brain. *Proc. Natl. Acad. Sci.* 118, e2017032118
31 Szinte, M. and Knapen, T. (2020) Visual Organization of the Default Network. *Cereb. Cortex* 30, 3518–3527
32 Chklovskii, D.B. and Koulakov, A.A. (2004) MAPS IN THE BRAIN: What Can We Learn from Them? *Annu. Rev. Neurosci.* 27, 369–392
33 Konkle, T. (2021) *Emergent organization of multiple visuotopic maps without a feature hierarchy*, Neuroscience.
34 Kanwisher, N. and Dilks, D.D. (2013) The functional organization of the ventral visual pathway in humans. In *The new visual neurosciences*
35 Levy, I. *et al.* (2001) Center–periphery organization of human object areas. *Nat. Neurosci.* 4, 7
36 Hasson, U. *et al.* (2002) Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron* 34, 479–490
37 Larsson, J. and Heeger, D.J. (2006) Two Retinotopic Visual Areas in Human Lateral Occipital Cortex. *J. Neurosci.* 26, 13128–13142
38 Brewer, A.A. *et al.* (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat. Neurosci.* 8, 1102–1109
39 Amano, K. *et al.* (2009) Visual Field Maps, Population Receptive Field Sizes, and Visual Field Coverage in the Human MT+ Complex. *J. Neurophysiol.* 102, 2704–2718
40 Arcaro, M.J. *et al.* (2009) Retinotopic Organization of Human Ventral Visual Cortex. *J. Neurosci.* 29, 10638–10652
41 Arcaro, X.M.J. and Livingstone, X.M.S. Retinotopic Organization of Scene Areas in Macaque Inferior Temporal Cortex. *Journal of Neuroscience*, *37*(31), 7373-7389.
42 Strong, S.L. *et al.* (2017) A Direct Demonstration of Functional Differences between Subdivisions of Human V5/MT+. *Cereb. Cortex* 27, 1–10
43 Silson, E.H. *et al.* (2013) Specialized and independent processing of orientation and shape in visual field maps LO1 and LO2. *Nat. Neurosci.* 16, 267–269
44 Nasr, S. *et al.* (2011) Scene-Selective Cortical Regions in Human and Nonhuman Primates. *J. Neurosci.* 31, 13771–13785
45 Silson, E.H. *et al.* (2016) Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J. Vis.* 16, 14
46 Kay, K.N. *et al.* (2015) Attention Reduces Spatial Uncertainty in Human Ventral Temporal Cortex. *Curr. Biol.* 25, 595–600
47 Finzi, D. *et al.* (2020) *Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections*, Neuroscience.
48 Silson, E.H. *et al.* (2015) A Retinotopic Basis for the Division of High-Level Scene Processing between Lateral and Ventral Human Occipitotemporal Cortex. *J. Neurosci.* 35, 11921–11935
49 Hong, H. *et al.* (2016) Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat. Neurosci.* 19, 613–622
50 Apurva Ratan Murty, N. and Arun, S.P. (2018) Multiplicative mixing of object identity and image attributes in single inferior temporal neurons. *Proc. Natl. Acad. Sci.* 115, E3276–E3285
51 Kravitz, D.J. *et al.* (2010) High-Level Visual Object Representations Are Constrained by Position. *Cereb. Cortex* 20, 2916–2925
52 Huberman, A.D. *et al.* (2008) Mechanisms Underlying Development of Visual Maps and Receptive Fields. *Annu. Rev. Neurosci.* 31, 479–509

