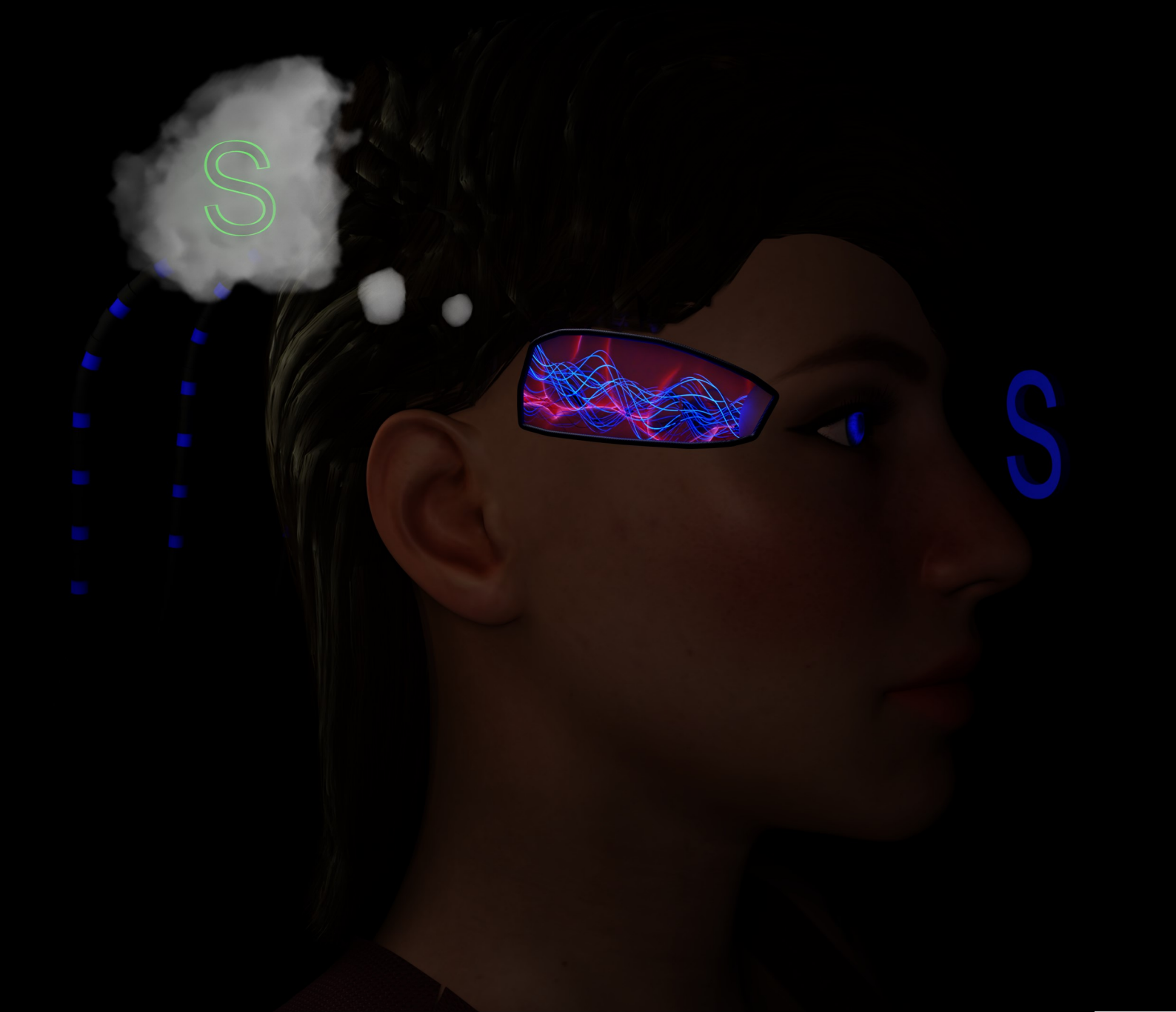
# The brain as image processor and generator

HENRICUS JOHANNES VAN HOOF

# THE BRAIN AS IMAGE PROCESSOR AND GENERATOR

## towards function-restoring brain-computer-interfaces



Maastricht University

DOCTORAL THESIS

---

# The brain as image processor and generator

*towards function-restoring brain-computer-interfaces*

---

*To obtain the degree of Doctor at Maastricht University*
*on the authority of the Rector Magnificus, Prof. Dr. Pamela Habiboviç,*
*in accordance with the decision of the Board of Deans,*
*to be defended in public on*
*Friday, 3rd of June 2022, at 10:00 hours*

*by*

Henricus Johannes (Rick) VAN HOOF

*Promotor:*
Prof. Rainer GOEBEL

*Co-promotors:*
Dr. Mario SENDEN
Dr. Joel REITHLER

*Assessment Committee:*
Prof. Dr. Beatrice DE GELDER *(chair)*

Prof. Dr. Eduardo FERNANDEZ JOVER
Miguel Hernández University

Prof. Dr. Brigitte RŐDER
University Hamburg

Dr. Judith PETERS



Printed by: Ridderprint, www.ridderprint.nl
Cover: Rick van Hoof
ISBN: 978-94-6458-281-9

# Contents

## Chapter 1

# General Introduction

The human neocortex has evolved to be a sophisticated information processor with roughly 89 billion neurons (Gazzaniga, Ivry, & Mangun, 2018). Astonishingly, about 20 percent of the entire cortical area is specifically devoted to vision (Maunsell & Newsome, 1987; Wandell et al., 2009). Yet, humans (and all other organisms for that matter) merely register a very small fraction of visual information available at any given time. Nonetheless, a representation of space can be obtained through several sensory modalities. It is the visual sense, that ultimately is central to the human representation of space (Marr, 1982) and what allowed early day homo sapiens to navigate terrain and avoid predators. In recent days, our survival depends on the ability to participate in society. This often translates to maintaining employment, and in many occupations, communication is vital and dependent on information conveyed via written text or digital screens. Visual information even appears to frame the way we think. Our memories and even much of our language refer to vision, when used to describe mental concepts with metaphors (Gazzaniga et al., 2018). Therefore, it is not surprising that losing the ability of sight or the ability to communicate can have a devastating effect on people's lives.

Globally, an estimated 42 million people are blind. The most common causes are cataract, glaucoma, age-related macular degeneration, corneal opacity, diabetic retinopathy, and trachoma (Flaxman et al., 2017). Blindness can occur when any part of the visual pathway is damaged or injured. In vision, eye movements regulate where we look, the size of the pupil adjusts to filter the light, and the cornea and lens serve to focus the light, which is then transmitted into electrical signals, much like a camera does. This analogy has inspired researchers to find ways to restore basic visual abilities in the blind by artificially bypassing parts of the visual pathway. This thesis focusses particularly on how the human brain processes visual information and how this knowledge might lead to technological advancements that, in the future, could be used to enhance or restore brain functioning.

## 1.1 The brain as an image processor

Hundreds of million years ago early organisms evolved photoreceptors that were capable of detecting light, and enabled image-forming eyes and visual systems to emerge (Lamb et al., 2007). In primates, the optical structures of the eye —the cornea, pupil, and lens—filter and focus the light that eventually reaches photoreceptor cells in the retina. Photoreceptors are specialized receptor cells whose pigment molecules absorb photons of specific wavelengths and transform the information conveyed by light energy into neural signals. Due to the distribution of the different types of photoreceptors in the retina, visual acuity (i.e., the ability to discriminate two stimuli separated in space at high contrast relative to the background) falls off rapidly as a function of eccentricity (i.e., the distance from the center of sight in degrees of visual angle). Consequently, reduced acuity outside the central retina means that the direction of gaze has to frequently be moved to different positions in visual space —by moving our eyes— in order to perceive the visual world clearly (Kniestedt & Stamper, 2003).

Photoreceptors in the retina are only the first step in a complex neural image processing network. Retinal signals are transmitted via a bundle of axons (the optic nerve) through the optic chiasm -where many nerve fibers decussate (cross from left to right)- to bilateral thalamic relay stations called the lateral geniculate nuclei (LGN). On each side of the brain, information is carried from the LGN to the primary visual cortex via the optic radiations (see figure 1.1.1). Interestingly, the information flow throughout the visual processing network retains the spatial relations of the matrix of photoreceptors. Areas of the brain receiving retinal input are therefore labeled as topographically organized structures. Because the geometric structure of the retina reflects the geometric structure of the external visual world as a two-dimensional planar projection, these brain regions are known to contain retinotopic "visual field maps". A retinotopic area, is a region of the brain that contains a representation of all or most of the contralateral visual hemifield in each hemisphere (Tootell, Hadjikhani, et al., 1998; Wandell et al., 2005).

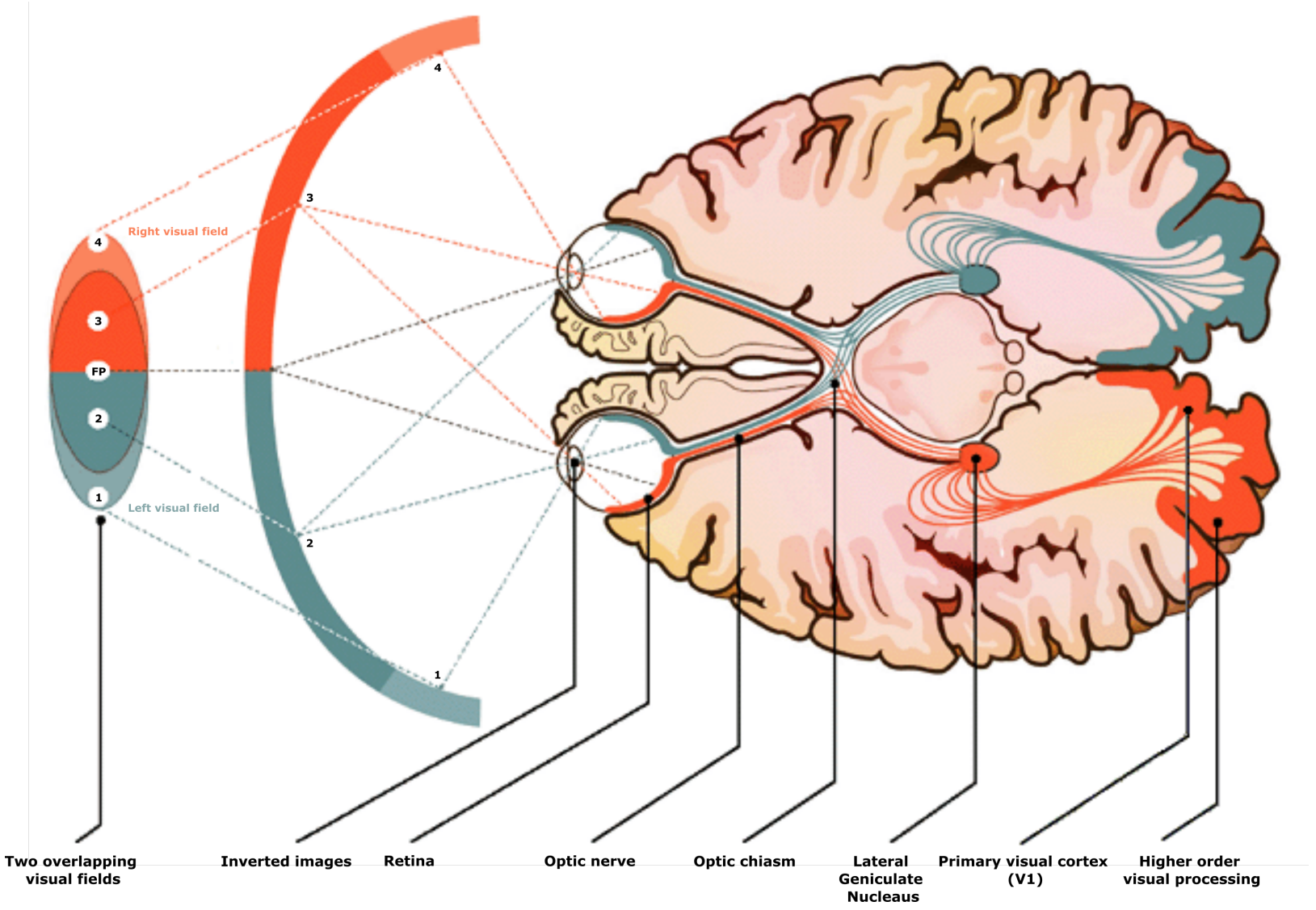



**Figure 1.1.1: simplified depiction of the human visual pathway.** Two overlapping scenes (captured by each eye) are focused and inverted by each eye's lens. The photoreceptors on the retina at the back of the eye convert light into electrical-chemical signals. The optic nerve connects neurons at the back of the retina to the midbrain. At the optic chiasm fibers from each eye cross over, before connecting to the lateral geniculate nucleus. Neuronal fibers connect LGN to the primary visual cortex (V1), where information from the left and right eyes are integrated into a single image. Finally, signals from V1 neurons spread to various areas, including those responsive to color, motion and category-specific stimuli. Figure adapted from (DeWitt, 2013).

The primary visual cortex, also known as striate cortex or area V1, is the first area in the cortex to receive visual inputs from the thalamus via axons that terminate in cortical layer IV. This dense input layer (also called stria of Gennari) makes V1 detectable in myelin-stained cross sections of the occipital lobe. The stria of Gennari runs parallel to the surface of the cerebral cortex on the banks of the calcarine sulcus (CS) and lead to the cytoarchitectonic name striate (which means 'striped') cortex. Primary visual cortex has six layers and begins the cortical encoding of visual features like luminance, spatial frequency, orientation, and motion. Neurons in layers I and V of the striate cortex project to extrastriate visual cortical areas (the human homologues of areas V2, V3, V4 located adjacent to V1). Studies recording single cell responses in the retina and LGN have shown that these regions respond best to small patches of light, while cells in V1 are sensitive to patterns of visual stimulation, such as edges. Higher up the visual system, in areas like V4 and occipitotemporal-cortex, the optimal stimulus becomes way more complex, in the form of shapes or even faces (Purves et al., 2008).

### 1.1.1 Retinotopic organization

The brain is partly shaped by topographic maps to extract higher order representations of the external world. Hubel and Wiesel (1959, 1968) were the first to discover receptive fields (RF) in striate cortex by presenting bars of light to cats and monkeys while recording responses of single cortical cells. Single neurons fired when presented with a bar of light, but only if that bar is shown in a specific region of space. This specific region is called the neuron's receptive field (RF). As an interrelated network, the neurons in retinotopically organized regions form an orderly mapping between an external dimension and the neuronal representation of that dimension. Over the years, neurons in visual cortex were also shown to be specifically tuned to different shapes, orientation, movement directions and ocularity of presented stimuli (Spillmann, 2014). With the onset of functional magnetic resonance imaging (fMRI), (population) receptive fields could also be quantified non-invasively in humans (Sereno et al., 1995) (see Visual responses measured by functional magnetic resonance imaging (fMRI)).

Two other important features of topographic organization in striate and extrastriate cortex are cortical magnification and receptive field size (Figure 1.1.2). Cortical magnification is the phenomenon in which the central visual field is processed by more neurons compared to peripheral parts. The cortical magnification factor (CMF) is measured as the distance along the cortex between two

points representing visual field positions apart by one visual degree (Daniel and Whitteridge, 1961). Hubel and Wiesel (1974) famously reported (in primates) a relationship between RF size and CMF in V1, where RF size decreases with increasing CMF. Our perception of visual space in the fovea is enriched with visual detail and might therefore require more complex neural processing. There is a consensus that more cortical (or subcortical) space is required for more complex neural processing, at least in sensory systems. Corresponding to CMF, receptive fields generally become larger with eccentricity, but RF size also increases along the visual system; from small receptive fields for cells in LGN, to receptive fields that may encompass an entire hemifield in the temporal lobe. Retinotopic maps of visual cortex are frequently visualized on inflated or flattened cortical maps, as a large portion of the cortical surface is buried within folds of the brain. At the border between adjacent early visual areas, the local cortical representation (polar angle) of the visual field is inverted at the horizontal and vertical meridians. This inversion is often called the "visual field sign" and signals the start of the mirror image (turning point of polar angle) of the visual field on the cortical surface (Sereno et al., 1994).

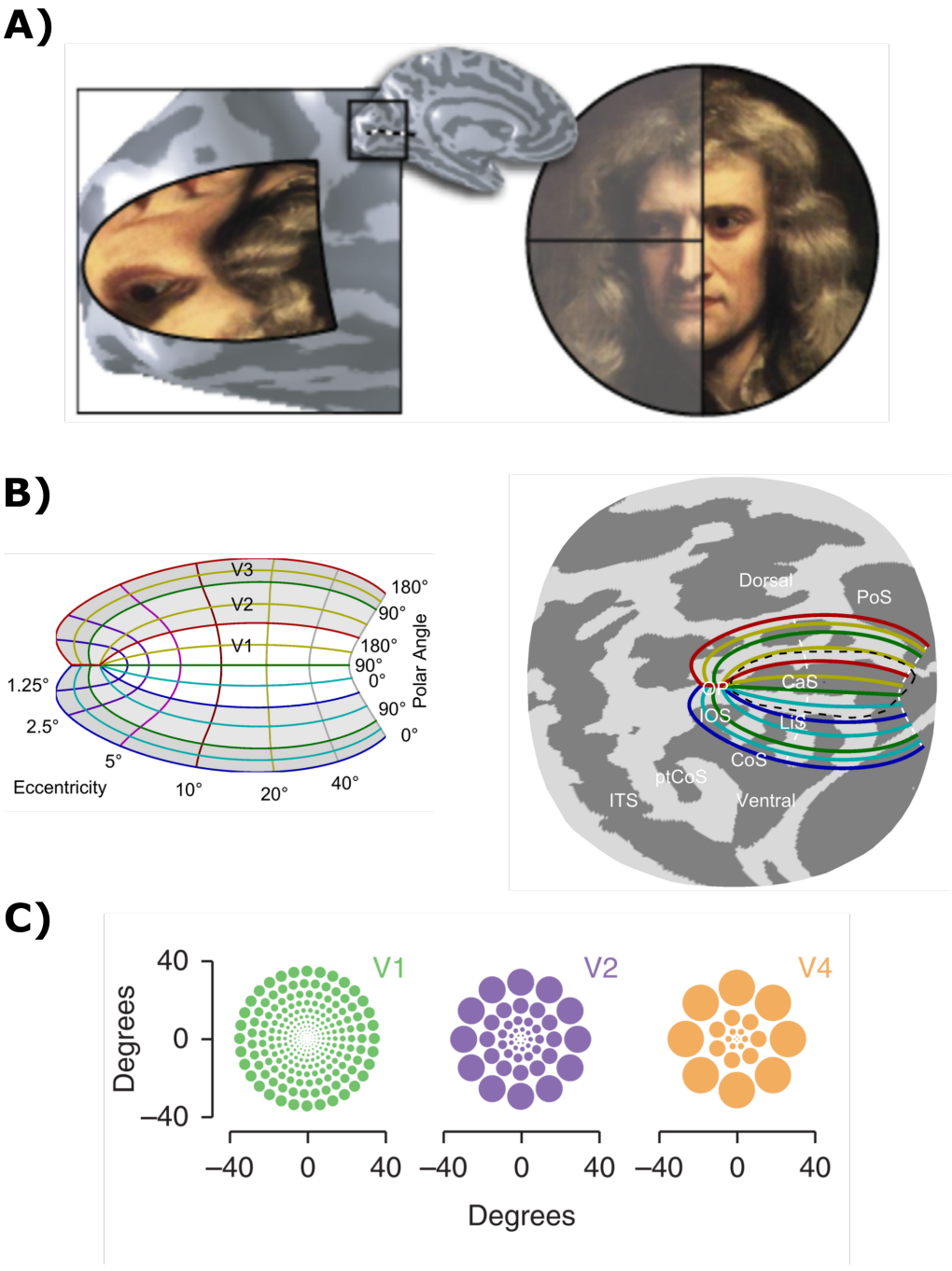



**Figure 1.1.2: principles of retinotopic organization in striate and extrastriate cortex.** A) Cortical magnification: the fovea is greatly overrepresented on the cortical surface. In addition, the image projected on the cortex corresponds to the contralateral visual field and is upside down. B) On the left, the cortical map is expressed in polar coordinates (polar angle and eccentricity). V1, V2, and V3 are colored white, light gray, and dark gray, respectively. On the right, the V1 border and polar angle lines are plotted on a flattened 2D surface. In addition, the Calcarine Sulcus (CaS), the Parietal-occipital Sulcus (PoS), the Lingual sulcus (LiS), the Inferior Occipital Sulcus (IOS), the Collateral Sulcus (CoS), the posterior Collateral Sulcus (ptCoS), the Inferior Temporal Sulcus (ITS), and the Occipital Pole (OP) are shown. C) Tiling of receptive field size increases along the visual hierarchy. Images A to C were adopted from Wandell et al. (2007), Benson et al (2014) and Freeman and Simoncelli (2011), respectively.

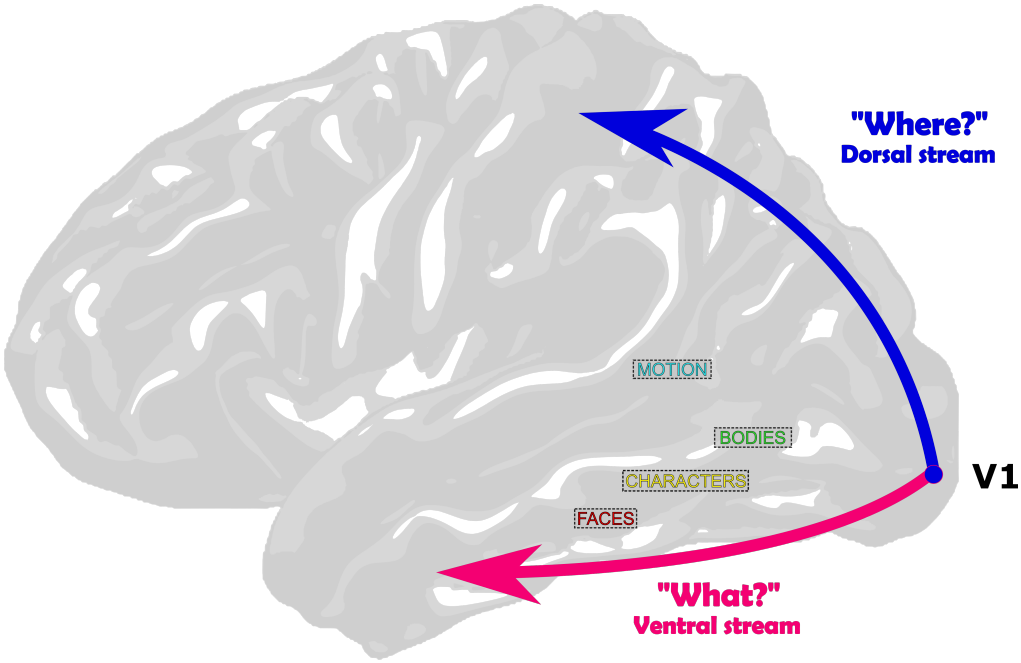



**Figure 1.1.3: association areas beyond V1.** The ventral "what" pathway terminates in the inferior temporal cortex and contains domain-specific modules with neuronal populations that respond selectively to category-specific stimuli. The dorsal "where" pathway terminates in the posterior parietal cortex.

Moving beyond early levels of the visual hierarchy, two types of visual information are integrated for the control of action, namely, "what" can be the subject of action and "where" the specific object among multiple ones is located in a specific place at a given time. Leslie Ungerleider and Mortimer Mishkin (1982) hypothesized that the brain contains two pathways which extract these fundamentally different types of information. In this simplified view, the ventral stream is specialized for object perception and recognition (i.e., what am I looking at?), while the dorsal stream is specialized for spatial perception (i.e., where is this object?), see figure 1.1.3. More recent evidence indicates that the separation of the two pathways is not limited to vision, but also extends to the auditory system (Lomber & Malhotra, 2008). It is hypothesized that there are "domain-specific modules" in human visual cortex which code specifically for objects such as faces, places, and body parts. Importantly, these modules perform computations specific to these classes of stimuli. A more detailed description of the functional organization of object representations in visual cortex and their robustness across individuals is described in chapter 2.

## 1.2 The brain as an image generator

*And in the castle there were three,*
*That one of them was blind and might not see*
*But it was with the eyes of his mind,*
*With which all men see after they have become blind*

— Chaucer
1387

The aforementioned extract dates back to Chaucer's Man of Law's Tale (Canterbury Tales) and is the first appearance of "the mind's eye" in English texts. Mental imagery (also known as 'seeing with the mind's eye') is the amazing ability that allows us to experience (or rather re-experience) objects, fears and pleasures, even travel through time and space, and test different virtual worlds and ideas. A modern definition of mental imagery refers to mental representations and the accompanying experience of sensory information without a direct external stimulus. Such representations are generated from memories and lead one to re-experience a version of the original images or some new combination of stimuli (Pearson, Naselaris, & Holmes, 2015).

Historically, there has been a long-standing debate about the nature and mechanisms of how the brain generates visual mental images. More recent cognitive scientists and philosophers argue either for an *analog* (pictorial), *propositional* (descriptive), or hybrid theory of imagery. On the analog side of the debate, mental representations are regarded as picture-like and retain the intrinsic spatial representational properties of seen images. In contrast, on the propositional side, relevant mental representations are deemed to be more like linguistic descriptions of visual scenes without inherently spatial properties of their own. However, some argue that based on recent evidence from cognitive neuroscience, the debate has been settled in favor of the analog theory of imagery (Joel Pearson, 2019). Thomas et al. (2020) provides an extensive review of mental imagery debates.

In contrast to the quote from Chaucer's Canterbury Tales, not all blind (or sighted) individuals experience mental images. One of the hallmarks of visual imagery is the large number of subjective reports and the wide range in vividness of individual's imagery. People who report to be entirely unable to voluntarily generate visual images in their mind are called aphantasics. Visual imagery is commonly assessed with the vividness of the visual imagery questionnaire (VVIQ), which measures the subjective vividness of an individual's visual imagery, with questions inquiring about the vividness of images of objects, persons and scenes that have to be rated on a Likert scale. Alternative (less subjective measures) include mental rotation (e.g., manikin and Shepard–Metzler) and binocular rivalry tasks. More recently, a study on congenital aphantasics (Zeman et al., 2015) found that aphantasics were able to solve mental rotation tasks. This suggests that they are able to create mental images needed to solve these tasks, while likely lacking the metacognition, or inability to introspect on their mental imagery. Since a mental rotation task could perhaps be performed merely using spatial (or kinaesthetic) imagery rather than 'low-level' visual object imagery, a new experiment using binocular rivalry was conducted on volunteers with aphantasia (Keogh & Pearson, 2018). Binocular rivalry is a phenomenon of visual perception in which perception alternates between two images presented to each eye. Participants are cued to imagine one of the two images (perceptual priming) and report on which image they perceived. In general, observers are biased to perceive the imagined (primed) image. Interestingly, this perceptual priming effect did not work for aphantasics as it would for a control group. The authors argued that binocular rivalry directly measures sensory priming for the mental image, and bypasses the need for any self-reports on vividness of mental imagery. On the other hand, binocular rivalry might not be the best diagnostic of imagery ability on the individual level, as some individuals with aphantasia showed better-than-chance priming (unlike the group-average).

Recent neuroscientific research suggests a strong link between image processing and image generation in the brain. Voluntary mental imagery might be based on flow of information which resembles the flow of visual perception, but in reverse. In this framework, imagery would be formed by a combination of information retrieved from stored memory. Two recent studies found evidence for this hypothesis; by showing the reversal of effective connectivity from frontal to early visual areas common to both perception and imagery (Dentico et al., 2014), and that increases in top-down connectivity to early visual cortex were associated with the experienced vividness during imagery (Dijkstra et al., 2017). A schematic representation of this hypothesis illustrated in figure 1.2.1.

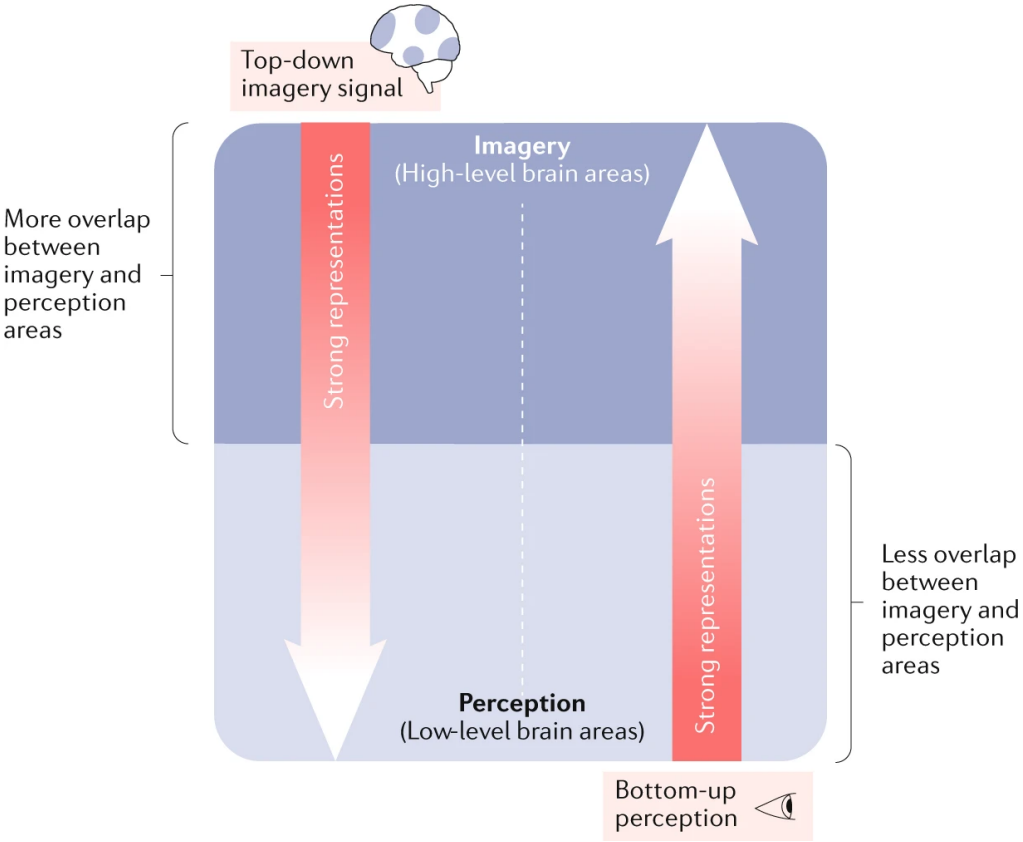



**Figure 1.2.1: brain signals shared between imagery and perception.** There might be a greater overlap in high-level areas (dark blue) than in lower-level areas (light blue) for voluntary imagery and perception. Image adopted from (Joel Pearson, 2019).

Especially the role of primary visual cortex in relation to imagery has been the subject of many neuroimaging studies in the last two decades. Due to differences in imagery tasks, content (simple vs. complex images), and differences in imagery vividness, some research groups found V1 BOLD (blood oxygenation level-dependent; see next section for an introduction to magnetic resonance

imaging) activations during imagery (Formisano et al., 2002; Ishai et al., 2000; Knauff et al., 2000; Sack et al., 2002; Trojano et al., 2000; Wheeler et al., 2000), while others did not (Bihan et al., 1993; Dijkstra et al., 2017; O'Craven & Kanwisher, 2000; Reddy et al., 2010). In fMRI research, neuroscientists predict how features of bottom-up processing of visual input are represented in neural activity (encoding). However, it has been a dream of many neuroscientists to extract the contents of the mind's eye (i.e., mind reading). This decoding problem refers to predicting the stimulus that is being viewed when a particular brain state is being observed inside the MRI scanner (see figure 1.2.2). With new advances in predictive data analysis techniques on conjunction with fMRI, there is now clear evidence that the content of mental imagery can be decoded and reconstructed from relatively low BOLD amplitude responses in early visual areas, including V1, V2 and V3 (see chapter 3 for details).

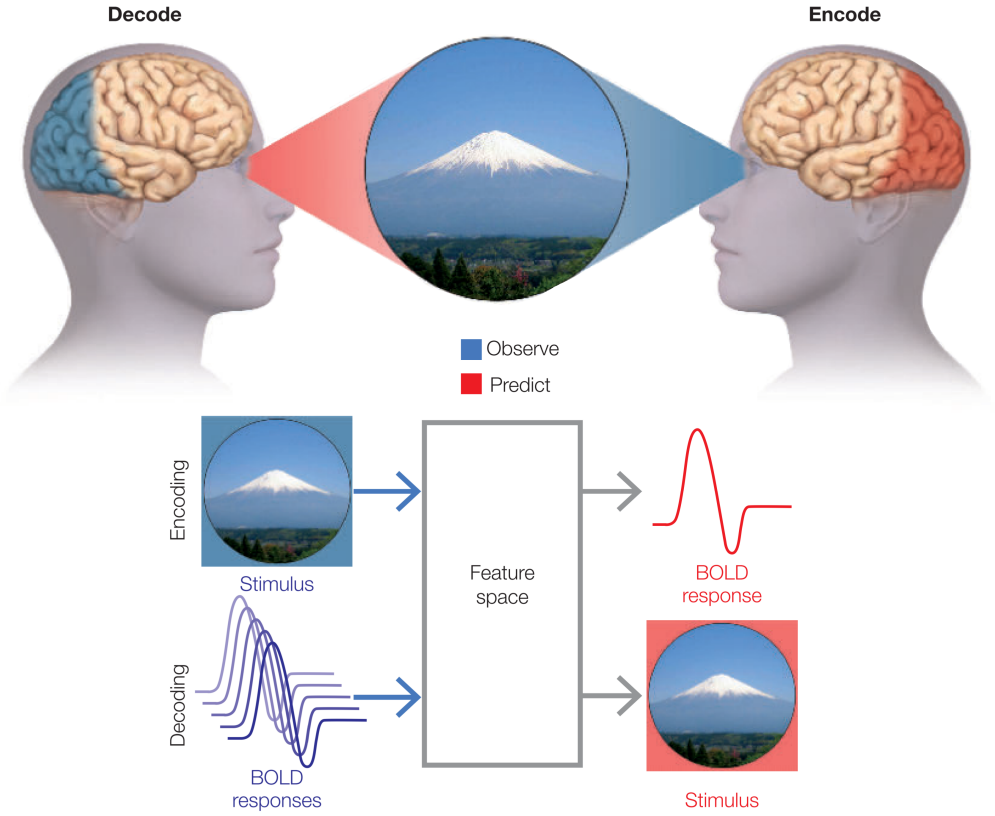



**Figure 1.2.2: encoding and decoding neural activity** – adopted from (Gazzaniga, M. S., Ivry, R. B., & Mangun, 2018).

## 1.3 Visual responses measured by functional magnetic resonance imaging (fMRI)

Magnetic resonance imaging (MRI) is based on the understanding of the nuclear magnetic resonance (NMR) exhibited by atomic nuclei within liquid and solid substances. The most commonly studied nuclei are hydrogen as they are not only highly abundant but also highly NMR-sensitive. Due to their odd number of protons, hydrogen atoms exhibit an intrinsic magnetic spin (Kuperman, 2000). A powerful external magnetic field can be used to align the spins that are otherwise randomly oriented within the water of the tissue being examined. Next, radio frequency pulses are sent to disturb the equilibrium of these spins, which causes a measurable voltage to be induced in a receiver radiofrequency coil. The frequency information contained in the signal from each location in the imaged plane is then converted (using a Fourier transformation) to corresponding intensity levels, which are displayed as shades of gray in an MR-image. Built on the same basic principles of atomic physics, functional magnetic resonance imaging (fMRI) indirectly measures brain activity of the person inside the MRI-machine. Active parts of the brain require more oxygen for glucose to be metabolized during the neural processing of information. This leads to increased blood flow to supply these brain regions with oxygenated blood. Crucially, deoxygenated hemoglobin (in the blood surrounding the active neurons) is weakly magnetic because of the exposed iron in the hemoglobin molecule. As oxygenated and deoxygenated blood behave differently in a magnetic field, it is the contrast between these two that gives rise to the blood oxygen level-dependent (BOLD) contrast (Hashemi et al., 2012). Having explained the basics of fMRI, we return to the retinotopic organization of human visual cortex, which has been subject to many studies by vision scientists who have developed sophisticated fMRI techniques (Wandell & Winawer, 2015).

Typically, retinotopic maps are obtained by presenting stimuli with checkerboard patterns which are slowly shifted around and away from the center of view (see figure 1.3.1A). The neural representation of the checkerboards in the visual field can be measured by tracking neural firing responses at corresponding locations that change over time in synchrony with the presented stimulus. fMRI measures the changes in blood oxygenation level dependent (BOLD) signal that occur after neurons have fired and are resupplied with oxygen. Note that the BOLD signal is influenced by multiple factors, yet it is generally seen as a reliable proxy to neural activity. Furthermore, the spatial resolution and signal-to-noise ratio (SNR) of fMRI images depend on the magnetic field strength

of the MRI machine (Logothetis, 2008; Winawer et al., 2013). Each voxel (a three-dimensional, volumetric, pixel) in an fMRI image contains the aggregated signal of thousands of neurons. Therefore, when inferences are made regarding spatial visual responses using fMRI, we talk about population receptive field (pRF) mapping, as each voxel's visuospatial preference represents the average preference of the neural population within that voxel (Dumoulin & Wandell, 2008).

Population receptive field estimation will reappear in the method sections of all chapters described in this thesis; however, it is important to note that receptive fields can be estimated from BOLD responses in several ways. In 2008, Dumoulin and Wandell (2008) introduced a forward modeling approach to fit the location and size of symmetrical 2D Gaussian pRFs. In this method, traversing bar stimuli are usually used rather than wedges and rings to map out visual space. The forward model minimized the difference between recorded and predicted responses. The responses are multiplied by the stimulus time-course profile and convolved with a hemodynamic response function (HRF) to model neurovascular coupling (Figure 1.3.1B). Since the development of the pRF method, it was shown that suppressive surrounds of pRFs in early visual cortex can also be well captured by a Difference-of-Gaussians (DoG) pRF profile (Zuiderbaan et al., 2012) and that mapping of more anterior extrastriate visual areas benefits from non-linear spatial summation across receptive fields (Britten & Heuer, 1999; Kay et al., 2013; Oleksiak et al., 2011). Both refinements have provided better fits than the originally proposed linear model and incorporate additional factors of visual organization. Finally, pRFs of any shape can alternatively be estimated using model-free pRF mapping methods (Bhat et al., 2021; Lee et al., 2013; Tangermann et al., 2012).

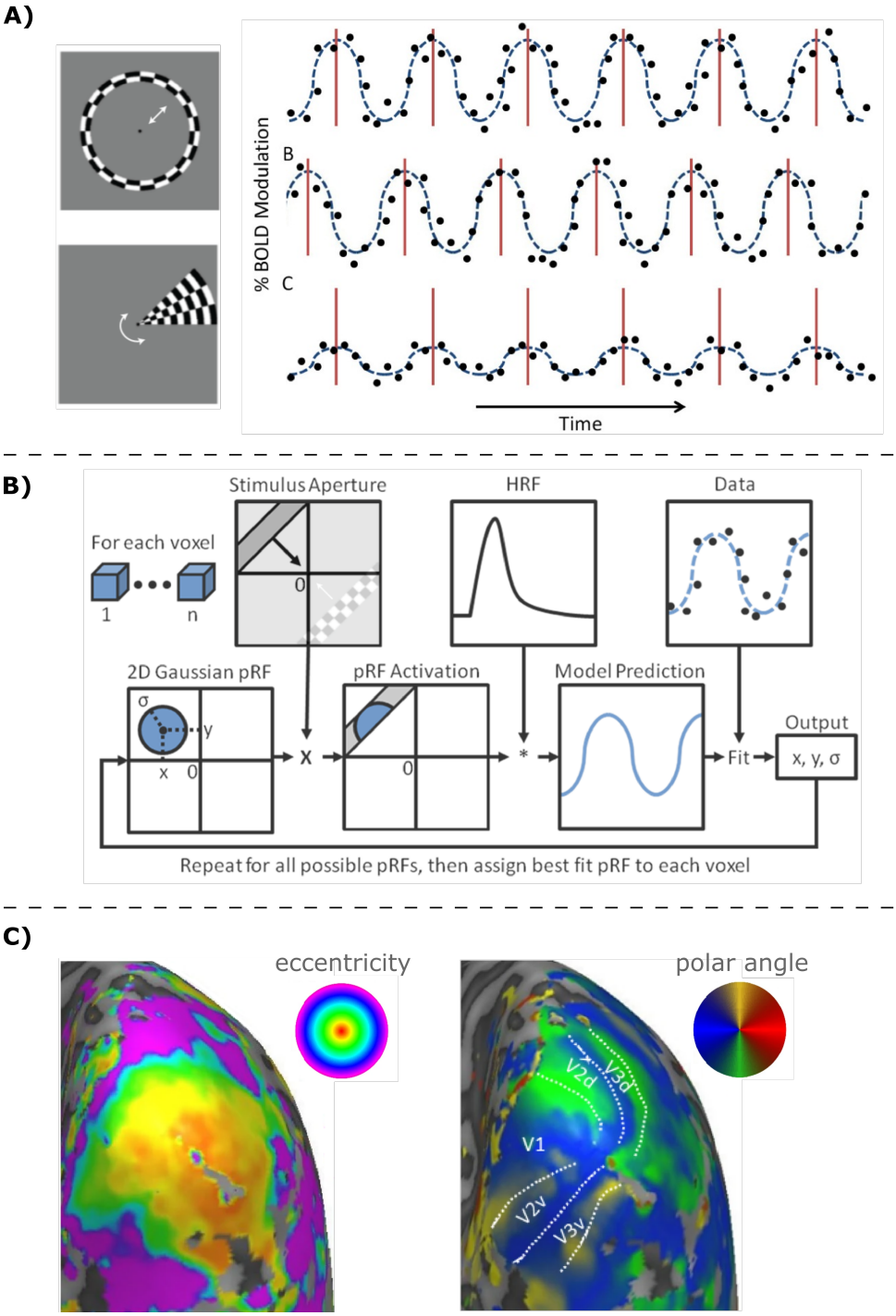



**Figure 1.3.1: population receptive field mapping.** A) Ring and wedge stimuli contract/expand and transverse in phase-encoded paradigms (left). The time-courses of three example voxels show BOLD modulation in phase with the preferred location of the stimulus in the visual field (right). B) A parameterized candidate model of the underlying neuronal population and the stimulus is computed. For each voxel a Gaussian model of the underlying neuronal population is computed with location parameters x, y as the center of the receptive field and  as Gaussian spread. C) Visual field signs are indicated in the polar angle map by the white dotted lines (right). Figures were adapted from (Brewer. & Barton, 2012)

## 1.4 Artificial images in the absence of visual input through brain-computer-interfacing

In chapters 3 and 4 we lay the groundwork for two highly relevant brain-computer-interface (BCI) systems. BCIs monitor brain activity and translate specific signal features that reflect a person's intent into commands that operate a device. Nowadays, BCIs are being deployed for treatment of many patients suffering from cognitive or physical impairments and offer the promise of greatly improving personal autonomy and mobility (Belkacem et al., 2020; Pichiorri & Mattia, 2020). Chapter 3 describes the current gaps in knowledge required for the development of vision-restoring cortical prostheses in human early visual cortex, and how virtual simulations of electrode implantation can bring cortical prostheses a step closer to becoming an effective clinical therapy for the late blind.