53 Kim, J. *et al.* (2020) Spontaneous Retinal Waves Can Generate Long-Range Horizontal Connectivity in Visual Cortex. *J. Neurosci.* 40, 6584–6599
54 Gribizis, A. *et al.* (2019) Visual Cortex Gains Independence from Peripheral Drive before Eye Opening. *Neuron* 104, 711-723.e3
55 Dekker, T.M. *et al.* (2019) Population receptive field tuning properties of visual cortex during childhood. *Dev. Cogn. Neurosci.* 37, 100614
56 Ellis, C.T. *et al.* (2021) Retinotopic organization of visual cortex in human infants. *Neuron* DOI: 10.1016/j.neuron.2021.06.004
57 Deen, B. *et al.* (2017) Organization of high-level visual cortex in human infants. *Nat. Commun.* 8, 13995
58 Arcaro, M.J. and Livingstone, M.S. (2017) A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife* 6, e26196
59 Arcaro, M.J. *et al.* (2017) Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* 20, 16
60 Arcaro, M.J. *et al.* (2019) Universal Mechanisms and the Development of the Face Network: What You See Is What You Get. *Annu. Rev. Vis. Sci.* 5, 341–372
61 Srihasam, K. *et al.* (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783
62 Gomez, J. *et al.* (2019) Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nat. Hum. Behav.* 3, 611–624
63 Op de Beeck, H.P. *et al.* (2019) Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. *Trends Cogn. Sci.* 23, 784–797
64 Mahon, B.Z. *et al.* (2009) Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron* 63, 397–405
65 Ratan Murty, N.A. *et al.* (2020) Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus. *Proc. Natl. Acad. Sci.* 117, 23011–23020
66 van den Hurk, J. *et al.* (2017) Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci.* 114, E4501–E4510
67 Mattioni, S. *et al.* (2020) Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *eLife* 9, e50732
68 Striem-Amit, E. *et al.* (2015) Functional connectivity of visual cortex in the blind follows retinotopic organization principles. *Brain J. Neurol.* 138, 1679–1695
69 Kamps, F.S. *et al.* (2020) Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc. Natl. Acad. Sci.* 117, 6163–6169
70 Saygin, Z.M. *et al.* (2016) Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* 19, 1250–1255
71 Li, J. *et al.* (2020) Innate connectivity patterns drive the development of the visual word form area. *Sci. Rep.* 10, 18039
72 Arcaro, M.J. and Livingstone, M.S. (2021) On the relationship between maps and domains in inferotemporal cortex. *Nat. Rev. Neurosci.* DOI: 10.1038/s41583-021-00490-4
73 Poltoratski, S. *et al.* (2021) Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions. *Nat. Commun.* 12, 4745
74 Sereno, M.I. and Huang, R.-S. (2006) A human parietal face area contains aligned head-centered visual and tactile maps. *Nat. Neurosci.* 9, 1337–1343
75 de Haas, B. *et al.* (2021) Inferior Occipital Gyrus Is Organized along Common Gradients of Spatial and Face-Part Selectivity. *J. Neurosci.* 41, 5511–5521
76 Rossit, S. *et al.* (2013) fMRI reveals a lower visual field preference for hand actions in human superior parieto-occipital cortex (SPOC) and precuneus. *Cortex* 49, 2525–2541
77 Filimon, F. *et al.* (2009) Multiple Parietal Reach Regions in Humans: Cortical Representations for Visual and Proprioceptive Feedback during On-Line Reaching. *J. Neurosci.* 29, 2961–2971

78 Gardner, J.L. *et al.* (2008) Maps of Visual Space in Human Occipital Cortex Are Retinotopic, Not Spatiotopic. *J. Neurosci.* 28, 3988–3999

79 Knapen, T. (2016) Oculomotor Remapping of Visual Information to Foveal Retinotopic Cortex. *Front. Syst. Neurosci.* 10, 12

80 Golomb, J.D. and Kanwisher, N. (2012) Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. *Cereb. Cortex* 22, 2794–2810

81 Andersen, A. and Mountcastle, B. (1983) THE INFLUENCE OF THE ANGLE OF GAZE UPON THE EXCITABILITY OF THE LIGHT-SENSITIVE NEURONS OF THE POSTERIOR PARIETAL CORTEX. *J. Neurosci.* 3, 532–548

82 Merriam, E.P. *et al.* (2013) Modulation of Visual Responses by Gaze Direction in Human Visual Cortex. *J. Neurosci.* 33, 9879–9889