Blind persons can potentially benefit greatly from biology-based technology if the malfunctioning parts of the visual system can somehow be replaced. Blindness causes enormous disruptions in physical (McLean et al., 2014; Crews, 2016) and emotional wellbeing (Stelmack, 2001; Hassell et al., 2006) and an alternative treatment is desired for those who struggle to adapt to a live without the visual sense. Since it was discovered that electrical stimulation of the eye could produce spots of light, or phosphenes, in one's visual field (Leroy, 1755; Volta, Alexander, 1832), many have attempted to simulate vision using electrical stimulation of various nodes in the visual pathway with so-called "electrode-based visual prostheses". A traditional visual prosthesis and its components is displayed in figure 1.4.1. The location of the target region along the visual pathway is dictated by the underlying pathophysiology and in most cases involves either the retina, optic nerve, LGN or visual cortex. The advantages and limitations of the different types of prosthetic implants have recently been reviewed by (Farnum & Pelled, 2020). Chapter 3 introduces the current challenges specific to (sub) cortical prosthetic implants and a solution to part of these challenges by simulating distributions of phosphene configurations for specific electrode designs.

**Figure 1.4.1: overview of a wearable cortical visual prosthesis.** A video-feed captured by a camera on the head is processed by a mobile device and converted into interpretable signals. These signals are then wirelessly transferred to the electrode arrays that interface directly with the neural tissue. Image adopted from (Li, 2013).

## Chapter 2

# A probabilistic functional parcellation of human occipito-temporal cortex

Human visual cortex contains many retinotopic and category-specific regions. These brain regions have been the focus of a large body of functional MRI research, significantly expanding our understanding of visual processing. As studying these regions requires accurate localization of their cortical location, researchers perform functional localizer scans to identify these regions in each individual. However, it not always possible to conduct these localizer scans. Here, we developed and validated a functional region of interest atlas of early visual and category-selective regions in human ventral and lateral occipito-temporal cortex. Results show that for the majority of fROIs, cortex-based alignment results in lower between-subject variability compared to non-linear volumetric alignment. Furthermore, we demonstrate that (1) the atlas accurately predicts the location of an independent dataset of ventral temporal cortex ROIs and other atlases of place-selectivity, motion-selectivity, and retinotopy. Next, (2) we show that the majority of voxel within our atlas are responding mostly to the labelled category in a left-out subject cross-validation, demonstrating the utility of this atlas. The functional atlas is publicly available (`download.brainvoyager.com/data/visfAtlas.zip`) and can help identify the location of these regions in healthy subjects as well as populations (e.g. blind people, infants) in which functional localizers cannot be run.

## 2.1 Introduction

Human visual cortex extends from the occipital lobe to the posterior parietal and temporal lobes, containing more than two dozen visual areas. Early and intermediate visual areas are typically defined by their representation of the visual field, where each visual area contains a topographic (retinotopic; Engel et al., 1994; Sereno et al., 1995) representation of the entire visual field across both hemispheres (referred to as a visual field map, Arcaro et al., 2009; Wandell et al., 2005; Wandell and Winawer, 2011; Wang et al., 2014). Higher visual areas are typically defined by their function and stimulus selectivity rather than the representation of the visual field. This includes preference to visual attributes such as motion (M I Sereno et al., 1995), shape (Grill-Spector et al., 1998; Malach et al., 1995), or color (Lafer-Sousa et al., 2016), as well as preference for certain visual stimuli over others. A well-documented characteristic of higher-level regions in ventral and lateral occipito-temporal cortex are regions that respond preferentially to ecologically-relevant stimuli such as faces (Kanwisher et al., 1997), places (Aguirre et al., 1998; Epstein & Kanwisher, 1998), bodies (Downing et al., 2001; Peelen & Downing, 2005), and words (Cohen et al., 2000) compared to other stimuli. These regions are referred to as category-selective regions.

To elucidate neural mechanisms of visual processing and perception, a central goal in neuroscience is to understand the function and computation in each of these regions. Indeed, tens of thousands of papers have investigated visual processing in specific visual areas, from visual field maps to category-selective regions. For example, according to google scholar, more than 7575 studies cite the study that discovered the fusiform face area (Kanwisher et al., 1997). The first step in this scientific endeavor is the identification of each visual region in each brain. The standard approach is to perform an independent scan, such as retinotopic mapping (S A Engel et al., 1997) or a functional localizer scan, in each individual to identify the relevant region of interest (ROI, Kanwisher et al. 1997; Saxe et al. 2006). Then, the main experiment of interest is performed, and the data are analyzed within the ROI identified using the independent scans. The ROI approach is advantageous for four reasons: (1) it allows hypothesis driven comparisons of signals within independently-defined regions of interest across many different conditions, (2) it increases statistical sensitivity in multi-subject analyses (Nieto-Castañón & Fedorenko, 2012), (3) it reduces the number of multiple comparisons present in whole-brain analyses (Saxe et al., 2006), and (4) it identifies ROIs in each participant's native brain space.

Nevertheless, there are also several limitations to the independent localizer approach. First, it is not always possible to obtain an independent localizer scan. This is especially the case in patient populations, for example in the congenitally blind (Bedny et al., 2011; Mahon et al., 2009; Striem-Amit, Cohen, et al., 2012; van den Hurk et al., 2017) or individuals with visual agnosia/prosopagnosia (Barton, 2008; Gilaie-Dotan et al., 2009; Schiltz & Rossion, 2006; Sorger et al., 2007; Steeves et al., 2006; Susilo et al., 2015). Second, performing a localizer scan before each experiment is costly in terms of scanning time, as well as mental effort and attention resources of the participant. The latter can result in fatigue during the main experiment of interest, leading to lower quality data. Third, as localizer scans are typically conducted in a subject-specific manner, and researchers vary in the manner they define the ROIs (e.g. whether smoothing was employed, if they use anatomical constraints, what thresholding methods were employed), it is hard to assess variability between participants and across studies.

To overcome these limitations, progress in the field of cognitive neuroscience has led to the development of cortical atlases, which allow localization of visual areas in new subjects by leveraging ROI data from an independent set of typical participants (Frost and Goebel 2012; ventral-temporal cortex category selectivity: Julian et al. 2012; Engell and McCarthy 2013; Zhen et al. 2017a; Weiner et al. 2018; visual field maps: Benson et al. 2012; Wang et al. 2014; motion-selective hMT: Huang et al. 2019; multimodal parcellation: Glasser et al. 2016; cytoarchitectonic parcellation of ventral visual cortex: Rosenke et al. 2018). In addition to providing independent means to identify ROIs, this approach enables quantification of between-subject variability. Further, the process of atlas creation also enables measuring the prevalence and robustness of each ROI across participants. Presently, atlases for the human visual system include atlases of visual field maps (Benson et al., 2012, 2014; Benson & Winawer, 2018; L Wang et al., 2014), and atlases of cytoarchitectonically-defined areas (Amunts et al., 2000; Caspers et al., 2013; Kujovic et al., 2013; Lorenz et al., 2015; Rosenke et al., 2017b; Rottschy et al., 2007). However, presently, there is no atlas of the full extent of visual category-selective regions in occipito-temporal cortex, or atlases that include both visual regions that are defined retinotopically as well as from stimulus selectivity.

To fill this gap in knowledge, in the present study we: (a) develop a functional atlas of category-selective visual cortex, (b) quantify inter-subject variability of category-selective regions in visual cortex, and (c) validate our approach by using the same procedure to define retinotopic regions and hMT+, which also allows us to compare our definitions to existing atlases. To generate the visual functional atlas (visfAtlas), 19 participants (10 female) underwent the following functional scans: (i) a localizer experiment to identify word, body, face, body, and place-selective regions in lateral occiptio-temporal (LOTC) and ventral temporal cortex (VTC), (ii) a visual field mapping experiment to delineate early visual cortex (V1-V3), and (ii) a motion localizer to identify hMT+. We identified each ROI in each participant's brain. We then used a leave-one -out cross-validation (LOOCV) approach and two anatomical alignment methods: (i) nonlinear volume-based alignment (NVA) and (ii) cortex-based alignment (CBA), to evaluate the accuracy of the atlas in predicting ROIs in new participants. The resulting visfAtlas is available with this paper in BrainVoyager (www.brainvoyager.com) and FreeSurfer (www.surfer.nmr.mgh.harvard.edu) file formats for cortical surface analyses, as well as in nifti format for volumetric analysis (download.brainvoyager.com/data/visfAtlas.zip).

## 2.2 Materials and Methods

### 2.2.1 Participants

To obtain functional data, a total number of 20 participants (average age 30 ± 6.61) were recruited at Maastricht University but one subject's functional MRI (fMRI) scans were excluded from further analysis due to self-reported lack of attention on the stimuli and intermittent sleep. Two participants were left-handed, and the sample consisted of 10 women and 9 men. All participants were healthy with no history of neurological disease and had normal or corrected-to-normal vision. Written consent was obtained from each subject prior to scanning. All procedures were conducted with approval from the local Ethical Committee of the Faculty of Psychology and Neuroscience.

### 2.2.2 Data acquisition

Participants underwent one scanning session of 1 hour at a 3T Siemens Prisma Fit (Erlangen, Germany). First, a whole brain, high resolution T1-weighted scan (MPRAGE) was acquired (repetition time/echo time = 2250/2.21 ms, flip angle = 9 , field of view = 256 x 256 mm, number of slices = 192, 1 mm isovoxel resolution). Following that, six functional runs were acquired using a T2*-weighted sequence with the following parameters: repetition time/echo time = 2000/30 ms, flip angle = 77 , field of view = 200 x 200 mm, number of slices = 35, slice thickness = 2 mm, in-plane resolution = 2 × 2 mm. fMRI included (i) three scans of the functional localizer (fLoc; Stigliani et al. 2015) (ii) two scan of an hMT+ localizer, and (iii) one scan of retinotopic mapping. Maximal diameter of the visual stimuli ranged from 30-36 in the fMRI experiments. Details for each localizer can be found in the section below.

### 2.2.3 Visual localizers

**Category-selective regions in ventral temporal cortex and lateral occipito-temporal cortex**

In order to identify category-selective regions that respond preferentially to characters (pseudowords, numbers), bodies (whole bodies, limbs), places (houses, corridors) , faces (child, adult) and objects (cars, instruments), we used stimuli included in the fLoc functional localizer package (Stigliani et al., 2015). Eight stimuli of one of the five categories were presented in each miniblock design, each miniblock holding a duration of 4 seconds. To assure participant's attention, they were asked to perform an Oddball task, indicating with a button press

when they saw a scrambled image instead of one of the categories. Each run consisted of 150 volumes, and each subject underwent three runs.

#### hMT+

To localize the motion-selective area in middle temporal cortex (hMT+, Dumoulin et al., 2000; Zeki et al., 1991), we used stimuli as in Emmerling et al. (2016) and Zimmermann et al. (2011), which were based on Huk et al. (2002). During the first 5 volumes participants were presented with a fixation dot in the center of the screen. In the following blocks, moving and stationary dot patterns alternated while the participants fixated on the fixation dot at the center of the screen. Moving dot blocks were 18 seconds long, while stationary blocks had a duration of 10 seconds. The active screen filled with dots was circular. In total, each run consisted of 12 blocks of moving dots and 12 blocks of stationary dots. Black dots on a gray background traveled towards and away from the fixation point (speed=1 pixel per frame, dot size=12 pixels, number of dots=70). In different blocks, dots were presented either in the center of the screen, in the left visual hemifield, or in the right visual hemifield. Stationary blocks were in the same three locations. The order of blocks was fixed (center moving, center static, left moving, left static, right moving, right static). Each subject underwent two hMT+ localizer runs.

#### Early visual cortex

We ran one visual retinotopic mapping run that consisted of 304 volumes (TR = 2s). In the first 8 volumes a fixation dot was presented, followed by a high-contrast moving bar stimulus (1.33° wide) revealing a flickering checkerboard pattern (10 Hz). The checkerboard pattern varied in orientation and position for 288 volumes, concluding the run with 8 volumes of fixation dot presentation. The fixation was presented during the entire run and changed color at random time intervals. To keep participants' motivation and attention they were asked to count these color changes. The bar stimulus moved across the visual field in 12 discrete steps and remained at each position for 1 TR. The 12 different stimulus positions were randomized within each bar orientation. Each combination of orientation (4) and direction (2) represented one cycle. These eight different cycles were repeated three times in random order throughout the run (Senden et al., 2014).

### 2.2.4 Preprocessing

If not stated otherwise, data were preprocessed and analyzed using BrainVoyager 20.6 (Brain Innovation, Maastricht, The Netherlands). Anatomical data were inhomogeneity corrected and transformed to Talairach space (TAL, Talairach and Tournoux, 1988) by identifying the anterior commissure (AC) and posterior commissure (PC) and fitting the data to TAL space. Functional data were slice scan time corrected, motion corrected with intra-run alignment to the first functional run to account for movement between runs, and high-pass filtered (3 cycles). Next, the preprocessed functional data were co-registered to the inhomogeneity corrected anatomical image. Using the anatomical transformation files, all functional runs were normalized to TAL space. Based on the normalized anatomical data, we segmented the grey-white matter boundary for each brain and created a cortical surface. Next, the volumetric functional data were sampled on the cortical surface incorporating data from -1 to +3 mm along the vertex normals. Ultimately, we computed two general linear models (GLM), one for the three localizer runs for category-selective regions in ventral temporal cortex, and one for the hMT+ localization.

### 2.2.5 Regions of interest

All ROIs where manually defined in individual subjects on their cortical surface reconstruction in BrainVoyager. For volumetric alignment and atlas generation, surface regions were transformed to volumetric regions by expanding them (-1 to +2 mm) along the vertex normals of the white-gray matter boundary. The final atlas includes all regions that could be defined in more than 50% of the subjects (N10, see Table 2.1 for number of subjects per atlas ROI).

### 2.2.6 Retinotopic areas in occipital cortex

Visual field maps were determined for each subject based on an isotropic Gaussian population receptive Field (pRF) model (Dumoulin & Wandell, 2008; Senden et al., 2014). The obtained pRF maps estimating the location and size of a voxel pRF were used to calculate eccentricity and polar angle maps. The polar angle maps were projected onto inflated cortical surface reconstructions and used to define six topographic regions in occipital cortex (V1d, V2d, V3d and V1v, V2v, V3v, where d = dorsal and v = ventral) by identifying the reversals in polar angle representation at the lower vertical meridian (LVM), upper vertical meridian (UVM) or horizontal meridian (HM; DeYoe et al., 1996; Engel et al., 1997; Sereno et al., 1995). We did not define visual areas beyond V3d and V3v

as visual field maps using the single run retinotopic mapping paradigm were noisy beyond V3.

### 2.2.7 Ventral and lateral category-selective areas

Each category (e.g. faces) was contrasted against the mean of all other categories to identify vertices that displayed a preference for the given category. Then we followed a two-step approach to define ROIs: First, for all categories we selected a statistical threshold of t = 3 for a whole brain map. Based on the thresholded activation map we identified ROIs in anatomically plausible locations (see details for each region below). Furthermore, in the case of an activation cluster transitioning into an adjacent one of the same visual category, we divided those clusters into separate ROIs by following the spatial gradient of t-values and separating the two areas at the lowest t-value. Based on insufficient activation pattern found for the 'objects' category, we dismissed that category from further analysis.

Face-selective regions (faces > all others) were identified in the mid lateral fusiform gyrus (mFus) and posterior lateral fusiform gyrus (pFus), which correspond to the fusiform face area (Kanwisher et al., 1997), as well as on the inferior occipital gyrus (IOG). Body-selective regions (bodies > all others) were observed in ventral temporal cortex on the occipital temporal sulcus (OTS), also known as fusiform body area (FBA, Peelen et al., 2009; Schwarzlose, 2005) and in lateral occipital cortex. There, we identified three different regions (Weiner & Grill-Spector, 2011) together forming the extrastriate body area (Downing et al., 2001), one anterior of hMT+ on the middle temporal gyrus (MTG), one posterior of hMT+ on the lateral occipital sulcus (LOS), and one ventral to hMT+, on the inferior temporal gyrus (ITG). Place-selective regions (places > all others) were observed in ventral temporal cortex on the collateral sulcus (CoS), corresponding to the parahippocampal place area (PPA, Epstein and Kanwisher, 1998), and on the transverse occipital sulcus (TOS, Hasson et al., 2003). Character-selective regions (characters > all others) were identified in the posterior occipital temporal sulcus (pOTS) and a left-lateralized region in the mid occipital temporal sulcus (mOTS). Furthermore, we identified one character-selective regions in the inferior occipital sulcus (IOS). In the following, we will refer to each ROI by its anatomical nomenclature, as described in Stigliani et al. (2015). For reference, Table 2.1 provides an overview about each ROI's anatomical as well as functional name.

**Table 2.1: nomenclature for functional regions-of-interest (fROIs) and number of subjects per fROI.** Each category-selective functional activation cluster can be described by functional category or anatomical location. In this article we describe category-selective ROIs using the anatomical nomenclature and provide this table as a reference. Functional abbreviations are as followed: FFA: fusiform-face area, FBA: fusiform-body area, EBA: extrastriate body area, VWFA: visual word form area, PPA: parahippocampal place area, hMT: human middle-temporal (cortex). Number of identified ROIs per hemisphere (N LH/N RH): Due to individual-subject variability and using a strict statistical threshold (t>3, vertex level), not every fROI was identified in all participants in both hemispheres. fROIs that were defined in more than half the participants (N10) were included in the atlas. Areas that were not included are indicated in gray subject counts. The last column, N, indicates the number of subjects in which a given fROI could be identified in at least one hemisphere. Abbreviations: LH: left hemisphere, RH: right hemisphere.

| ROI | Functional nomenclature | N (LH) | N (RH) | N |
|---|---|---|---|---|
| mFus - faces | FFA-2 | 13 | 15 | 18 |
| pFus - faces | FFA-1 | 17 | 15 | 19 |
| IOG - faces | - | 15 | 15 | 18 |
| OTS - bodies | FBA | 14 | 13 | 17 |
| ITG - bodies | EBA | 17 | 17 | 19 |
| MTG - bodies | EBA | 16 | 15 | 18 |
| LOS - bodies | EBA | 15 | 16 | 19 |
| pOTS -characters | VWFA-1 | 16 | 5 | 17 |
| IOS - characters | - | 11 | 1 | 11 |
| TOS - places | - | 9 | 12 | 13 |
| CoS - places | PPA | 18 | 19 | 19 |
| hMT - motion | hMT | 18 | 16 | 19 |
| V1d | - | 19 | 19 | 19 |
| V2d | - | 19 | 19 | 19 |
| V3d | - | 14 | 17 | 17 |
| V1v | - | 19 | 19 | 19 |
| V2v | - | 19 | 19 | 19 |
| V3v | - | 19 | 19 | 19 |

#### hMT+

Motion selective regions were identified by contrasting left, right and central visual field motion conditions vs. the equivalent stationary conditions and using a thresholded statistical map with a minimum t-value of 3. Two subjects only showed functional activation for the contrasts at a t-value of 2.5 in one hemisphere, which we allowed for these subjects. hMT+ was consistently located in the posterior inferior-temporal sulcus (pITS).

#### Visual functional atlas (visfAtlas) generation

After ROIs were defined for each subject in each subject's space, we utilized two normalization techniques to bring the data into a common space: (1) nonlinear volumetric alignment (NVA) for volume and (2) cortex-based alignment (CBA) for surface space. Furthermore, as it is common that not every ROI can be identified in each of the subjects, we decided that an ROI had to be present in more than 50% of the subjects (N > 10) to be considered for a group atlas. The ROIs which were ultimately used for the group atlases and in how many subjects they were defined can be found in Table 2.1.

#### Nonlinear-volumetric alignment (NVA)

First, surface regions that were defined on each subject's cortical surface were mapped to volumetric regions by expanding them (-1 to +2 mm) along each vertex normal of the white-gray matter boundary. Second, the volumetric regions were transformed back to native ACPC space. Next, the individual brains were registered to the MNI152 group average brain using the Advanced Normalization Tools (ANTS; https://sourceforge.net/projects/advants/). Finally, the resulting nonlinear transformation matrices were used to warp the functionally-defined regions of interest (fROIs) into the same orientation and reference frame. The specific code for the affine volume registration and nonlinear transformation can be found here: download.brainvoyager.com/data/visfAtlas.zip. The resulting NVA-aligned regions were further processed in NifTi format using MATLAB 2014b and 2019a (www.mathworks.com), see details below.