83 Pouget, A. *et al.* (2002) A computational perspective on the neural basis of multisensory spatial representations. *Nat. Rev. Neurosci.* 3, 741–747

84 Klein, B.P. *et al.* (2014) Attraction of Position Preference by Spatial Attention throughout Human Visual Cortex. *Neuron* 84, 227–237

85 Womelsdorf, T. *et al.* (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat. Neurosci.* 9, 1156–1160

86 van Es, D.M. *et al.* (2018) Spatial sampling in human visual cortex is modulated by both spatial and feature-based attention. *eLife* 7, e36928

87 Arcaro, M.J. *et al.* (2019) Body map proto-organization in newborn macaques. *Proc. Natl. Acad. Sci.* 116, 24861–24871

88 Yamamoto, S. and Kitazawa, S. (2001) Reversal of subjective temporal order due to arm crossing. *Nat. Neurosci.* 4, 759–765

89 Rigato, S. *et al.* (2014) The neural basis of somatosensory remapping develops in human infancy. *Curr. Biol. CB* 24, 1222–1226

90 Begum Ali, J. *et al.* (2015) Human infants' ability to perceive touch in external space develops postnatally. *Curr. Biol.* 25, R978–R979

91 Roder, B. *et al.* (2007) Developmental vision determines the reference frame for the multisensory control of action. *Proc. Natl. Acad. Sci.* 104, 4753–4758

92 Crollen, V. *et al.* (2017) How visual experience impacts the internal and external spatial mapping of sensorimotor functions. *Sci. Rep.* 7, 1022

93 Cohen, Y.E. and Andersen, R.A. (2002) A common reference frame for movement plans in the posterior parietal cortex. *Nat. Rev. Neurosci.* 3, 553–562

94 Finzi, D. *et al.* (2021) Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections. *Nat. Commun.* 12, 2278

95 Kaiser, D. *et al.* (2019) Object Vision in a Structured World. *Trends Cogn. Sci.* 23, 672–685

96 Kaiser, D. and Cichy, R.M. (2018) Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J. Neurophysiol.* 120, 848–853

97 Bonner, M.F. and Epstein, R.A. Coding of navigational affordances in the human visual system. *Proc. Natl. Acad. Sci.114*(18), 4793-4798.

98 Epstein, R.A. and Baker, C.I. (2019) Scene Perception in the Human Brain. *Annu. Rev. Vis. Sci.* 5, 373–397

99 Groen, I.I.A. *et al.* (2017) Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160102

100 Chan, A.W.-Y. *et al.* (2010) Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat. Neurosci.* 13, 417–418

101 Quek, G.L. and Finkbeiner, M. (2014) Face-sex categorization is better above fixation than below: Evidence from the reach-to-touch paradigm. *Cogn. Affect. Behav. Neurosci.* 14, 1407–1419

102 Peterson, M.F. and Eckstein, M.P. (2012) Looking just below the eyes is optimal across face recognition tasks. *Proc. Natl. Acad. Sci.* 109, E3314–E3323

103 Bonner, M.F. and Epstein, R.A. Computational mechanisms underlying cortical responses to the affordance properties of visual scenes.

104 Turk-Browne, N.B. (2019) The hippocampus as a visual area organized by space and time: A spatiotemporal similarity hypothesis. *Vision Res.* 165, 123–130

105 Huang, R.-S. and Sereno, M.I. (2013) Bottom-up Retinotopic Organization Supports Top-down Mental Imagery. *Open Neuroimaging J.* 7, 58–67

106 Wynn, J.S. *et al.* (2021) *Neural correlates of subsequent memory-related gaze reinstatement*. *bioRxiv*.

107 Wandell, B.A. and Winawer, J. (2011) Imaging retinotopic maps in the human brain. *Vision Res.* 51, 718–737

108 Haak, K.V. *et al.* (2013) Connective field modeling. *NeuroImage* 66, 376–384

109 Gomez, J. (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* *9*(1), 1-12.

110 Zuiderbaan, W. *et al.* (2012) Modeling center-surround configurations in population receptive fields using fMRI. *J. Vis.* 12, 10–10