#### Cortex-based alignment (CBA)

To generate a surface group average brain of the subjects, we used cortex-based alignment (CBA) to generate a dynamic average (subsequently called BVaverage, publicly available at download.brainvoyager.com/data/visfAtlas.zip

and usable as surface template for future studies). CBA was performed for both hemispheres separately after inflation to a sphere with overlaid curvature information at various levels of resolution (Frost & Goebel, 2012; R Goebel et al., 2006). First, during a rigid alignment, the spheres of each subject's hemisphere was rotated along three dimensions to best match the curvature pattern of a randomly chosen target hemisphere. The lower the variability between the two folding curvature patterns, the better the fit after rigid sphere rotation. Following the rigid alignment for all subjects, a non-rigid CBA was performed. Curvature patterns of each subject were used in four different levels of anatomical detail. Starting from low anatomical detail, each subject's hemisphere was aligned to a group average out of all subjects. During this process, the group average was dynamically updated to most accurately average all hemispheres. This sequence was repeated for all levels of curvature detail, until the group average was updated based on the highest level of anatomical detail per subject. During the alignment, we (1) derived a group average for each hemisphere (BVaverage), as well as (2) a transformation indicating for each vertex on a single-subject cortical surface where it maps to on the group average. These transformation files were then used to map each individual subject's fROIs to the BVaverage.

**Probabilistic maps for occipitotemporal cortex in volume and surface space**

We generated probabilistic maps of all regions after NVA as well as CBA, where each of the following was done in both group spaces: after individual subject fROIs were projected to the MNI152 and BVaverage, respectively, each group fROI was defined. For each voxel/vertex of a group fROI, the number of subjects sharing that voxel/vertex in the fROI was divided by the total number of subjects of the fROI ($voxel probability = \frac{number\ of\ subjects\ sharing\ voxel\ /\ vertex}{total\ number\ of\ subjects\ in\ fROI}$). Thus, a value of 0 at a vertex in the group fROI indicates a vertex did not belong to that fROI in any subject, a value of .5 means that it belonged to the fROI in half the subjects, a value of 1 indicates that it belonged to that functional region in the entire study population (Fig. 5.4.1).

**Cross-validated predictability estimation and atlas generation**

One interesting feature of those fROIs is the possibility to serve as a prior to estimate the localization of corresponding ROIs in a new subject's brain, eliminating the need for a dedicated localizer run in the new subject. To allow for a probabilistic estimate to find this region in a new subject, we performed an exhaustive leave-1-subject-out cross-validation analysis after the volumetric

(NVA) as well as surface (CBA) alignment to establish how well our atlas can predict fROIs in new subjects. For each fold of the LOOCV, we generated a group probabilistic fROI (G) and a left-out subject's individual fROI (I). We estimated the predictability of the group probabilistic fROI by calculating the Dice coefficient (DSC), a measure of similarity of two samples:

$$dsc = \frac{2|I \cap G|}{|I| + |G|} \tag{2.1}$$

A Dice coefficient of zero indicates no predictability and a Dice coefficient of 1 indicates perfect predictability. As we did in previous work (Rosenke et al., 2017b), we applied different threshold levels to the group probabilistic fROI (G) to predict the location of the left-out-subject (Fig. 2.3.2). That means we created a liberal group probabilistic fROI including each vertex that was present in at least 1 subject. Then we sequentially increased the threshold up to the most conservative threshold where all subjects had to share a voxel/vertex for it to be included in the group map. For statistical assessment, we compared Dice coefficients across the two alignment methods using a repeated measures analysis of variance (ANOVA) with individual regions as different entries, alignment method (CBA vs. NVA) as within-subject factor, and hemisphere as between-subject factor. We ran this comparison on two different thresholds: once on unthresholded group maps, and once on a threshold that produced - across regions and methods - the highest predictability. To determine this threshold, we averaged Dice coefficient values across alignment methods, hemispheres, and ROIs, resulting in one Dice coefficient per threshold level (as previously done in Rosenke et al. 2018). Comparison across thresholds revealed that a threshold of 0.2 produced the highest predictability. Additionally, we ran paired permutation tests within each region on Dice coefficient results at threshold 0.2 to establish whether the specific region showed a significant Dice coefficient for either alignment (NVA or CBA). Finally, we calculated the mean ROI surface area (in mm2) for each hemisphere and ROI (Fig. 2.3.3) and used a paired t-statistic to assess whether there was a systematic hemispheric difference in size across ROIs.

### Generating a visual functional atlas (visfAtlas) by assigning each voxel and vertex to a unique fROI

The processes described below provide a non-overlapping tiling of the functionally defined regions in occipito-temporal cortex in surface as well as volume space (Fig. 2.3.5).

*Cortex-based alignment:* The probability maps determine the probability that each vertex belongs to a given fROI. However, it is possible that a point on the brain may belong to more than one probabilistic fROI. This overlap is more likely to occur along boundaries of neighboring functional regions. In order to assign a unique functional label to each vertex in the atlas, we generated a maximum-probability map (MPM) of each area, once in volume space (NVA) and once in surface space (CBA). Using the probabilistic fROIs, we determined which vertices were shared by more than one probabilistic fROI and assigned these vertices to a single fROI based on the area which showed the highest probability at that vertex (Eickhoff et al., 2005). In cases where two areas held the same probability value for one vertex, we averaged the probabilistic values of neighbors of that vertex for each of the fROIs. The degree of neighbors averaged was increased until the vertex had a higher probability value in one of the areas. Lastly, after all vertices were assigned in each of the MPM areas, we searched for individual vertices that were not connected to other vertices of the same ROI. We used a decision threshold where a minimum of at least one 3rd degree neighbor for each vertex had to be in the same group ROI for that vertex to be part of the group ROI. In cases where single vertices where detected, they were assigned to the ROI with the second-highest probabilistic value and same-ROI vertices in the immediate neighborhood.

*Nonlinear volume alignment:* The creation of a maximum probability map in volume space was identical to that for CBA as described above, except for the neighborhood search. The neighborhood search was implemented differently as the 3D nature of the volume atlas would lead to inevitable differences in the MPM creation when compared to the surface atlas. Neighborhood search was only performed for 1 immediately adjacent voxel in all three dimensions.

### A visual functional atlas available in volume and surface space

The unique tiling of functionally defined visual regions provides a functional atlas (visfAtlas) which we make available (1) in volume space, and (2) in surface space. In addition, we make this atlas available in multiple file formats. Volume: we publish the volumetric visfAtlas in MNI space in BrainVoyager

file format (VOI file) and NifTi format, which can be read by a variety of software packages. Surface: we publish the visfAtlas in file formats compatible with Brain Voyager as well as FreeSurfer. Note, however, that the surface atlases are generated slightly differently for each software. For BrainVoyager, we generated a publicly available dynamic group average brain (BVaverage, Fig. 2.3.5C) that will be available with the distributed atlas, details are described above. Since FreeSurfer (https://surfer.nmr.mgh.harvard.edu/) is commonly used with the fsaverage brain, an average surface of 39 individuals, we converted the individually defined fROIs from each subject to cortical surface space in FreeSurfer after running each subject through the recon-all pipeline. Then, we used the FreeSurfer CBA algorithm to bring each subject's fROIs to the fsaverage space. Further processing was done as described above and the same for both software packages. All files can be downloaded here: download.brainvoyager.com/data/visfAtlas.zip.

**Evaluating whether fROI size and reproducibility are related to inter-subject consistency**

There are several factors that can influence consistency across subjects. First, region of interest size has been shown to influence across-subject consistency measures using the Dice coefficient (Rosenke et al., 2017b). Therefore, we determined if there is a correlation between the cross-validated Dice coefficient and average fROI surface area. Second, we established whether categories differ in reproducibility of cortical responses within a subject. We reasoned that across-subject variability cannot be expected to be lower than within-subject variability over time (reproducibility), hence it can be used as a proxy for noise ceiling. To measure reproducibility, we first defined two regions of interest, ventral temporal cortex (VTC) and lateral occipito-temporal cortex (LOTC). VTC was manually defined by tracing well known anatomical: the occipitotemporal sulcus (OTS), posterior transverse collateral sulcus (ptCoS), parahippocampal gyrus (PHG) and the anterior tip of the mid-fusiform sulcus (MFS). LOTC was defined as previously described in Weiner and Grill-Spector (2013). Posteriorly, the LOTC ROI was defined at the convergence of the intraparietal sulcus (IPS) and the descending limb of the superior temporal sulcus (STS). The superior boundary was defined at the dorsal lip of the STS, and inferior boundary at the occipitotemporal sulcus (OTS). We then computed general linear models for all three individual fLoc localizer runs we acquired and computed t-statistic contrast maps identical to those used for our ROI definitions (e.g. faces vs all other categories, see ROI definition section for details), resulting in 3 contrast maps

for each subject for each of the 4 categories: characters, bodies, faces, and places. Consequently, we computed the Dice coefficient between each pair of runs for each subject, hemisphere, and ROI, separately (run 1 and 2, 1 and 3, and 2 and 3 within VTC and LOTC). We then took the average across those three splits as the Dice coefficient for that subject. Ultimately, we performed this analysis with a liberal statistical threshold of $t > 0$ (any vertex holding a positive t-value is included) and once with a threshold of $t = 2.2$ ($p<0.01$) for vertices to be included in the contrast map (Fig. 2.3.4). Together, these measures result in a lower and upper bound estimation of our Dice coefficient noise ceiling.

**Validation of the visfAtlas with an independent dataset of category-selectivity in ventral temporal cortex and with an increasing number of subjects**

Common consideration in building atlases are (i) the number of subjects that are used to build the atlas and (ii) how well it can predict new datasets. To address whether our sample size is sufficient to achieve generalizability to new data, we tested how well the visfAtlas predicts fROIs of 12 new subjects. These data were acquired using a similar localizer in a different scanning facility, identified by independent experiments, and have been published previously (Stigliani et al., 2015; Weiner et al., 2017). We compared their fROI definitions of mFus-faces, pFus-faces, OTS-bodies, pOTS-characters and CoS-places to our visfAtlas definitions in the following ways: (1) we visualized our visfAtlas MPMs in relation to their probability maps of each of the fROIs (Fig. 2.3.6), and (2) we calculated how well our visfAtlas predicted each of their individual subjects' fROIs using the Dice coefficient.

Lastly, to address how the number of subjects affects the accuracy of our visfAtlas, we calculated the Dice coefficient for different iterations of the visfAtlas in which we incrementally increased the number of subjects from 2 to 19; specifics are in the Supplemental Materials.

**Functional responses of atlas fROIs in left out data**

When using a probabilistic atlas, it is of great interest not only to know how likely one would find a new subject's fROI in the same location, but also what signals would be picked up for that subject within an atlas-fROI. For example, are voxel in face-selective atlas fROIs responding mostly to faces? To test the generalizability of our atlas, we performed a leave-subject-out maximum responsivity analysis. The analysis calculates the percentage of voxel responding highest to each condition within a given fROI, where the fROI is defined on

all subject's data except the one dataset used for the responsivity computation. This was repeated for all possible leave-subject-out combinations. First, for each subject individually we created a maximum probability map (MPM) based on the other N-1 subjects (leaving the target subject out). Then, for each individual voxel within each fROI in this MPM, we estimated the average response amplitude to each category across trials using the optimized Least Squares – Separate (LS-S) trial estimation approach as described by Mumford et al. (2012). Then, we created a 'winner map' for each fROI per subject, in which the condition index that yielded the strongest response was assigned to each voxel within the fROI. Per condition, we counted the number of winning voxels within the ROI, which we expressed as a percentage of the total number of voxels in the fROI. This procedure was repeated for each subject (Fig. 2.3.7).

**Comparison of our visfAtlas to existing publicly available atlases and relevant fROIs**

How does the visfAtlas compare to published atlases? While there is no complete occipitotemporal atlas of visual areas yet, atlases of retinotopic areas have been published by Wang et al. (2014) and Benson et al. (2012, 2014). To compare our atlas to the Benson atlas where there is no separation between ventral and dorsal quarterfields, we merged our dorsal and ventral V1-V3. Additionally, there is a published probabilistic atlas of CoS-places (Weiner et al. 2018), and motion selective hMT+ (Huang et al. 2019). We compared our surface visfAtlas to the existing surface maps by assessing their correspondence in the FreeSurfer fsaverage space. For each published atlas we (i) qualitatively assessed the spatial correspondence by visualizing the atlas definitions on a common brain space of the FreeSurfer average brain (Fig. 2.3.8) and (ii) quantitatively assessed the correspondence by calculating the Dice coefficient between each of our individual subject's fROIs and the respective other atlas as we do not have access to the individual subject data in the Wang, Benson or Huang atlases.

## 2.3 Results

Using data from 19 healthy participants we aimed at generating a probabilistic atlas of occipito-temporal and ventral temporal cortex. Individually defined regions were normalized to group space using either (1) cortex-based alignment (CBA) or (2) nonlinear volumetric alignment (NVA).

### 2.3.1 Superior spatial overlap after cortex-based alignment for retinotopic and category selective regions

In order to determine whether nonlinear volumetric (NVA) or cortex-based alignment (CBA) result in higher accuracy and predictibility of our atlas, we aimed at comparing both alignment techniques across all functional regions of interest (fROIs). Figure 2.3.1 displays three example regions, one early visual retinotopic region in occipital cortex (V1d), as well as two higher-order category-selective regions in ventral temporal cortex (CoS-bodies and mFus-faces). Qualitatively, a higher degree of consistency across subjects is observable when group maps were normalized using CBA as compared to NVA. Both V1d and Cos-places display a high consistency in the group map center as indicated by yellow colored vertices, while centers are more variable after NVA alignment, most evident in V1d. For mFus-faces, both group maps display a greater degree of variability across subjects than the other two regions.

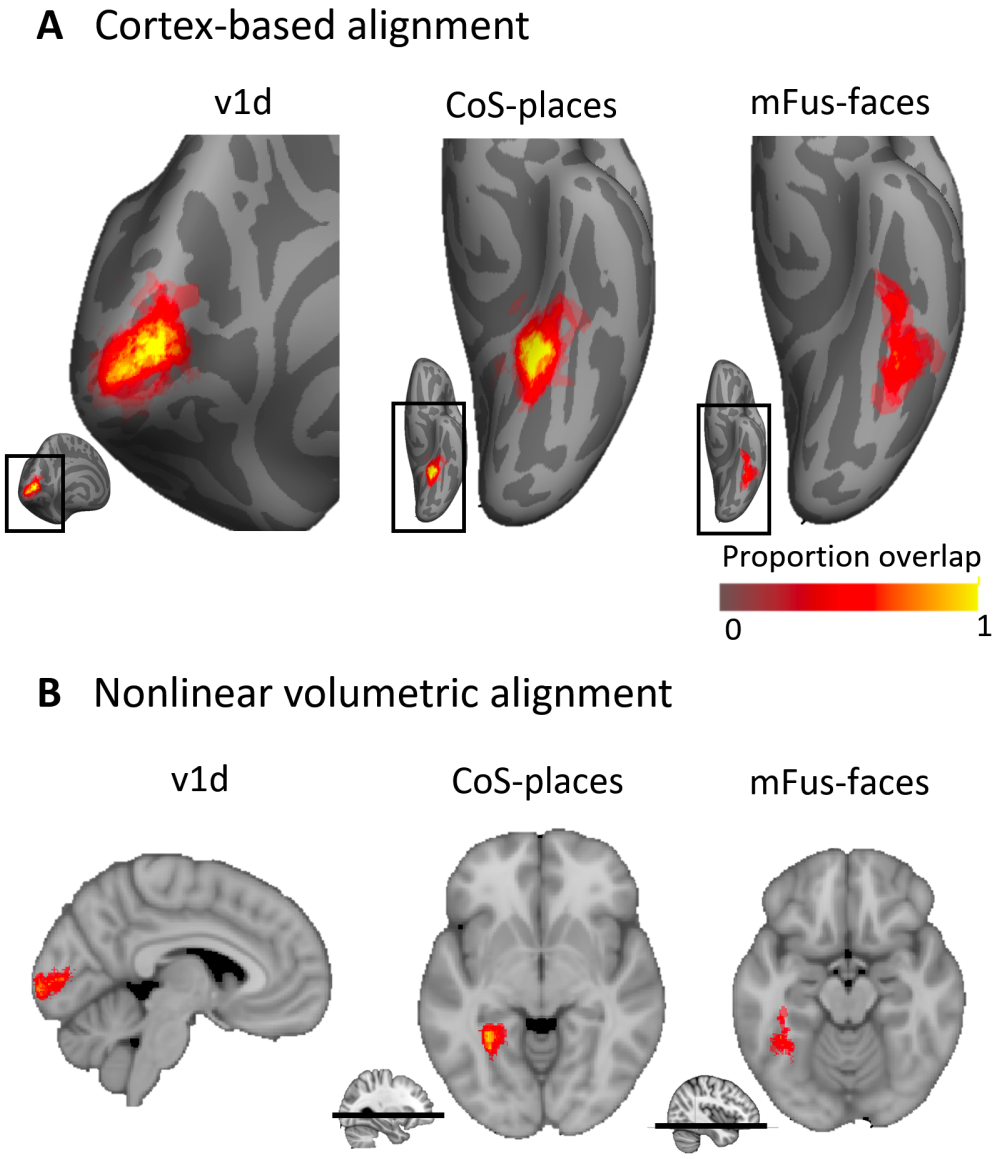



**Figure 2.3.1: example probabilistic group maps in the left hemisphere after two brain alignments.** (A) Three example regions-of-interest (ROIs) are displayed where the most left column, v1d, shows an early visual cortex map and the middle and right columns display two higher-order visual category-selective regions in ventral temporal cortex, Cos-places and mFus-faces. Probability values range from 0 to 1 where 0 indicates no subject at a given vertex and 1 that all subjects in the probabilistic maps shared the given vertex. mFus-faces reveals less consistency as shown by a lower percentage of yellow-colored vertices. Bottom inset displays zoomed in location of the main figure. (B) Same ROIs as in A but after nonlinear volumetric alignment (NVA). Bottom inset for CoS-places and mFus-faces indicates the location of the axial slice in the volume.

To quantify which group alignment resulted in higher consistency and therewith predictability, we used the Dice coefficent (DSC) and a leave-one-out cross-validation (LOOCV) procedure to determine the predictability of finding the same region in a new subject. Moreover, we calculated the Dice coefficient using different thresholds for the probabilistic group map, ranging from a liberal unthreshold (one subject at a given voxel/vertex is enough to assign it to the group map) map to a conservative threshold where all N-1 subjects had to share a voxel/vertex to be assigned to the group map (Fig. 2.3.2). For retinotopically defined regions, DSC's varied between 0.35 and 0.59 for peak probability after CBA, and between 0.30 and 0.42 after NVA. Especially regions with a lower predictability overall tended to show higher predictability after NVA for more conservative group thresholds (e.g. Fig. 2B, mFus-faces, TOS-bodies). For CBA, peak predictibility (DSC) for each region ranged from 0.1 to 0.60, while it ranged from 0.1 to 0.42 for NVA, with character-selective regions showing the lowest consistency for both alignments, closely followed by mFus- and IOG-faces.

Quantitatively, CBA displayed an overall greater predictability across regions and thresholds (except for V3d LH, see Fig. 2.3.2A), which was confirmed by a significant difference in alignment for both unthresholded ($F_{(1,34)} = 20.12, p < .001$) and thresholded (0.2; ($F_{(1,34)} = 174.84, p < .001$) probability maps, see Methods for details on threshold selection. Additionally, there was no significant main effect for hemisphere (unthresholded: p = .90; thresholded: p = .56) and no interaction between alignment and hemisphere (unthresholded: ($F_{(1,34)} = .85, p = .36$), thresholded: ($F_{(1,34)} = 0.35, p = .56$). We followed up with a paired permutation test (across alignments) for the unthresholded DSC within each fROI. As there was no main effect for hemisphere (see above) and no significant difference in region size across hemispheres (t(17) = -0.48, p = .64, Fig. 2.3.3), permutation tests were performed on Dice coefficients using an unthresholded group map prediction and averaged across hemispheres. Results show that CBA alignment has a higher predictability than NVA for all regions (p < .05), except for unthresholded: pOTS-characters (p = 1), IOS-characters (p = .81), v3d (p = .05), IOG-faces (p = .05) and thresholded: V3d (p = .05), mFus (p = .70), IOG (p = .55), pOTS (p = 1), IOS (p = 1), OTS (p = .14).

As shown in the previous section and displayed in Figure 2.3.2, different category-selective regions in VTC and LOTC show different levels of Dice coefficients. One factor that may contribute to this variability is the region's size, which also varies across fROIs (Fig. 2.3.3). To test if this relationship is significant, we measured the correlatation between the Dice coefficient and surface area of the fROIs. Results indicate a significant correaltion (left hemisphere: $r = 0.83$, $p < 0.01$; right hemisphere: $r = 0.85$, $p < 0.01$), suggesting that larger regions have higher Dice coefficients. We also examined if differences in Dice coefficient are related to differences in noise ceiling across ROIs. As a measure of noise ceiling, we calculated the within-subject Dice coefficient across the 3 runs of the fLoc. We reasoned that if there are between-ROIs differences in the noise ceiling estimated from within-subject Dice coefficients, they would also translate to the between-subject Dice coeffient. When using a lenient t-map threshold, results (Fig. 2.3.4) indicate that within-subject Dice coefficient for a lenient t-map threshold ($t>0$) range from 0.4 – 0.77 across categories. We find a higher Dice coefficient for bodies and faces in left VTC, and a higher Dice coefficient for places in the right VTC. In LOTC, the highest within-subject Dice coefficient is for place-seletivity in the left LOTC, and body-selecitivty in the right LOTC. Given that within- and between-subjects Dice coefficients are in the same range and vary similarly across fROIs, we believe that the precision of the visfAtlas will allow to identify fROIs in individual participants.

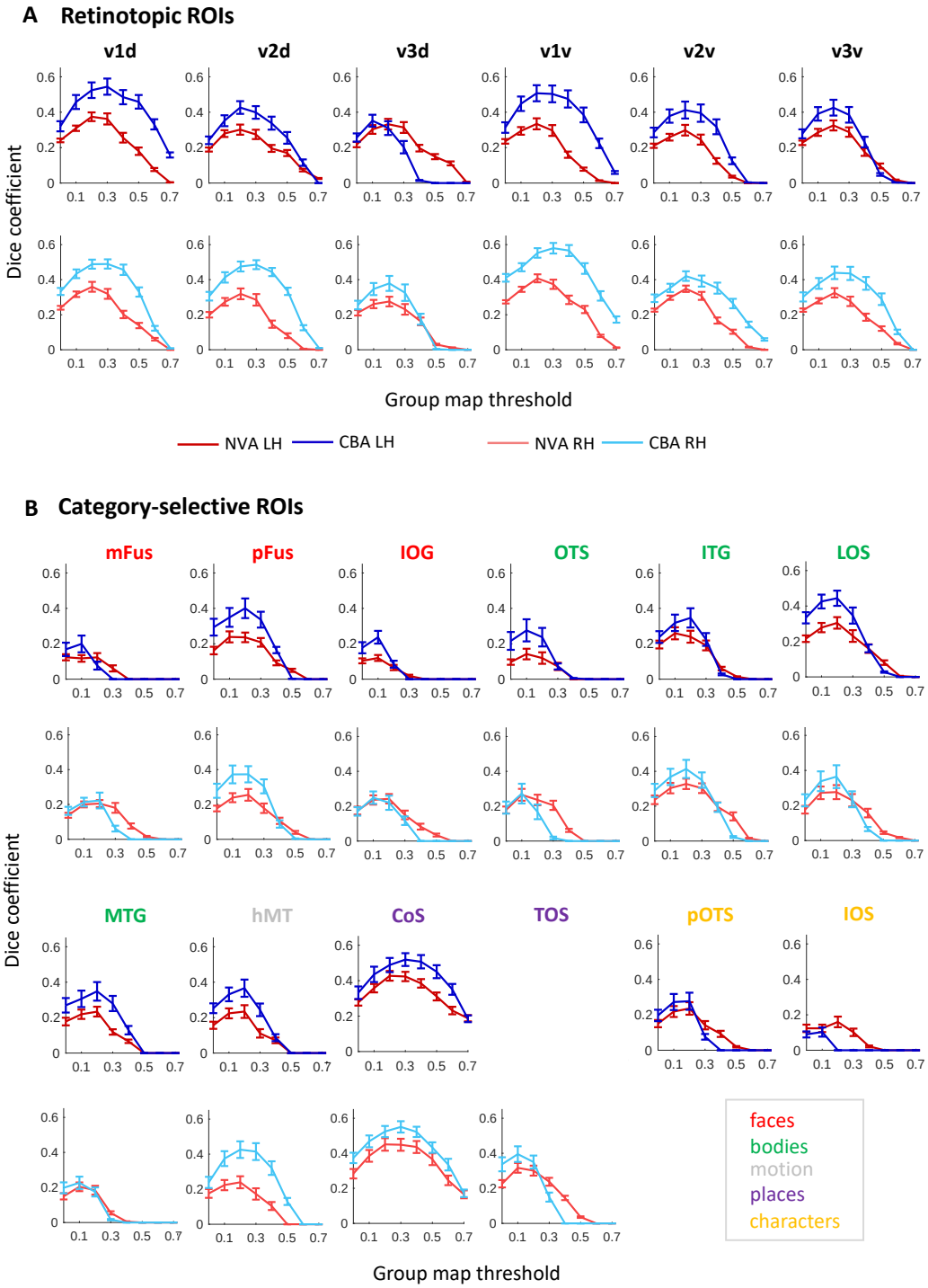



**Figure 2.3.2: leave-one-out cross-validation predictability analysis using the Dice coefficient (DSC) for retinotopic regions (A) and category-selective regions (B).** x-axis: threshold of the probability map generated using N-1 subjects, y-axis: DSC. A DSC value of 1 indicates perfect overlap between the N-1 group map and the left-out subject, 0 indicates no overlap. Blue lines: DSC after CBA, red lines: DSC after NVA. Dark colors/top rows correspond to left hemisphere data, light colors/bottom rows to right hemisphere data. Red: face-selective ROIs, green: body-selective ROIs, yellow: character-selective ROIs, gray: motion-selective ROI, error bars: standard error (SE) across the N-fold cross-validation.

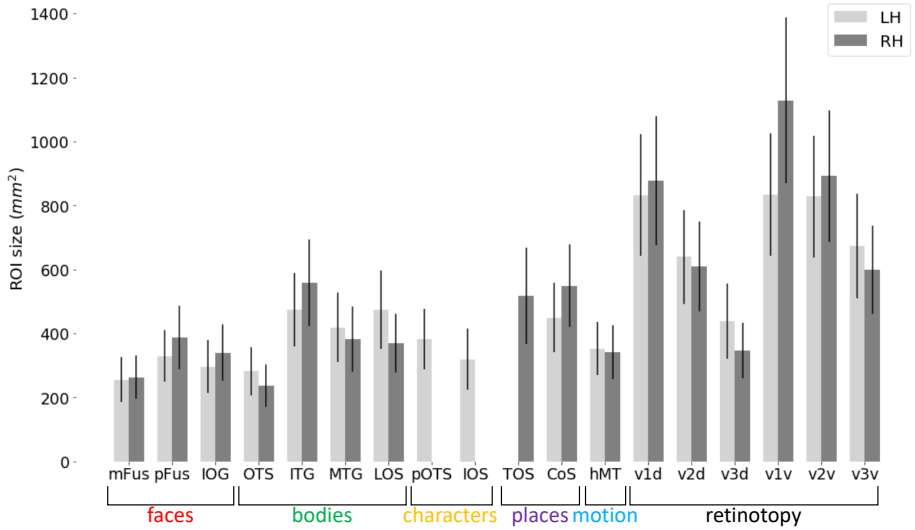



**Figure 2.3.3: fROI size across occipito-temporal cortex.** Average ROI size in surface space separately for the left hemisphere (LH, light gray) and right hemisphere (RH, dark gray). Error bars: standard error across subjects. Regions of X-axis are organized by category.

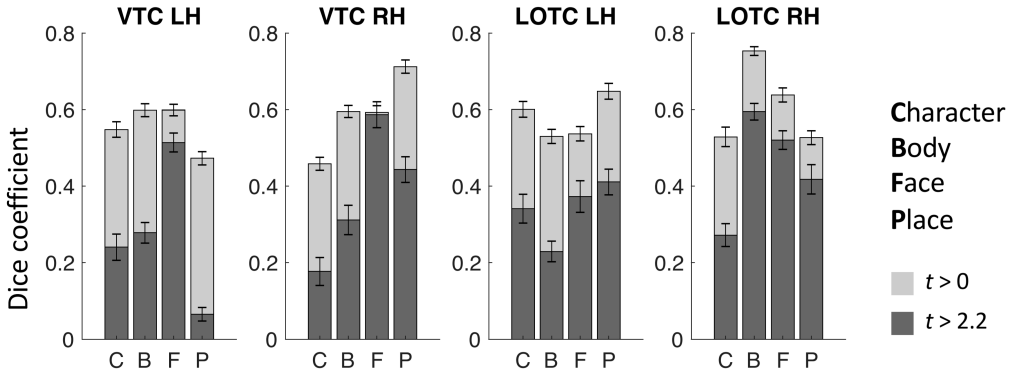



**Figure 2.3.4: reproducibility of category-selectivity responses.** For the two cortical expanses that contain the category-selecitve regions of the visfAtlas, VTC and LOTC, the reproducibility of category responses was computed across the t-contrast maps of single runs for each respective category (see Materials and Methods for details). Dark gray bars represent the Dice coefficient results based on t-contrast maps that were thresholded with $t > 2.2$, which equals $p < 0.01$, while light gray bars were based on t-contrast maps that were thresholded at $t > 0$. Errorbars represent standard errors across subjects.

### 2.3.2 A functional atlas of occipito-temporal cortex in volume and surface space

By systematically varying the group map threshold for predicting a left-out subject's fROI, we established that a group map threshold of 0.2 allows for greatest predictability across regions. Using the 0.2 threshold, we generated a functional atlas of occipito-temporal cortex by generating a maximum probability map (MPM, see Methods for details). Figure 2.3.5 displays the resulting unique tiling of category-selective regions in stereotaxic space for surface (Fig. 2.3.5A) and volume (Fig. 2.3.5B) space. The visfAtlas is publicly available in both surface as well as volume space to allow usage in a variety of analyses and in file formats for BrainVoyager and FreeSurfer for surface space as well as in volume space using the NifTi format. In addition, we publish a BrainVoyager average brain (BVaverage, Fig. 2.3.5C; `download.brainvoyager.com/data/visfAtlas.zip`).

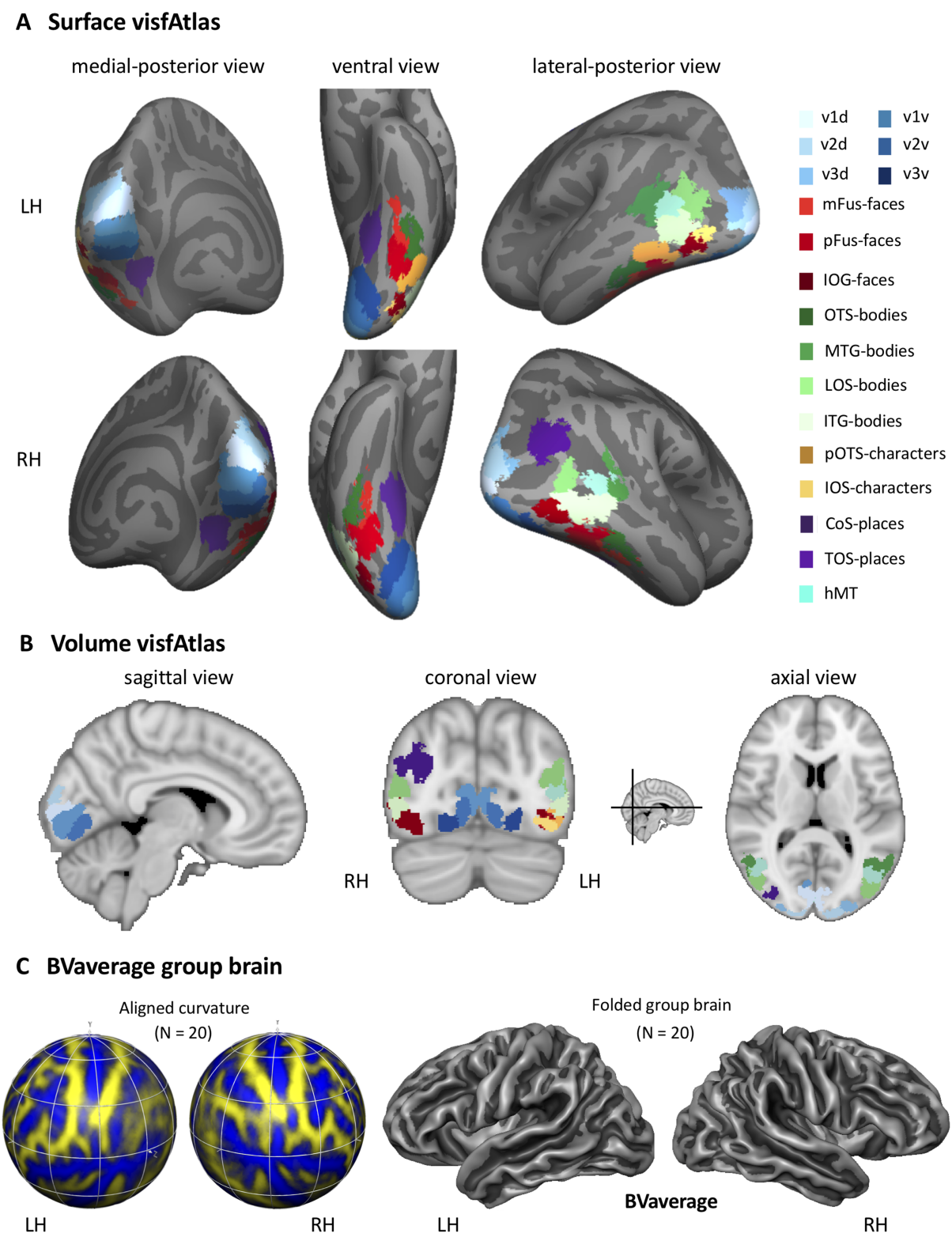



**Figure 2.3.5: maximum-probability map (MPM) of occipito-temporal cortex functional regions-of-interest (fROIs).** (A) visfAtlas in surface space after cortex-based alignment. Each color displays a unique fROI group map thresholded at 0.2 of all subjects in which the given fROI could be identified. (B) Volume atlas using the same color coding as in surface space. Inset between coronal and axial view displays the slice location for coronal and axial slices, respectively. LH: left hemisphere, RH: right hemisphere. (C) A new group average brain (BVaverage) published in BrainVoyager, based on 20 adults. This average brain can be used for future studies as a common reference brain.

### 2.3.3 Atlas validation using an independent dataset and an increasing number of subjects

How well does the visfAtlas localize regions in new subjects scanned at a different scanner and facility? To answer this question, we compared the ventral visfAtlas ROIs with a dataset acquired at Stanford University (Stigliani et al., 2015; Weiner et al., 2017) using different subjects and a functional localizer experiment similar to ours. Figure 2.3.6 shows unthresholded probabilistic maps of Weiner's MPMs (across 12 participants) and our respective visfAtlas MPMs. Qualitatively, the location of their probabilistic maps, especially peak probabilities, correspond to our respective visfAtlas ROIs. To quantify the similarity, we tested how well our data predict the fROIs of these 12 independent subjects by calculating the Dice coefficient between our MPM fROIs and each of the independent subjects' fROIs (Fig. 2.3.6B). The mean Dice coefficients (+/- SE) for left and right hemispheres, respectively, are in a similar range as the Dice coefficient of the leave-one-out-cross-validation results of our data (compare Fig. 2.3.2 threshold 0.2 with Fig. 2.3.6B).

Additionally, we explored how the number of subjects used for generating our atlas affects its accuracy(Supplemental Fig. 5.4.1). Results indicate that in general, having more participants generates better accuracy in the LOOCV, but the number of requireed subjects varies across ROIs. Overall, across all ROIs, the highest Dice coefficient plateaus between 12 and 14 subjects, suggesting that our atlas based on an average of 16 subjects per ROI (see Table 2.1 for details) is sufficient.

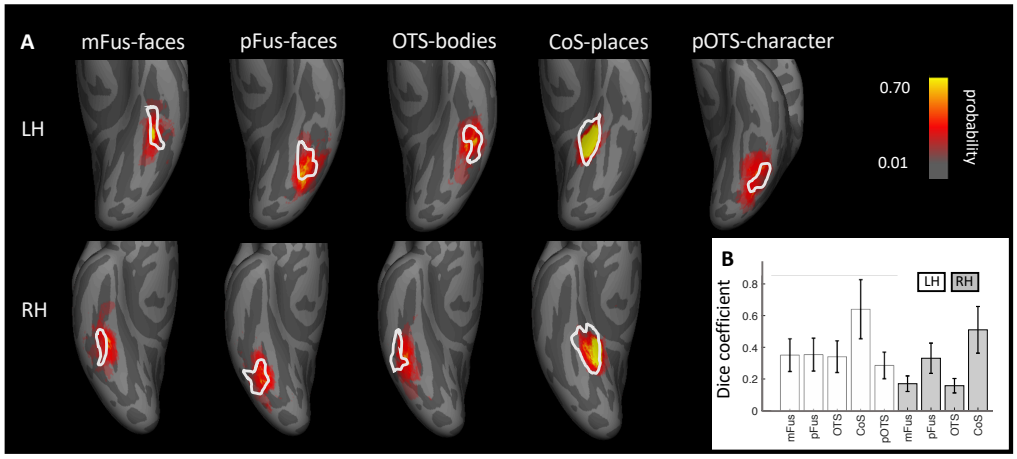



**Figure 2.3.6: correspondence between the visfAtlas and 5 ventral temporal cortex probabilistic maps from independent data.** (A) We compared visfAtlas MPM fROIs (white outlines) in VTC with probabilistic maps (colored regions) of 6 functional regions from an independent dataset that used a similar localizer, which has been published previously (Stigliani et al. 2015; Weiner et al. 2017). Top: left hemisphere; Bottom: right hemisphere. (B) Average Dice coefficient between fROIs of the individual subjects from Stigliani and Weiner and colleagues and the MPMs of our visfAtlas fROIs. Errorbars: standard errors across subjects. LH: left hemisphere; RH: right hemisphere.

### 2.3.4 Generalizability of functional atlas: functional responsivity in left out data

One of the advantages of a probabilistic atlas is the ability to locate a region of interest with a degree of certainty (as established using the Dice coefficient analysis) in a new subject without the need to run a localizer itself. In order to quantify the atlas' generalizability, the category responsivity of the category selective areas in new participants is a crucial metric. Therefore, we performed a leave-subject-out responsivity analysis in volume space to assess category- responsivity. For each fROI, we established the percentage of voxel that showed the strongest response to each available category (Fig. 2.3.7, see Methods for details of responsivity estimation). For all category selective regions, we confirmed that the category it is selective for indeed yields the highest percentage of maximum voxel responsivity across subjects. Face-selective fROIs (Fig. 2.3.7, top left) contain 52-72% (lowest to highest fROI) face-selective voxel responses (red). The second-highest maximum responsivity is body-selective (green) with 10-43% on average across subjects, followed by character-selective

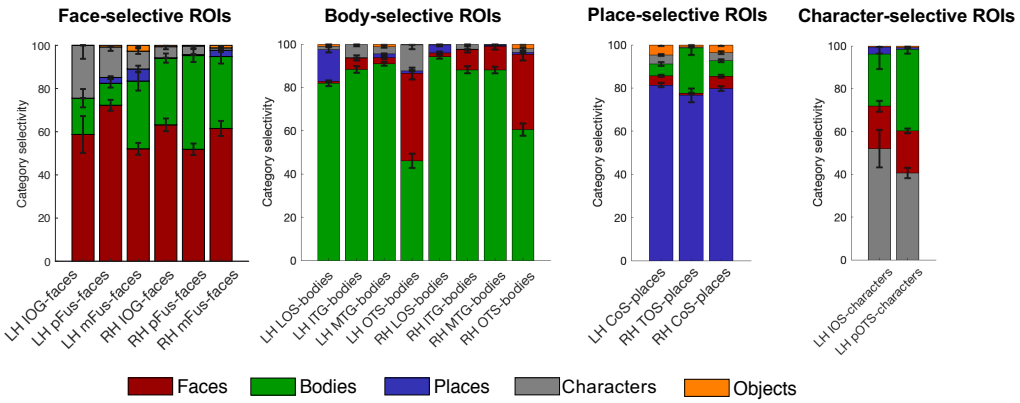



**Figure 2.3.7: proportion of voxels that show maximum responsivity in left out subjects are largely their own category.** Using our volumetric atlas data we generated a cross-validated estimate of voxel maximum responsivity in a left out subject. N-1 times, we generated a volumetric maximum probability map and calculated the proportion of voxel that were maximally responsive for the ROI's category, e.g. face response voxel in mFus-faces. This gives an estimate for the expected specificity of the atlas. For each major category - faces, bodies, places, characters – proportions of category responsivity are displayed with each region's preferred category as the bottom bar of each stacked bar graph. Error bars: Proportion own category selectivity across all left-out subjects.

regions (gray) with 2-25%. Body-selective regions (Fig. 2.3.7, top right) contain the highest proportion of body as maximum voxel responsivity for lateral body-selective regions (80-94%), with lowest proportions for ventral OTS-bodies in left and right hemisphere (46-55%). The second-largest number of voxel-maximum- responsivity is faces (1- 40%). Place-selective fROIs (Fig. 2.3.7, bottom left) show a large proportion of voxels with their preferred place responses (purple, 77-82%), followed by up to 21% body-maximum voxel responsivity. Character-selective ROIs (Fig 2.3.5., bottom right) on the other hand contain 41 - 52% character-response voxel, followed by up to 38% body-response voxels.

### 2.3.5 Similarities between previously published atlas areas and our visfAtlas

In order to establish the correspondence of our probabilistic functional atlas to other atlases, we made quantitative comparisons to existing atlases of one or multiple regions localized with comparable stimuli. As retinotopic atlases are frequently used to define early visual cortices in new subjects, we wanted to compare our retinotopic areas V1-V3 dorsal and ventral to a group atlas of retinotopic visual areas aligned to the fsaverage brain by Wang et al. (2014). To assess the correspondence between the two atlases we computed the Dice

coefficient (see Methods for details) between the existing group atlas and each of our visfAtlas subjects (Fig. 2.3.8) separately. Qualitatively, V1d and V1v from both atlases show a high degree of overlap and correspondence decreases when moving to the dorsal and ventral V2 and V3 (Fig. 2.3.8A). However, for each of the probabilistic maps of our visfAtlas regions, the peak probability location falls within the MPM published by Wang et al. (2014). This observation is confirmed by high Dice coefficients for V1d and V1v in the left and right hemisphere (average Dice coefficient 0.4 – 0.5, see Fig. 2.3.8E), and lower Dice coefficients in V2 and V3 (average Dice coefficient 0.15 – 0.4, Fig. 2.3.8E). Next, we also compared our visfAtlas retinotopic regions to an anatomical prediction of V1-V3 by Benson et al. (2012), which shows a similar pattern of correspondence with a greater overlap in V1 (0.4 – 0.42) and a decrease in V2 and V3 (0.2 – 0.29).