111 Kay, K.N. *et al.* (2013) Compressive spatial summation in human visual cortex. *J. Neurophysiol.* 110, 481–494

112 Silson, E.H. *et al.* (2018) Differential Sampling of Visual Space in Ventral and Dorsal Early Visual Cortex. *J. Neurosci.* 38, 2294–2303

113 Lerma-Usabiaga, G. *et al.* (2021) Population Receptive Field Shapes in Early Visual Cortex Are Nearly Circular. *J. Neurosci.* 41, 2420–2427

114 Aqil, M. *et al.* (2020) Biologically inspired unification of population receptive field models provides new insights into cortical computations. *J. Vis.* 20, 377

115 Lerma-Usabiaga, G. *et al.* (2020) A validation framework for neuroimaging software: The case of population receptive fields. *PLOS Comput. Biol.* 16, e1007924

116 Ekman, M. *et al.* (2017) Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* 8, 15276

117 Naselaris, T. *et al.* (2015) A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* 105, 215–228

118 Farahbakhsh, M. *et al.* (2020) A demonstration of cone function plasticity after gene therapy in achromatopsia. *medRxiv* DOI: 10.1101/2020.12.16.20246710

119 Dumoulin, S.O. and Knapen, T. (2018) How Visual Cortical Organization Is Altered by Ophthalmologic and Neurologic Disorders. *Annu. Rev. Vis. Sci.* 4, 357–379

120 Ritter, M. *et al.* (2019) Correspondence between retinotopic cortical mapping and conventional functional and morphological assessment of retinal disease. *Br. J. Ophthalmol.* 103, 208–215

121 Silson, E.H. *et al.* (2018) Comparing Clinical Perimetry and Population Receptive Field Measures in Patients with Choroideremia. *Investig. Opthalmology Vis. Sci.* 59, 3249

122 Benson, N.C. *et al.* (2021) *Variability of the Surface Area of the V1, V2, and V3 Maps in a Large Sample of Human Observers*, Neuroscience.

123 Benson, N.C. *et al.* (2014) Correction of Distortion in Flattened Representations of the Cortical Surface Allows Prediction of V1-V3 Functional Organization from Anatomy. *PLoS Comput. Biol.* 10, e1003538

124 Wang, L. *et al.* (2015) Probabilistic Maps of Visual Topography in Human Cortex. *Cereb. Cortex* 25, 3911–3931

125 Benson, N.C. and Winawer, J. (2018) Bayesian analysis of retinotopic maps. *eLife* 7, e40224

126 Rawat, W. and Wang, Z. (2017) Deep Convolutional Neural Networks for Image

Classification: A Comprehensive Review. *Neural Comput.* 29, 2352–2449
127 Lindsay, G.W. (2020) Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *J. Cogn. Neurosci.* DOI: 10.1162/jocn_a_01544
128 Kietzmann, T.C. *et al.* Deep Neural Networks in Computational Neuroscience.
129 Mohsenzadeh, Y. *et al.* (2020) Emergence of Visual Center-Periphery Spatial Organization in Deep Convolutional Neural Networks. *Sci. Rep.* 10, 4638
130 Dwivedi, K. *et al.* (2020) Unraveling Representations in Scene-selective Brain Regions Using Scene-Parsing Deep Neural Networks. *J. Cogn. Neurosci.* DOI: 10.1162/jocn_a_01624
131 Yu, C.-P. *et al.* (2019) Modelling attention control using a convolutional neural network designed after the ventral visual pathway. *Vis. Cogn.* 27, 416–434
132 Mnih, V. *et al.* Recurrent Models of Visual Attention.
133 Cheung, B. *et al.* (2017) Emergence of foveal image sampling from learning to attend in visual scenes. *ArXiv161109430 Cs* at <http://arxiv.org/abs/1611.09430>
134 Wang, P. and Cottrell, G.W. (2017) Central and peripheral vision for scene recognition: A neurocomputational modeling exploration. *J. Vis.* 17, 9
135 Wu, K. *et al.* (2018) , Learning scene gist with convolutional neural networks to improve object recognition. , in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, pp. 1–6
136 Deza, A. and Konkle, T. (2021) Emergent Properties of Foveated Perceptual Systems. *ArXiv200607991 Cs Eess Q-Bio* at <http://arxiv.org/abs/2006.07991>
137 Lindsey, J. *et al.* (2019) *A Unified Theory Of Early Visual Representations From Retina To Cortex Through Anatomically Constrained Deep CNNs*, Neuroscience.