Similiar to the retinotopic regions, we compared a category-selective region - the CoS-places fROI - to a published probabilistic version by Weiner et al. (2018) which used a very similar localizer for their study. Both atlases display a high correspondence, with a slightly higher Dice coefficient in the left hemisphere than in the right hemisphere (Fig. 2.3.8E). On lateral occipito-temporal cortex we compared a recently published motion selective group area of hMT+ that has been defined using data from 509 adults (Huang et al., 2019).As Huang's et al (2019) group fROI was not bounded by body-selective regions but ours was defined by maximum probability map (MPM) that takes into account the neighboring face and body-part areas, the visfAtlas is smaller than Huang's definition Nonetheless, also here, the locus of our hMT+ probabilistic map is within the hMT+ atlas published by Huang et al (2019).

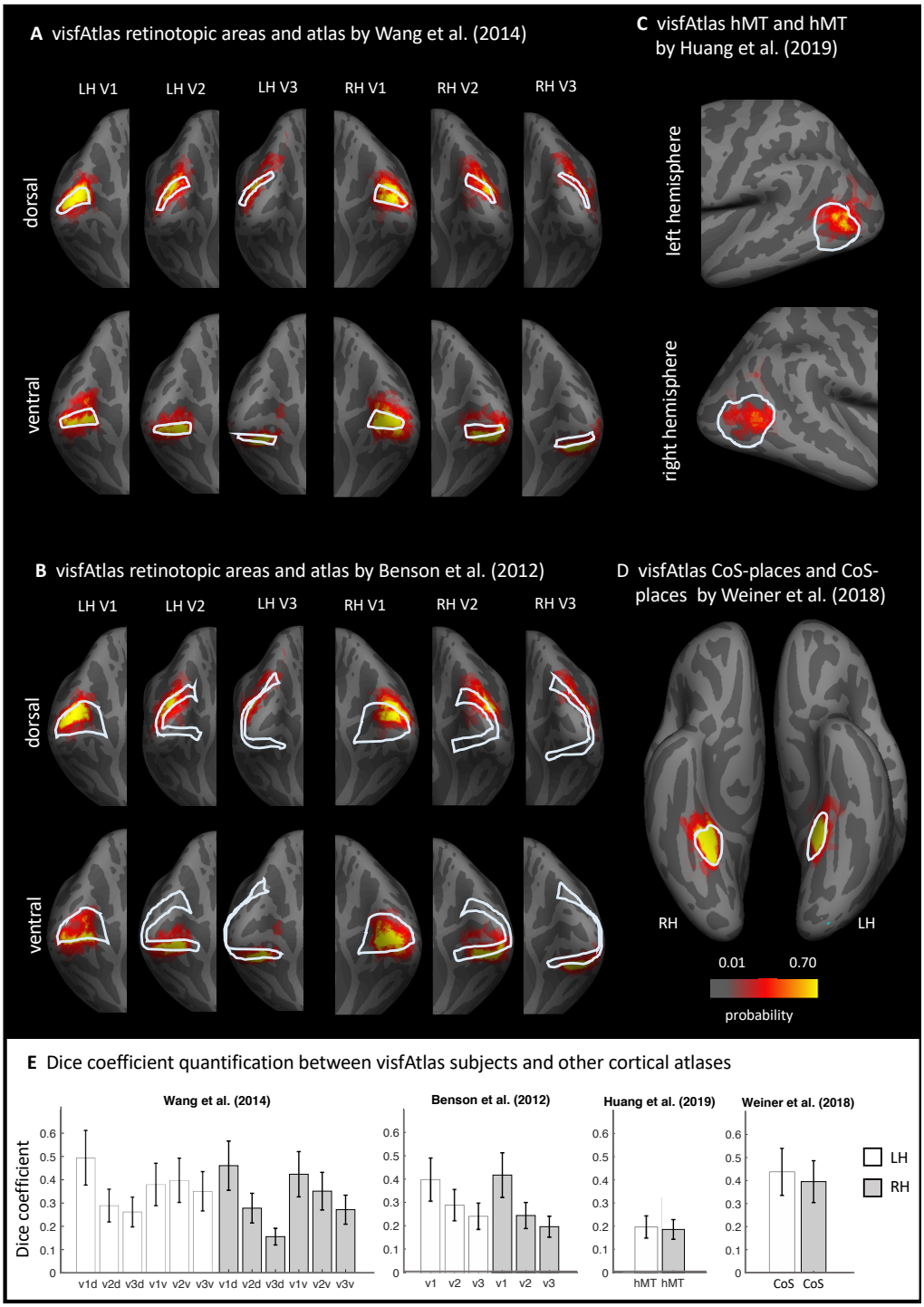



**Figure 2.3.8: comparison of the visfAtlas to other probabilistic atlases.** In A-D each red-yellow map is the probabilistic map of unthresholded individual regions of the visfAtlas ROI and the outline is the fROI of the relevant atlas; all images are show in the fsaverage brain. (A) Comparison of V1-V3 dorsal and ventral of the retinotopic atlas published by Wang et al. (2014) and our respective visfAtlas regions. Regions are presented on a medial-occipital view of the fsaverage group brain. (B) Comparison of V1-V3 dorsal and ventral to the anatomically estimated V1-V3 (Benson et al., 2012). (C) Comparison of motion-selective hMT+ published by Huang et al. (2019) to visfAtlas hMT+ probabilistic map. (D) Comparison of CoS-places published by Weiner et al. (2018) to the visfAtlas CoS-places map. (E) Dice coefficient between the visfAtlas fROI and the same fROI defined by other atlases. Errorbars: Standard error across 19 visfAtlas subjects. LH: left hemisphere; RH: right hemisphere.

## 2.4 Discussion

In the present study, we generated a cross-validated functional atlas of occipito-temporal visual cortex, including early-visual cortex retinotopic regions as well as category-selective regions. Additionally, we evaluated how accurately this atlas predicts category-selectivity in left-out subjects. We found that cortex-based alignment (CBA) outperforms nonlinear volumetric alignment (NVA) for most ROIs. Importantly, using CBA our probabilistic category-selective ROIs accurately identify 40% - 94% of category-selective voxels in left-out subjects (Fig. 2.3.7). We make this functional atlas (visfatlas) of occipito-temporal cortex available on cortical surfaces of the fsaverage (FreeSurfer) and BVaverage (BrainVoyager), and volume formats in MNI space compatible with the majority of software tools.

In the following we will discuss the implications of our results for theories of anatomical and functional coupling in visual cortex, how our atlas relates to other atlases in the field, whether it can be validated by independent data, and how future research can expand on our atlas with new methodological approaches.

### 2.4.1 Cortex-based alignment improves the consistency of group fROIs: implications

Spatial consistency in both retinotopic and category-selective regions was on average higher after CBA as compared to NVA (Fig. 2.3.2). The higher performance of CBA is in agreement with previous studies that reported that CBA results in atlases with higher accuracy than volumetric atlases (Coalson et al., 2018; Frost & Goebel, 2012), and specifically of retinotopic visual areas (Wang et al., 2014, Benson 2012) and cytoarchitectonic regions (Rosenke et al., 2017a, 2017b). Since CBA specifically aligns macroanatomical landmarks, the higher accuracy of CBA suggests a coupling between macroanatomical landmarks and functional regions. These results are consistent with prior research showing striking functional-macroanatomical coupling in visual cortex including: (i) V1 with the calcarine sulcus (O. P. Hinds et al., 2008), (ii) V3A and the transverse occipital sulcus (Nasr et al., 2011; Tootell et al., 1997), (iii) hV4 and the posterior transverse collateral sulcus (Witthoft et al., 2014), (iv) motion-selective hMT+ and the posterior inferior temporal sulcus (Dumoulin et al., 2000; Weiner & Grill-Spector, 2011), (v) mFus-faces and the mid-fusiform sulcus (Grill-Spector & Weiner, 2014) and (vi) CoS-places and the intersection of the anterior lingual

sulcus with the collateral sulcus (Weiner et al., 2018). One interesting observation regarding the Dice coefficient results (Fig. 2.3.2) is that in some fROIs, NVA produces a higher Dice coefficient than CBA for high threshold values (e.g., pOTS-characters LH, mFus-faces RH). We hypothesize that since NVA is operating in 3D volume space and CBA in cortical surface space, shifts around crowns of gyri or fundi of sulci may produce a large impact on CBA than NVA. This hypothesis can be tested in future research.

Historically, the prevailing view (Glasser & Van Essen, 2011; Haxby et al., 2011; Orban et al., 2014; Osher et al., 2015) was that higher-level functional visual regions have greater variability across participants as well as relative to macroanatomical landmarks compared to early visual areas such as V2 and V3. However, as we summarize in the prior paragraph, improvements in measurements and analysis methods argue against this prevailing view. In fact, our leave-one-out cross-validation procedure shows that five high-level visual regions (pFus-faces, LOS-bodies, ITG-bodies, CoS-places, motion-selective hMT+) have similar correspondence across subjects comparable to early visual cortex. However, some functional regions (mFus-faces, pOTS-characters, MTG-bodies, Fig. 2.3.2, see also Frost and Goebel, 2012), show more variability across participants. This diversity suggests that other factors may affect our ability to predict high-level visual regions. First, the shape and size of the ROI may impact across-subject alignment. Indeed, we found that larger and more convex ROIs tend to align better across participants than smaller ROIs, reflected in the finding of a positive correlation between the Dice coefficient and the size of the fROI. Second, the degree of macroanatomical variability differs across anatomical landmarks. In other words, stable macroanatomical landmarks may be better predictors of functional ROIs than variable ones. For example, the anterior tip of the mid-fusiform sulcus (MFS) is a more stable anatomical landmark than its posterior tip, as the length of the MFS substantially varies across people. Consequently, the anterior tip of the MFS better predicts face-selective mFus-faces than the posterior tip predicts pFus-faces (Weiner et al., 2014). Third, the quality of cortex-based alignment may vary across cortical locations (see Frost and Goebel 2012, 2013). Thus, more fragmented and less salient macroanatomical landmarks, such as the partially fragmented occipito-temporal sulcus (OTS), may align less well across participants with CBA. This in turn impacts the registration of functional ROIs that are associated with these landmarks. Fourth, the reliability of functional ROIs across sessions within an individual, which indicates a noise ceiling, may vary across ROIs. To evaluate the latter, we performed a reproducibility analysis for our category-selective regions by analyzing all

three localizer runs independently (Fig. 2.3.4). This analysis highlights that running the same experiment multiple times within the same subject will not result in the exact same cortical activation pattern. Here, reproducibility estimates (Dice coefficients) ranged between 0.4 and 0.75 in VTC as well as LOTC, similar to Dice coefficient estimates by other studies (Bugatus et al., 2017; Weiner et al., 2016; Weiner & Grill-Spector, 2010). Notably, the reproducibility analysis together with the analysis of an independent dataset indicate that reproducibility and variability of our Dice coefficient are within the range expected by previous studies (Weiner et al., 2018). However, one has to note that our reproducibility estimation is conservative since we used the three runs that comprised our category-selectivity localizer individually, which means that each split had less trials and a lower signal-to-noise ratio (SNR) than the analysis used to establish between-subject variability (3 runs per subject each). Future work should run the same experiment for an additional full 3 runs to establish a noise ceiling that is not impacted by SNR and trial number differences.

Future research can also improve the inter-subject alignment by improving CBA methods. For example, CBA may be improved by weighting microanatomical landmarks by their consistency and saliency. Other directions for improving the predictions of the model may include incorporating additional features, such as spatial relationships between ROIs, or adding some functional data (Frost & Goebel, 2013) to improve predictions. For example, adding one retinotopic run improves predicting early visual areas relative to macroanatomical landmarks alone (Benson & Winawer, 2018).

### 2.4.2 Category-preferred responses within visfAtlas regions and reasons for variability across areas

As the main purpose of a functional atlas is to allow generalization to new individuals, confirmation and validation of the functional responses of the predicted regions is crucial. We used a leave-one-out-cross-validation approach to quantify the generalizability of our maximum probability map and demonstrate that voxels within the predicted ROI are displaying maximum responsivity to the preferred category of that ROI (Fig. 2.3.7). The highest proportion of own category-responsive voxels was in lateral body-selective regions and the lowest own category response was in character-selective regions. One possible explanation for this variability is the proximity of ROIs to regions selective for other categories. For example, in ventral temporal cortex, the body-selective region on the OTS is small and located between two larger face-selective regions, but in lateral occipito-temporal cortex, body-selective ROIs are larger and some

of them distant from the face-selective regions on the IOG. Close proximity between ROIs selective for different categories increases the likelihood of overlapping atlas boundaries, which may reduce the predictions of category-selectivity in a new subject.

Another reason for variability across areas could be that areas are differentially affected by the number of subjects they require to reach a stable prediction. To test this, for each ROI we calculated Dice coefficients with N=2 to max N for that ROI and evaluated how the overlap changed with increasing number of subjects (Supplemental Fig. **??**). Interestingly, our analysis suggests that not all ROIs benefit from an increasing number of subjects equally. More specifically, only 5 of the 18 ROIs displayed such an increase, and those suggest to plateau between 12 and 16 subjects. For other ROIs, the number of subjects did not impact the Dice coefficient. Generally, the assumption is that as the number of subjects increases, the level of noise decreases and one gets closer to the true between-subject variability. One interesting note is that using the data of our visfAtlas, none of the ROIs displayed a positive trend in Dice coefficient that continues past the number of subjects included in our atlas. Follow up work should evaluate whether this is local plateau or the global maximum Dice coefficient for each region.

Additionally, our approach can be extended to generate atlases of additional high-level visual regions that have other selectivities by including stimuli and contrasts for: (i) dynamic vs. still biological stimuli to identify regions selective for biological motion in the superior temporal sulcus (Beauchamp et al., 2003; Grossman & Blake, 2002; Pitcher et al., 2011; Puce et al., 1996), (ii) objects vs. scrambled objects to identify object-selective regions of the lateral occipital complex (LOC; Malach et al. 1995; Grill-Spector et al. 1998; Vinberg and Grill-Spector 2008), and (iii) colored vs. black and white stimuli to identify color-selective regions in medial ventral temporal cortex (Beauchamp et al., 1999; Lafer-Sousa et al., 2016). Furthermore, future studies may explore the possibility to generate more sophisticated atlases, which contain not only a unique tiling of cortical regions, but also allow for multiple functional clusters to occupy overlapping areas and indicate probabilities for multiple categories at each voxel, perhaps building a hybrid of probabilistic maps of single regions and a maximum probability map.

### 2.4.3 Consistent definitions of visual areas across different atlases

In generating our visfAtlas is was important for us to include early visual areas and hMT+ in addition to category-selective regions for two reasons: (1) it allowed us to benchmark and test our approach to atlases of retinotopic areas (e.g. Wang et al. 2014) and (2) it allowed us to generate a more comprehensive atlas of the visual system that includes the most studied visual regions spanning early and higher-level visual regions. Finding that our approach generates similar ROIs to other atlases (e.g., V1-V3 in the Wang et al. (2014) atlas, Benson et al. (2012) atlas) and hMT+ (Huang et al., 2019) is important as it illustrates that these ROIs are robust to experimental design, stimuli type, and number of subjects that were used for generating atlases, all of which varied across studies. For example, we defined hMT+ by contrasting responses to expanding and contracting low contrast concentric rings to stationary ones in 19 subjects but Huang et al. (2019) defined hMT+ by contrasting responses to dots moving in several directions vs. stationary dots in 509 subjects. Despite these differences, where hMT+ is predicted to be, largely corresponds across both studies (Fig. 2.3.8C), even as the predicted spatial extend of hMT+ is substantially smaller in our atlas as compared to Huang's. For retinotopic regions, we found the best correspondence between our data and Wang et al. (2014) for V1d and V1v, especially in the left hemisphere (Fig. 2.3.8A). Right hemisphere V1 of our visfAtlas extends more dorsally compared to Wang's atlas, consequently shifting right hemisphere V2d and V3d further compared to Wang et al. (2014). For both, the comparison to Benson et al. (2012) and Wang et al. (2014), we observe a reduction in overlap that corresponds to a reduction in Dice coefficient when quantifying V1 vs. V2 and V3 (see Fig. 2.3.2 for details), indicating that these may be individual differences across subjects that are independent of anatomical coupling, but still display less individual variability than previously assumed (see Discussion section).

### 2.4.4 Cortex-based alignment improves the consistency of group fROIs: implications).

Ultimately, the visfAtlas showed close correspondence to the comparison atlases, highlighting the robustness of our approach and the utility of functional atlases for future neuroimaging studies.

### 2.4.5 Conclusion and future uses

To this date, no probabilistic atlas has been published which contains such an extensive set of functional regions in occipito-temporal cortex. The present study shows that most of the category-selective regions can be predicted in new subjects.

This functional atlas of occipito-temporal cortex is available in both surface and volume space and can be used in commonly used data formats such as BrainVoyager and FreeSurfer. We hope that this atlas may prove especially useful for (1) predicting a region of interest when no localizer data is available, saving scanning time and expenses, (2) comparisons across modalities and (3) patient populations, such as patients who have a brain lesion (Barton, 2008; de Heering & Rossion, 2015; Gilaie-Dotan et al., 2009; Schiltz & Rossion, 2006; Sorger et al., 2007; Steeves et al., 2006) or are blind (Bedny et al., 2011; Mahon et al., 2009; Striem-Amit, Dakwar, et al., 2012; van den Hurk et al., 2017).

## Chapter 3

# Bayesian optimization of high-channel cortical implant locations for functional phosphene vision

## 3.1 Introduction

Millions of people around the world suffer from blindness and are faced with the associated challenges such as those involved in navigation -the detection and avoidance of obstacles and looking for items- on a daily basis. The idea of using a visual prosthesis to restore a rudimentary form of vision for the late blind has been posed for many years (Brindley & Lewin, 1968; W. H. Dobelle et al., 1974; Schmidt et al., 1996), yet recently gained a lot of momentum (see recent reviews by (Farnum & Pelled, 2020; Nowik et al., 2020; Roelfsema, 2020)). In essence, a part of the visual pathway is replaced by a brain-computer-interface (or visual prosthetic implant) that typically translates a camera feed into electrical signals which are conveyed to the brain via electrodes. The passage of current via electrodes placed on or inside the brain activates neurons and elicits dot-like visual sensations (known as phosphenes). These visual sensations are evoked even in the absence of visual stimuli and, importantly, do so in the same degree for sighted and blind subjects (Brindley 1972; Brindley and Lewin 1968a; Dobelle and Mladejovsky 1974; Dobelle et al. 1974; Lee et al. 2000; Maynard 2001; Schmidt et al. 1996). The brain's visual system is retinotopically organized, which means that the neurons respond selectively to a limited region of the visual field (their receptive field). These regions are predictably organized to represent the visual world. Therefore, a consistent pattern of phosphenes can be evoked by stimulating several electrodes simultaneously (X. Chen et al., 2020; Wm H. Dobelle et al., 1976). Phosphene patterns created in real-time from a camera feed can help late blind individuals recover some of the visual functions or abilities that were lost or severely impaired after becoming blind. While theoretically sound, there are several practical challenges that need to be solved before visual prostheses can reach their full clinical potential.

One important issue researchers have to deal with, is the trade-off between implant functionality and invasiveness. In the 70s, Dobelle et al. (1974) were the first to connect a camera system to a set of surface electrodes on the medial side of the occipital lobe. Amazingly, they showed that a blind volunteer could read phosphene 'braille' at a much faster rate compared to tactile braille (Dobelle et al. 1976). The surface electrodes (subdural) used by Dobelle, however, could not yield the number of phosphenes, nor the contiguous distribution required for more advanced functional phosphene vision like navigation or face recognition (Ruyter van Steveninck et al., in prep). Compared to state-of-the-art electrodes, Dobelle's were large and far away from target neurons, resulting in activation of extensive volumes of brain tissue and therefore producing a limited spatial

phosphene resolution. In addition, the limitations of Dobelle's setup are inherent in later reports of system failures, infections and seizures (Naumann, 2012). If visual prostheses are to benefit blind patients in everyday life activities, the acquired functionality should outweigh the possible hazards of such an invasive device. Intracortical electrodes offer several advantages compared to surface electrodes. For one, the current required for phosphene generation is two orders of magnitude lower because the stimulation electrodes are much closer to the target neurons (Schmidt et al., 1996). Higher currents can lead to safety issues (e.g., seizures) and generate larger phosphenes, therefore requiring larger interspacing when attempting to create phosphene patterns by simultaneously stimulating multiple nearby electrodes. In addition to a higher spatial resolution along the cortical surface, intracortical electrodes can also reach neurons in areas that are contained within folded sulci, practically unreachable by surface electrodes. In spite of this, the amount of tissue disturbed (invasiveness) might become an issue when targeting higher eccentricities within V1, located deep inside the calcarine fissure (Wandell et al., 2007).