Box 1: Measuring visuospatial coding in the human brain

Several excellent retinotopy reviews already exist [2,107], so here we provide only a brief overview of visuospatial mapping techniques and additional considerations: ***Travelling-wave:*** Fixating observers typically view high-contrast flickering stimuli that either rotate around or expand through the visual field. By measuring when during the stimulus cycle the largest response at each cortical location occurs, the preferred polar coordinates can be inferred (angle, eccentricity). ***pRF modelling:*** Fixating observers typically view a traversing bar. This stimulus information is combined with hypothetical receptive fields described by a centre location (x, y coordinate), size and the hemodynamic response function to generate predicted responses. Parameters are adjusted to minimise the difference between the data and the prediction. The final model provides the best pRF estimate at each brain location. ***Connective field (CF) modelling:*** An extension of pRF modelling [30,108], defined not in visual space but in cortical surface coordinates. To generate model time-courses, CF modelling uses activation patterns across brain regions. Its parameters are the surface location of the center of the connective field, and its size. Because it is brain-referred and not stimulus-referred, this method can be applied to data from any experimental paradigm. ***Stimulus materials**:* Mapping stimuli have been modified to better capture naturalistic viewing conditions, embedding high-level categories to elicit responses in face-, scene- and word-selective regions [10,12,46–48,55,109]. ***Modelling strategies:*** Different pRF models now incorporate suppressive surrounds [110], compressive spatial summation [111] changes in pRF shape [112,113] and divisive normalisation [114]. Importantly, systematic comparisons of different implementations are beginning to take place [115]. ***Reconstructing Representations:*** These models are increasingly used to reconstruct how different visual regions and cortical layers represent real/imagined visual stimuli [116,117]. A clinically relevant application is their use to assess visual field loss and recovery in visual disorders [118–121]. ***Map delineation:*** Challenges remain with respect to delineating retinotopic maps in individuals: Labs may adopt different heuristics regarding the borders between maps. Recent work indicates that such basic features such as the areal extent of early visual cortex may vary 2-3 fold across individuals [122]. That said, taking into account an individual's particular anatomy allows one to predict the location of early visual maps from anatomy alone [123]. One promising way to effectively deal with inter-individual variability at higher levels is to combine an anatomical projection of a learned atlas (e.g., a group-average [10,124]) with fMRI measurements of retinotopy using Bayesian approach [125].

Box 2: Visuospatial coding in deep convolutional neural networks

Popular contemporary computational models of vision are deep convolutional neural networks (CNNs, see **Glossary**), which achieve human-level performance on many visual recognition tasks [126]. CNNs also predict brain responses to natural images in multiple visual brain regions with unprecedented accuracy (see [127,128] for recent reviews). Do CNNs use visuospatial coding? The feature maps inside convolutional CNN layers have a 2D layout and thus an inherent form of retinotopy, but they also have properties inconsistent with visuospatial coding in biological brains, such as complete weight sharing across the visual field [128]. Nevertheless, some signatures of visuospatial coding observed in higher-level visual cortex have also been found in CNNs. For example, object position can be decoded not just from low- but also high-level CNN layers [49,50], and face- and scene-specialized CNNs exhibit center-periphery biases similar to face- and scene-selective brain regions [129]. Conversely, CNNs trained to predict the spatial location of image components better predicted fMRI responses in scene-selective regions than CNNs trained on whole-scene category labels ([130]), providing additional support for the presence of visuospatial representations in higher visual cortex. Might building visuospatial information into CNNs more explicitly also benefit CNN task performance? Visuospatial coding in brains is characterized by systematic organization of receptive field sizes, which increase from lower to higher areas and from fovea to periphery. Convolutional filter sizes and pooling regions also increase across CNN layers by design, but building fovea-periphery organization into feature maps is technically difficult [131]. However, different filter sizes within a feature map have been used when building attention mechanisms into CNNs [131,132], and fovea-periphery organization emerged naturally in a neural network model that spatially sampled information by translating the retina across the visual input (similar to the eye making saccades) [133]. Computational advantages of foveal-periphery organisation are also evident in dual-stream architectures [134,135], and recent work suggests that training CNNs with foveation via peripheral texture-based computation yields robust representations compared to non-foveated systems [136]. Exploring the computational advantages of explicit visuospatial coding in CNNs is likely a promising avenue for future research, especially when applied to artificial visual systems with limited computational power (e.g., robots) that, like biological brains, may suffer from metabolic or implementation bottlenecks [137].

Glossary

Category-selectivity: The phenomenon that some brain regions respond more strongly to images of specific stimulus classes, such as faces, scenes, objects, bodies and words, compared to others.

Connective field modelling: A computational modeling approach that explains signal time-courses as resulting from connectivity with a localized patch on the surface of a brain region such as V1. This provides a way of mapping retinotopic brain organization regardless of stimulus paradigm.

Convolutional neural network (CNN): Artificial neural networks with multiple layers containing a computational hierarchy that stacks 2D convolutions, non-linear activation functions and response pooling in order to perform a learned task (e.g. visual object classification on images).

Contralateral visual field: The visual field opposite to the cortical hemisphere (e.g. left hemisphere - right visual field and vice-versa).

Cortical magnification: Describes the relationship between visual field locations and their cortical representations. Positions near the centre of gaze are processed by a disproportionately large amount of cortex as compared to larger portions of the periphery.

Default mode network: A series of brain regions that operate as a network and are principally associated with internally generated cognition, such as memory and/or mind-wandering.

Eccentricity: The distance from the centre of gaze is mapped systematically in cortex. Within early visual areas (i.e. V1-V3) eccentricity is represented along a posterior-anterior axis, such that locations close to the centre of gaze correspond to the occipital pole and more peripheral visual field locations correspond to more anterior areas.

Foveal vision: Vision in the center of the field of vision, where visual acuity is highest.

Gaze-reinstatement: When the pattern of eye-movements made during perception of a stimulus (e.g. when looking at a face or scene) is replayed when recalling that same stimulus from memory.

Navigational affordances: High-level representations of scenes that are thought to encode features of the local environment including, available routes of egress, objects to interact with and obstacles to avoid.

Peripheral vision: Vision outside the point of fixation, i.e. away from the center of gaze. The vast majority of the visual field is included in the notion of peripheral vision.

Polar angle: Positions in the visual field are mapped systematically in cortex in polar coordinates. Within early visual areas (i.e. V1-V3) polar angle is represented along a dorsal-ventral axis, such that locations in the lower visual field are mapped dorsally and locations in the upper visual field are mapped ventrally. In early visual areas in particular, eccentricity and polar angle are mapped orthogonally.

Population receptive field modelling: A computational modelling approach that estimates the effective receptive field of individual fMRI voxels.

Receptive Field: In visual cortex, this refers to the location and extent of the visual field over which the activity of the neuron can be modulated.

Resting-state functional connectivity: Refers to a fMRI analysis technique which takes into account the similarity in spontaneous activity of brain regions at rest (i.e. in the absence of an overt task). It is commonly assumed that the spontaneous activity of regions that share functionality (e.g. OFA/FFA) will be highly positively correlated.