The next challenge is related to resolution and complexity of phosphene patterns required for functional phosphene vision, which will determine the number and distribution of electrodes necessary for a functional prosthesis. With an increasing number of electrodes, the phosphene patterns can become increasingly complex. Recent developments in image processing techniques, allow the most relevant features of a scene to be intelligently converted into optimal phosphene patterns so that useful visual information is efficiently communicated. In theory, this allows crucial every-day life activities such as accurate emotion expression recognition (Bollen et al. 2019), navigation (Wang et al. 2008; Perez-Yus et al. 2017; Vergnieux et al. 2017), object recognition (Lu et al. 2014; Macé et al. 2015; Xia et al. 2015; Li et al. 2018; Sanchez-Garcia et al. 2018) and even motion detection (Perez-Yus et al. 2017; Chen et al. 2020) to be reestablished after vision has been lost. For example, an end-to-end convolutional neural network (CNN) was recently developed to optimize prosthetic vision in a navigation task (Ruyter van Steveninck et al., in prep). Furthermore, Lozano et al. (2020) developed and implemented Neurolight, a coherent framework for a deep learning-based neural interface for cortical visual prostheses, able to integrate state-of-the-art deep learning visual processing algorithms and computational neural models of artificial retinas.

Despite these developments, most current phosphene simulator studies assumed high density, uniformly spaced, full field phosphene maps (Avraham et al., 2021; Bollen et al., 2019; S. C. Chen et al., 2009; Sanchez-Garcia et al.,

2020; Steveninck et al., 2020; J. Wang et al., 2021). The density and location in which individual phosphenes can be evoked in visual space by a prosthesis is described by a so called "phosphene map" (PM), see figure 3.2.1D. The perceptual features of a single phosphene are related to the stimulation parameters, the design of the electrode and the characteristics of the underlying tissue. State-of-the-art prostheses, however, only cover a small portion of the visual field (Niketeghad and Pouratian 2019; Fernandez Jover et al. in preparation). Prospective users of a visual prosthesis likely require different spatial configurations of visual percepts (i.e. phosphene maps) for specific daily life activities. For example, a dense foveal coverage would be crucial during reading, or for recognizing an object in front of you, while peripheral vision is important for context awareness during navigation. Importantly, these idealized phosphene configurations are unlikely to match reality. Recently, Fernandez Jover et al. (in preparation) introduced a 96-channel cortical prosthesis (Utah-array) in a human subject, proving a safe and reliable interface for a period of 6 months. However, the 88 phosphenes evoked only covered up to 4 visual degrees eccentricity of the lower left part of the visual field. This phosphene pattern is still fundamentally different from typically simulated phosphene vision. In order to study the functional properties of phosphene vision in a realistic configuration, a simulation of stimulation-evoked phosphene sizes and locations based on individual anatomy is required to obtain phosphene configurations closer to idealized maps used in functional phosphene simulation studies.

Finally, scientists and neurosurgeons face practical limitations in surgical electrode placement, phosphene fusion due to insufficiently local stimulation (current spread), and unwanted axonal stimulation creating activation in distal sites. An ideal phosphene coverage throughout the visual field would likely require many stimulation sites in a brain structure that encodes the entire visual field. Candidate structures are the retina, lateral geniculate nucleus (LGN) and cortical areas V1 to higher order regions like hMT. Targeting multiple sites across distant areas will not always be feasible due to surgical limitations. Retinal electrostimulation, while less invasive compared to a (sub) cortical implant, has a narrow stimulating current range between retinal neuron excitation and inhibition which may lead to inferior performances of visual prostheses due to interference between stimulating electrode sites (Barriga-Rivera et al., 2017). LGN is a small thalamic structure deep in the brain and would require long electrode shanks with a high number of extremely densely placed electrodes at the tip. In addition, a study suggested that when two electrodes are too closely spaced in area V1, evoked phosphenes will likely merge (Schiller et al., 2011).

Similarly, the spatial resolution of a phosphene pattern evoked with LGN stimulation might be limited while the amount of affected tissue is large. An interesting candidate for a visual prosthesis is primary visual cortex (V1), the first cortical region in the visual processing stream. Importantly, the functional organization of basic visual features such as columnar orientation selectivity and color selectivity, is well established. When targeting a specific region of the visual field, V1 allows more lenient hardware constraints in terms of electrode spacing, since it has the largest cortical surface area. In contrast to higher visual areas, future technical advancements might allow for stimulation of single orientation or color columns, leading to percepts beyond simple white dots of light. In addition, providing artificial stimulation in a lower part of the visual hierarchy might allow higher level areas to continue to perform complex visual processes, such as movement (Salzman et al., 1990), and face perception (Mundel et al., 2003).

As a foundation for a visual prosthesis in humans, Chen et al. (2020) proved the feasibility of a 1024 channel count cortical prosthesis in non-human primates, contributing to the design of a human implant. Compared to other primates, human cortex has a lot more gyrification and great individual differences in surface area between visual cortices (Benson et al., 2021). The functional organization of early visual cortex, however, has shown to be predictable across individuals as long as the anatomy is known (Rosenke et al. 2021, Benson et al. 2018, Wang et al., 2015). The ability to derive function from anatomy is especially important in the blind, as conventional (visual) localizers based on visual input cannot be used. The functional organization of an individual's brain, together with the electrode design and placement, will ultimately determine the phosphene patterns a visual prosthesis can evoke and the extent of visual function recovery it can establish. Therefore, it is crucial to find suitable insertion angles and implant locations to assure maximal functional benefits while minimizing surgical risks. This comprises a difficult challenge without prior knowledge of individual (functional) anatomy.

Here, we present a pipeline for the exploration and optimization of electrode placement that uses the individual brain anatomy to predict stimulation-evoked phosphene sizes and locations. Within predetermined practical constraints, our pipeline automatically finds the electrode configuration that optimally matches a pre-set 'ideal' phosphene distribution. Because the pipeline uses the individual brain anatomy as a starting point, it can be applied in blind subjects for whom anatomical brain scans are available. The optimal location and insertion angles of the electrodes for a visual cortical prosthesis are based

on a Bayesian optimization that efficiently minimizes a cost function which considers the electrode yield in grey matter, visual field coverage of the phosphene distribution, and the relative entropy between the desired phosphene distribution and the predicted phosphene map. In the implementation we present here, model parameters were optimized for a thousand-electrode array implanted in primary visual cortex, but the procedure can easily be used with other electrode designs or extended to other brain areas as well. We test our method on 362 human hemispheres using anatomical and retinotopy data from the Human Connectome Project 7 Tesla retinotopy dataset (Benson et al., 2018). The software and example data will be made publicly available via github soon.

## 3.2 Methods

### 3.2.1 Preprocessing of the Human Connectome Project 7T retinotopy dataset

T1-weighted (T1w) and T2-weighted (T2w) structural scans at 0.7-mm isotropic resolution were processed using the FreeSurfer image analysis suite (version 7.2; http://surfer.nmr.harvard.edu). Subject brains were inflated and aligned to FreeSurfer's anatomical fsaverage atlas. The inferences made in this work are based on the retinotopic data described by Benson et al., (2018). In brief, MRI data was acquired using a Siemens 7T Magnetom actively shielded scanner and a 32-channel receiver coil array with a single channel transmit coil (Nova Medical, Wilmington, MA) at a 1.6mm isotropic resolution and 1s TR (multiband acceleration 5, in-plane acceleration 2, 85 slices). The population receptive field (pRF) maps describe the location and the size of the receptive field for each 1mm isotropic voxel (see Benson et al. (2018) for descriptions of pRF models used). pRF surface maps based on empirical data in Freesurfer fsaverage space were warped to individual space using Freesurfer's mri_surf2surf function. Bayesian inference of the retinotopic maps was performed using Neuropythy's register_retinotopy command (https://github.com/noahbenson/neuropythy). This procedure harmonizes the anatomical inference of the pRFs, the Benson 2014 atlas (Benson et al., 2014) and experimental data (Benssson et al. 2018) comprised of retinotopic responses to visual stimuli up to 8 degrees eccentricity. Note that the Benson 2014 atlas predicts maps up to 90 degrees eccentricity, however only the inner 20 degrees of eccentricity in V1, V2, and V3 have been validated.

### 3.2.2 Pipeline for optimization of electrode placement

We set out to minimize the difference between a desired phosphene distribution and the predicted phosphene map inferred after electrode placement by calculating the loss with a Bayesian search algorithm. An example of the electrode design (and its simplification) used for the simulations is displayed in figure 3.2.1A. The main input variables of the optimization function were angles alpha (pitch) and beta (yaw), which define the trajectory of the virtually inserted implant. The insertion trajectory was determined by finding the intersection between the insertion point on the cortical sheet and the centroid of the calcarine sulcus (CS) at the angles alpha and beta. The centroid, or geometric center, was calculated using the medians along the three dimensions of the CS volume. Importantly, the CS is a reliable estimate of the location and total volume of the human primary visual cortex (Gillesen & Zilles, 1996). To test whether

this statement holds for the HCP dataset, the CS volume was determined by Freesurfer's cortical parcellation (Desikan-Killiany Atlas) and compared to the volume of V1 (comprised of voxels in the cortical ribbon parcellated by the pRF model), see figure 3.3.1. The range of insertion angles were restricted so that the insertion trajectory cannot intersect with the other hemisphere and excludes unfeasible surgical approaches (e.g. anterior - posterior direction).

For many combinations of angles which are antiparallel to the cortical sheet, a portion of the electrode grid would be placed outside of cortex. In these cases, the configuration was flagged as invalid, and a cost penalty was assigned in the optimization procedure. An extra parameter (shank offset) was added to the optimization function to enable the distance to vary between the insertion point and the first contact point. To vary cortical depth, the final parameter shank length was included to the search for optimal electrode coverage, density and yield. We constrained electrodes to not exceed a cortical depth of 8cm, as the average width and length of HCP brains are about 17cm and 13cm (Yang et al., 2020). Furthermore, the search can be biased towards a specific region of interest. Here, we chose to add a cost penalty when contact points were located outside of primary visual cortex. By targeting V1, a denser phosphene map should be feasible as neurons in V1 have a smaller receptive field size compared to higher visual areas (Benson et al., 2021). Moreover, the interactions between phosphenes created by simultaneously stimulating in different visual areas are not yet sufficiently understood.

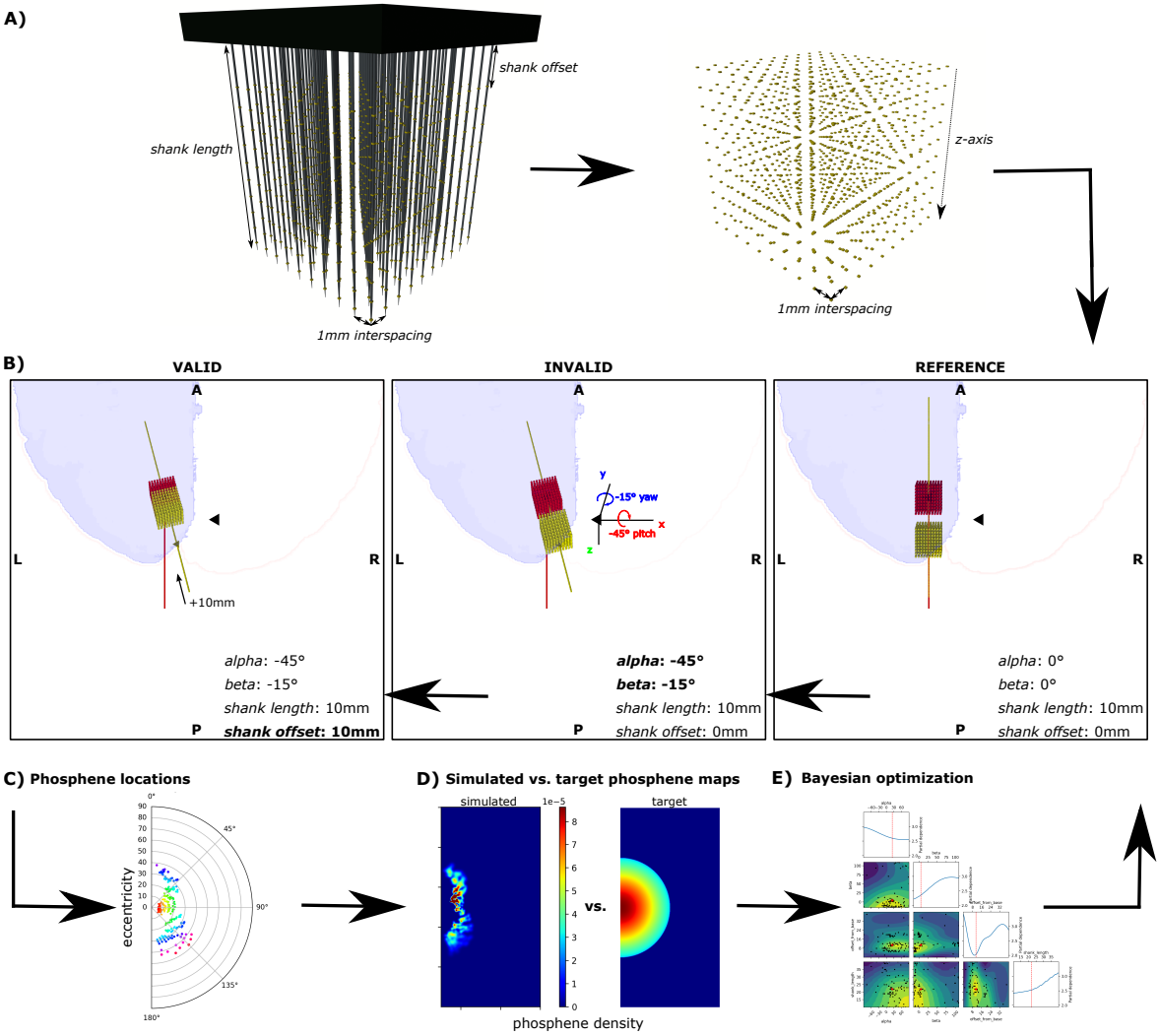



**Figure 3.2.1: Overview of electrode optimization pipeline.** A) example electrode-grid configuration (left) and corresponding simplification of conact points (right). B) The red grid only serves as a reference and is centered on the center of the left calcarine sulcus (black triangles). The implant position (yellow grid) is calculated based on alpha and beta (relative to the reference grid). For a new set of parameters, the resulting configuration can be either valid (left) or invalid (center) when contact points are located outside of the brain. In the 'valid' example, the first contact point is located 10mm from the point where the shank penetrates the cortex. C) Each contact point potentially evokes a phosphene in the polar angle plot (left). The individual phosphenes are modeled as 2-D Gaussians on a n-by-m *phosphene map*. Color codes for eccentricity. D) Sorensen dice coefficient is computed by comparing the binarized version (not shown) of the phosphene map with one of the binarized target phosphene maps. Relative entropy is calculated between the probability distributions of the simulated phosphene map (left) and the target phosphene map (right). E) The Bayesian search algorithm determines the next set of parameters and the process is repeated until the optimal set of parameters is found for a specific target phosphene configuration.

For each set of parameters, the electrode-grid is repositioned to match the trajectory set by the insertion angles. The resulting phosphene map can be created based on the coordinates of the electrodes and the matching voxels of the retinotopic map. Each individual electrode intersecting a voxel in the retinotopic map yields a phosphene on a 1000-by-1000-pixel image-grid. The spread of stimulation current is assumed to be relative to voxel size (1mm3). Depending on electrode spacing it may happen that two electrodes obtain the same pRF properties when they are located in the same voxel. In our simulations, we assume that tissue around each contact point will largely remain intact and phosphenes are modeled as 2-D Gaussian circular spots of light with standard deviation ranging from 0.2 up to 3 visual degrees, depending on receptive field size (Tehovnik et al., 2005; Lin Wang et al., 2008). Finally, the process is repeated for a new set of parameters up to 100 times (separately for each hemisphere), or when the total loss reached a certain threshold (early stopping). The entire procedure is guided by the images in figure 3.2.1 and the flowchart diagram in figure 3.2.2.

### 3.2.3 Target phosphene maps

The choice of target visual field coverage is a configurable component of our pipeline. It might be the case that specific visual tasks require targeting of specific subregions of the visual hemifield. We tested our pipeline on four distinct target visual field coverages by specifying phosphene maps composed of large partial 2-D Gaussians: 1) inner target: dense representation starting from the fovea, covering 0-45° eccentricity. 2) outer target: somewhat dense representation in the periphery, covering 45-90° eccentricity. 3) upper target: dense representation around the fovea, covering 0-90° eccentricity and 0-45° polar angle of the upper visual field. 4) lower target: dense representation around the fovea, covering 0-90° eccentricity and 135-180 ° polar angle of the lower visual field (see figure 3.2.1D and figure 3.3.8 for reference).

### 3.2.4 Bayesian optimization

Bayesian optimization has become an attractive method to optimize expensive to evaluate black box, derivative-free and possibly noisy functions (Shahriari et al. 2016). The algorithm is executed in an iterative fashion and evaluates a probabilistic model for which a cheap probability function f based on the posterior distribution is optimized before sampling the next point. The function objective is considered as a random function (a stochastic process) on which a

prior is placed. In our case, the prior is defined by a Gaussian process capturing our beliefs about the function behavior. Function evaluations are treated as data and used to update the prior to form the posterior distribution over the objective function. The convergence of the optimization algorithm was accelerated by setting an initial sampling point for which it was known that the resulting grid would hit some portion of V1. We chose to position the center of the electrode grid at the centroid of the calcarine sulcus with initial parameters 0° alpha, 0° beta and 20mm shank length and 25mm shank offset. However, the model can also be run without prior knowledge of the objective function.

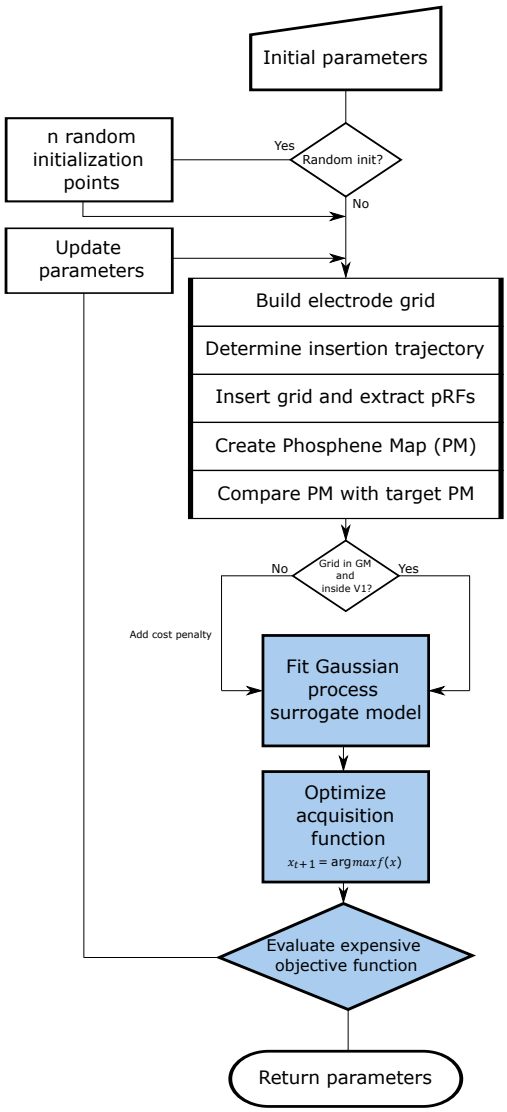



**Figure 3.2.2: flowchart diagram Bayesian optimization pipeline.** This diagram provides a detailed overview of the electrode placement pipeline. Rectangular boxes indicate processing steps and diamond shapes indicate model decisions.

### 3.2.5 Loss functions

Three loss functions (and their linearly weighted combinations) were proposed to minimize the difference between a desired phosphene map and the phosphene map derived from pRF parameters corresponding to 1000 electrode locations in the brain. For combinations of loss, their values were first normalized and then linearly combined. We chose to weigh the contribution of loss functions equally, however depending on the desired outcome, weights can either be set by the electrode designer, or optimized as an additional hyperparameter. Next, we will explain the loss functions used and their meaning in the context of the optimization goals.

### 3.2.6 Sørensen–Dice coefficient (dice)

The dice coefficient can be seen as a measure of overlap between two sets. It was computed on two discretized sets of data X and Y, where X is the binarized target phosphene map and Y is the binarized version of the predicted phosphene map. Each map's pixels are set to 1 if they contain phosphene activation and 0 if they do not. |X| and |Y| represent the number of elements in each set. The Sørensen index equals twice the number of elements common to both sets divided by the sum of the number of elements in each set. Dice is included to obtain phosphene maps that are localized in the desired visual region.

$$DC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{3.1}$$

### 3.2.7 Yield

The ratio between the total number of electrode contact points and voxels within grey matter in which a contact point is placed and for which retinotopic data is available.

$$Y = \frac{\text{number of hits}}{\text{number of contact points}} \tag{3.2}$$

The meaning of the yield loss its straightforward: we will reward solutions where the maximum number of electrodes can evoke a phosphene, thus allowing the implant to achieve a high phosphene count. Contact points outside of the targeted region were penalized.