Retinotopic (or visual field) map: The systematic mapping of the retina and thus the visual field across the cortical surface. Retinotopic maps typically represent either the entire contralateral hemifield (e.g. V1) or a quadrant of the contralateral visual field (e.g. V2d).

Visual field biases: Coarse-level measurements that demonstrate a preference for particular visual field positions but do not represent a systematic mapping of visual space. When derived from pRF modelling experiments these summarise the sensitivity of a given area to positions in the visual field.

visuospatial coding: Refers to the general principle that the brain appears to encode spatial elements in the world with respect to where those elements fall in the visual field and thus the retina. We posit that visuospatial coding comes in two main forms: retinotopic maps and visual field biases.

World-centered spatial processing (spatiotopy): Refers to a coding scheme in which the brain represents the absolute location of elements in the world.

Figure legends

Figure 1: Forms of visuospatial coding: A) Retinotopic organisation of early visual cortex (V1, V2d, V3d) **Left:** V1 (solid-lines) exhibits a complete and inverted representation of the contralateral visual field. Unlike V1, V2d/V2v (dashed-lines) exhibit a quadrant representation of the visual field (V2d = lower visual field, V2v = upper visual field). **Middle:** Representations of eccentricity demonstrate that the fovea is represented posteriorly in V1, V2d/V2v with the periphery represented anteriorly. Cortical magnification is also evident with the amount of V1, V2d/V2v dedicated to the fovea being disproportionately larger than that dedicated to the periphery. **Right:** pRF sizes vary systematically from smallest posteriorly in V1, V2d/V2v to largest more anteriorly. Eccentricity and pRF gradients are positively correlated. Data taken from the HCP initiative and analyzed by T. Knapen. **B)** Schematic of pRF modelling and visual field coverage of the reading circuitry. **Top:** Overlaid onto a schematic of the visual field are many individual receptive fields (white dot = RF centre, white circle = RF). Single fMRI voxels (at conventional resolutions) contain many thousands of neurons and thus many thousands of receptive fields. The pRF model produces a single estimate of these individual receptive fields described by a centre location (x, y coordinates = black dot) and size (diameter of the pRF = white arrow). **Bottom:** An enlarged ventral view of the left hemisphere is shown. Overlaid in hot-colours are regions of the brain that preferentially respond to written words over line drawings of objects - collectively referred to as the ventral occipitotemporal reading circuitry (VOTC) [12]. When pRFs in the VOTC are combined across hemispheres, their visual field coverage (VF coverage) extends further horizontally than vertically, which mirrors the location in the visual field of words during reading (e.g. English). VF coverage is schematic but based on [12]. **C)** Relationship between retinotopic maps and category-selective regions of cortex. **Top:** Schematic of the historical view that retinotopic maps were the purview of early visual areas only. The borders of V1-V3 (taken from [9]) are shown in white and appear posterior of regions selective for scenes (pink), faces (blue), bodies (yellow), objects (green), motion (blue) and color (red). **Bottom:** Schematic of the contemporary view of the relationship between retinotopic and category-selective regions. The borders of multiple retinotopic maps (taken from [9]) are shown in white and spatially overlap many category-selective regions.

Figure 2: visuospatial coding outside the traditional visual system. A) Connective Field Modeling (**see Glossary and Box 1)** during movie watching demonstrates retinotopic organization throughout the cerebral cortex and **B)** subcortex. **C)** In the subcortex, polar angle representations of retinotopic preferences show contralateral visual position preferences in thalamus, superior colliculus, and caudate nucleus; in cerebellum, the same visualization demonstrates ipsilateral visual field preferences. The sagittal slice shows the axial planes visualised here, as well as the polar-angle color scale used in **C & D**. **D)** Contralateral visual field preferences in hippocampus. C, D, & E all show the results of a pRF analysis on the bar runs of the Human Connectome Project retinotopic mapping dataset [10]. **E)** Contralateral visual field preferences of DMN deactivations in the angular gyrus of the parietal lobe, re-presentation of results from [31]. ROI positions are indicated by the white portion on the flatmap insets. In this rendition, the dashed line indicates a putative polar angle reversal at the lower vertical meridian. The lower panels show polar angle along the surface in the direction perpendicular to the putative reversal boundaries, with the dashed vertical lines highlighting the reversal at the lower vertical meridian. Subpanel E courtesy of Martin Szinte.

Figure 3: Computational/developmental benefits of visuospatial coding. A) Computational benefits of OPAs lover field bias. **Top left:** Routes of egress are highlighted on an example scene (red-lines). **Bottom left:** OPA is biased to the lower visual field. **Top right:** Response patterns to natural scenes in OPA show high correlation with human ratings of the same scenes in terms of navigational affordances, redrawn with permission from [97]. **Bottom right:** Lower field bias in predictive accuracy of navigational affordances in a convolutional neural network (CNN) and the OPA. The CNN effects were driven mostly by features in the lower visual field, redrawn with permission from [103]. **B)** Retinotopy present early in development. **Top:** Retinotopic organisation in infant macaques demonstrated through resting-state functional connectivity with V1. Retinotopic organisation shows adult-like structure even at <18 days of age. Colors correspond to seed eccentricity bands in V1 (see inset). Black arrows indicate three regions most correlated with the foveal representation. **Bottom:** Statistical maps of the contrast Faces vs Objects at <123 days **(left)** and >1.5 years **(right)**. Black arrows show the foveal regions highlighted above. At <123 days no face-selectivity is detectable, whereas robust face-selectivity is present at >1.5 years. Emerging face-selectivity is localised to regions already foveally tuned, redrawn with permission from [58]. **C)** Developmental changes in visual field coverage. Group-average visual field coverage from pFus-faces overlaid onto a face seen at ~1m for children **(top)** and adults **(bottom)**. The 50% density contour is shown by the white line and covers more of the average-sized face at typical viewing distance in adults than children, redrawn with permission from [13]. **D)** Spatial overlap of neural sensitivity to face-part and visual space. Face-selective inferior occipital gyrus (IOG) is outlined in black. Neural tuning to different face parts changes from upper-lower along the posterior-anterior gradient within IOG. Neural tuning to different visual field positions also changes from upper to lower along the same axis. **E)** Somatosensory evoked potentials in infants. **Top:** Grand averaged SEPs from central electrodes contralateral to the stimulated hand shown for 8-month old infants in the crosser **(top)** and noncrosser **(bottom)** groups. Crossed (dashed black), uncrossed (solid black) and difference (gray) waveforms are shown. The shaded area indicates the time course of reliable effects of posture on somatosensory

processing. There was no effect of posture in the noncrossers, but the crossers showed an effect between 298 and 392 ms. Adapted with permission from [89].

Figure 4: Visuospatial coding as a scaffold for human cognition. A) Visuospatial coding is the format in which visual information is constantly received, and different types of visual information are typically present in different parts of the visual field. For example, while having a conversation with a friend in a café, we will typically place facial information centrally in the visual field, while our hands holding the coffee cup will be positioned in the lower visual field, and the spatial layout of the room will appear in the visual periphery. **B)** Category-selective brain regions will more often receive early visual inputs from stereotypical parts of the visual field and may therefore have developed visuospatial biases for those parts (e.g. foveal bias for face-selective regions, lower field bias for object/tool-selective regions , and peripheral bias for scene-selective regions. A medial/ventral view of the right hemisphere is shown, but the similar visuospatial biases are also present in category-selective regions considered part of the dorsal stream **C)** Shared visuospatial coding between regions creates a common reference frame that facilitates *internal alignment* of communication, allowing them to bind visual features together (e.g. object identity and location), and with other sensory information (e.g., auditory, somatosensory, memory) facilitating multi-modal interactions. **D)** Since actions in the world need to occur at specific spatial coordinates, visuospatial coding also facilitates continuous *external alignment* of perceived visual information, visually-guided actions and their subsequent effects.