### 3.2.8 Kullback–Leibler divergence (KL or relative entropy)

KL describes the difference between two probability distributions over the same variable x. The Kullback–Leibler divergence, (also called relative entropy), is a measure of how one probability distribution is different from a second, reference probability distribution. The KL loss term will reward parameter sets for which the virtually implanted electrodes yield a phosphene map with the desired density distributions, and penalize density distributions that diverge from our target map. This is important, since phosphene density is likely to be a key element for pattern recognition.

$$D_{KL}(P \vee Q) = \sum_{x \in X} P(x) \log \left( \frac{P(x)}{Q(x)} \right) \tag{3.3}$$

### 3.2.9 Equality metric (EM)

The relative importance of loss functions can be set by weighing them accordingly. However, often the relationship between loss functions and their influence on the optimization process is difficult to understand. For a better understanding of the relative contribution of each loss function to the total cost, a measure of equality can be computed which summarizes the distance of each loss function's contribution to its fair share. EM is the mean standard deviation from the equal contribution, and can be calculated as follows:

$$EM = \sqrt{\frac{\sum (X-a)^2}{n}} \tag{3.4}$$

Where X is the relative contribution (percentage) of all loss functions to the sum of loss, $n$ is the total number of loss functions, and a is the equal contribution (1 / $n$ / 100) to the sum of loss. Loss functions contribute equally to total loss when EM is zero.

### 3.2.10 pRF polar density estimation

A non-parametric kernel-density estimation (KDE) was performed on the group average PMs to estimate probability density functions (PDFs). The group average PMs were created by summing simulated phosphenes of the optimized electrode locations in all subjects. PMs were created separately for each region of interest (GM, V1, V2, V3) before KDE was computed using SciPy 1.0 (Virtanen et al., 2020). For visualization purposes, a less convoluted overview of phosphene density is achieved by coloring the pRF locations of their corresponding phosphenes in the polar plots in figure 3.3.7, based on the PDF. Density was scaled by the number of samples to allow for a fair comparison between regions of interest.

### 3.2.11 Group average electrode configuration

Up to this point, we assumed that individual differences in anatomy necessitate an individual approach, i.e. determine the best electrode configuration for each person individually. However, perhaps an electrode configuration based on the average brain would suffice. To test this hypothesis, we applied the electrode optimization pipeline to the fsaverage brain and the group average retinotopy. Phosphene maps obtained with the 'average' parameter estimates and 'individual' parameter estimates were compared to establish the potential benefit of the individual optimization approach (see figure 3.3.9).

## 3.3 Results

Optimal electrode placement was determined in both hemispheres of 181 subjects from the HCP 7T dataset to obtain stimulation-evoked phosphene sizes and locations based on individual brain structure and match the desired phosphene distributions as closely as possible. First, we justify why the CS is a reliable landmark (and thus a good starting point) for optimizing electrode placement in V1 (Fig. 3.2.1). Next, the results of the optimization pipeline are illustrated using data from an example subject (Fig. 3.3.2 and 3.3.3). Then, an overview of the optimal parameters across subjects and their corresponding phosphene locations is presented (Fig. 3.3.4 to 3.3.7). Finally, we assess phosphene distributions when the electrode configuration is optimized in each individual versus a more general approach where the electrode configuration is based on the optimal solution for the group average brain (fsaverage), see figures 3.3.8 and 3.3.9.

### 3.3.1 Calcarine sulcus volume is a reliable estimate for primary visual cortex volume

The centroid of the CS was chosen as the reference and start location (where alpha and beta angles are zero) for the Bayesian search. We tested whether the size of the CS correlates with the size of V1 and found that Pearson correlation in 181 individuals was nearly perfect for the left hemisphere (r = 0.95) and very high for the right hemisphere (r = 0.87).

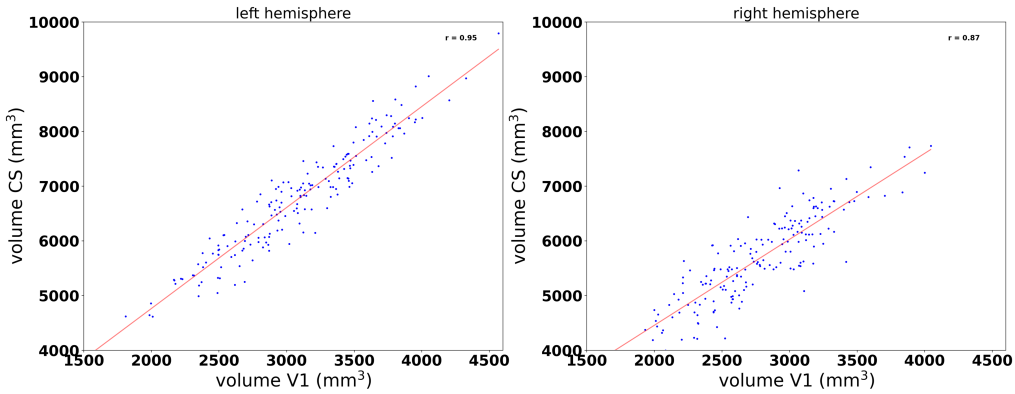



**Figure 3.3.1: volume comparison between CS and V1 for the left (left) and right (right) hemisphere.**

### 3.3.2 Bayesian optimization for an example subject

Here, the optimization pipeline is further illustrated using an example subject. Figure 3.3.2A demonstrates the influence of each search-space dimension on the objective function. In figure 3.3.2B, the order in which the hyperparameter space was sampled (indicated by color) is shown when optimizing electrode placement in the left hemisphere for a PM with a high phosphene density in the foveal region. The process was repeated for 100 iterations.

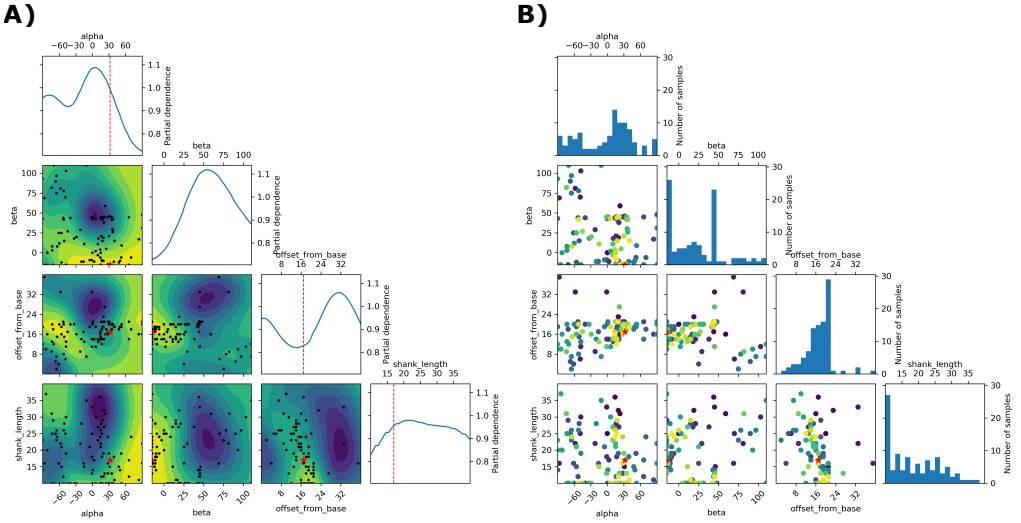



**Figure 3.3.2: the Bayesian optimization process in an example subject.** A) The influence of each search-space dimension on the objective function. The diagonal shows the effect of a single dimension on the objective function, while the plots below the diagonal show the effect on the objective function when varying two dimensions. The red stars indicate the best observed minimum. B) Each point's color indicates the order in which samples were evaluated (blue to yellow), and a red star shows the location of the best-found parameters. On the diagonal the frequency distributions of samples for each search-space dimension are shown. The optimal loss for this particular subject was 0.412 for dice, 0.334 for yield and 0.005 for KL. This corresponds to 58,8% overlap between best and target PM (dice), and 66,7% of the electrodes being located inside the cortical ribbon (yield).

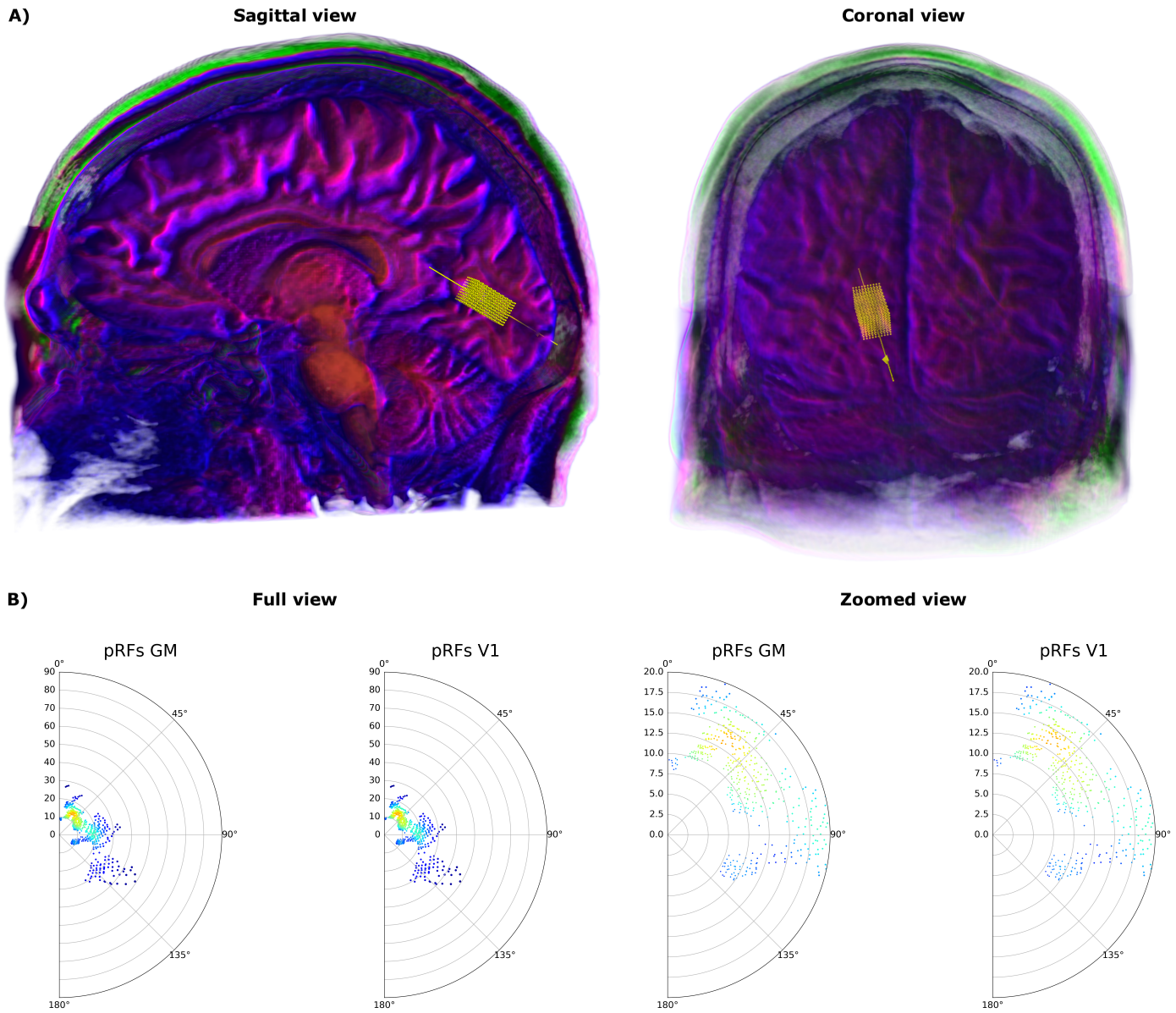



**Figure 3.3.3: electrode configuration optimized for the inner target PM in an example subject.** A) Position of electrode grid relative to the subject's anatomy with parameters: [alpha: 32°, beta: -15°, shank offset: 17mm, shank length: 22mm]. B) pRF parameters (location and size) corresponding to 1000 simulated electrode locations. Nearly all functional electrodes are located within V1.

### 3.3.3 Optimization results per combination of loss function

The optimization process was repeated for all possible combinations of loss functions. In figure 3.3.4A an overview of model loss per combination of loss function on the group level is displayed. Individual PMs for each combination of loss were compared with the target PMs. Optimization performance is assumed to be high when loss is low for dice, yield and KL, and when a low score on one metric does not go at the cost of another (indicated by the equality metric (EM) on top of each bar). KL was relatively well optimized in the inner target compared to the other target PMs. Average loss was smaller for the inner target across loss functions [stats], compared to the outer target. Overall, the combination of dice, yield and KL resulted in the lowest loss scores.

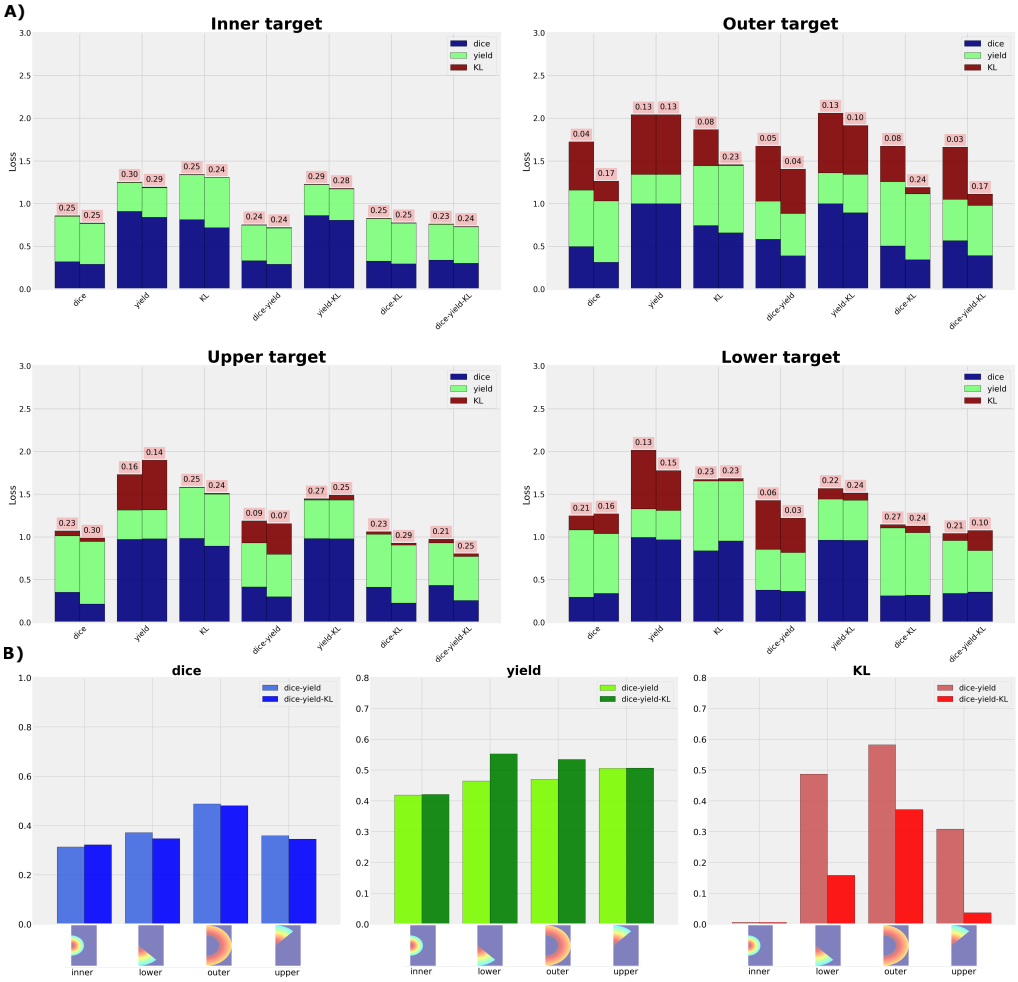



**Figure 3.3.4: phosphene map loss comparison.** A) Loss between optimized PMs and target PMs per combination of loss functions were averaged across subjects. On top of each bar the equality metric indicates the relative contribution of loss functions to the total loss. B) Average loss for combinations dice-yield and dice-yield-KL. When only optimizing for dice and yield, KL is sacrificed at the cost of yield (for lower, outer and upper target PM).

### 3.3.4 Cost and parameter distributions across subjects when equally weighting loss functions

A summary of the best selected (optimal) models per subject is plotted in figure 3.3.5 and 3.3.6. The histograms show the distributions for loss (A) and the best parameters (B) across subjects. Group-average coverage for the 181 subjects (dice) of the target regions was 57.4% (± 1.4% SE). On average 48.5% (± 0.1 SE) of the electrode locations were located inside grey matter (yield). Relative entropy was on average 0.17 (± 0.1 SE) and varied greatly between target regions.

Mean alpha was 3.91, 7.25, -8.51 and 16.57 for the inner, outer, upper and lower targets respectively in the left hemisphere and -8.88, 7.71, -14.34 and 12.13 in the right hemisphere. Similarly, mean beta was -4.62, -7.10, -8.32, -8.58 in LH and 8.48, 3.72, 4.05, 5.56 in RH. Mean shank offset was 27.62, 18.73, 28.97, 30.70 in LH and 33.31, 19.03, 33.85, 25.33 in RH. Finally, mean shank length was 16.75, 22.71, 18.55, 12.98 in LH, and 13.80, 19.86, 14.61, 16.43 in RH. Based on the organization of the visual cortex, we would expect a sign inversion for alpha angles when optimizing either for the lower or upper visual field. Similarly, beta angles are expected to be reversed for left and right hemisphere. These reversals can be seen in figure 3.3.3B. Moreover, phosphene coverage in higher eccentricities within V1 should require higher cortical depth, which can be seen when comparing the offset from GM-CSF between the inner and outer PM targets.

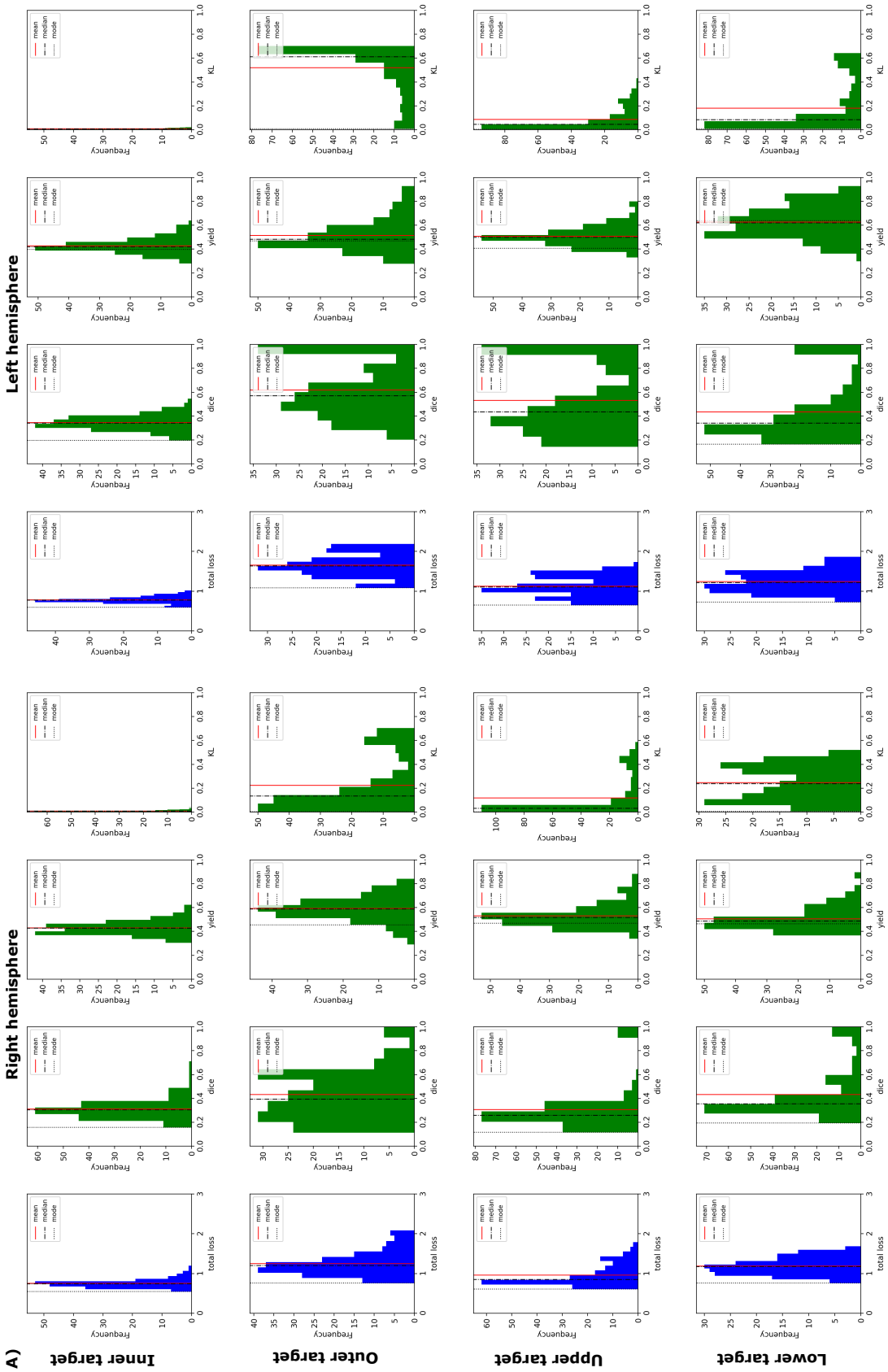



**Figure 3.3.5: Bayesian optimization cost distributions.** A) Histograms of optimal loss (when dice, yield and KL are linearly combined and equally weighted) across subjects. Overall loss is smaller for the inner compared to the outer PM target, and smaller for upper compared to lower PM target.

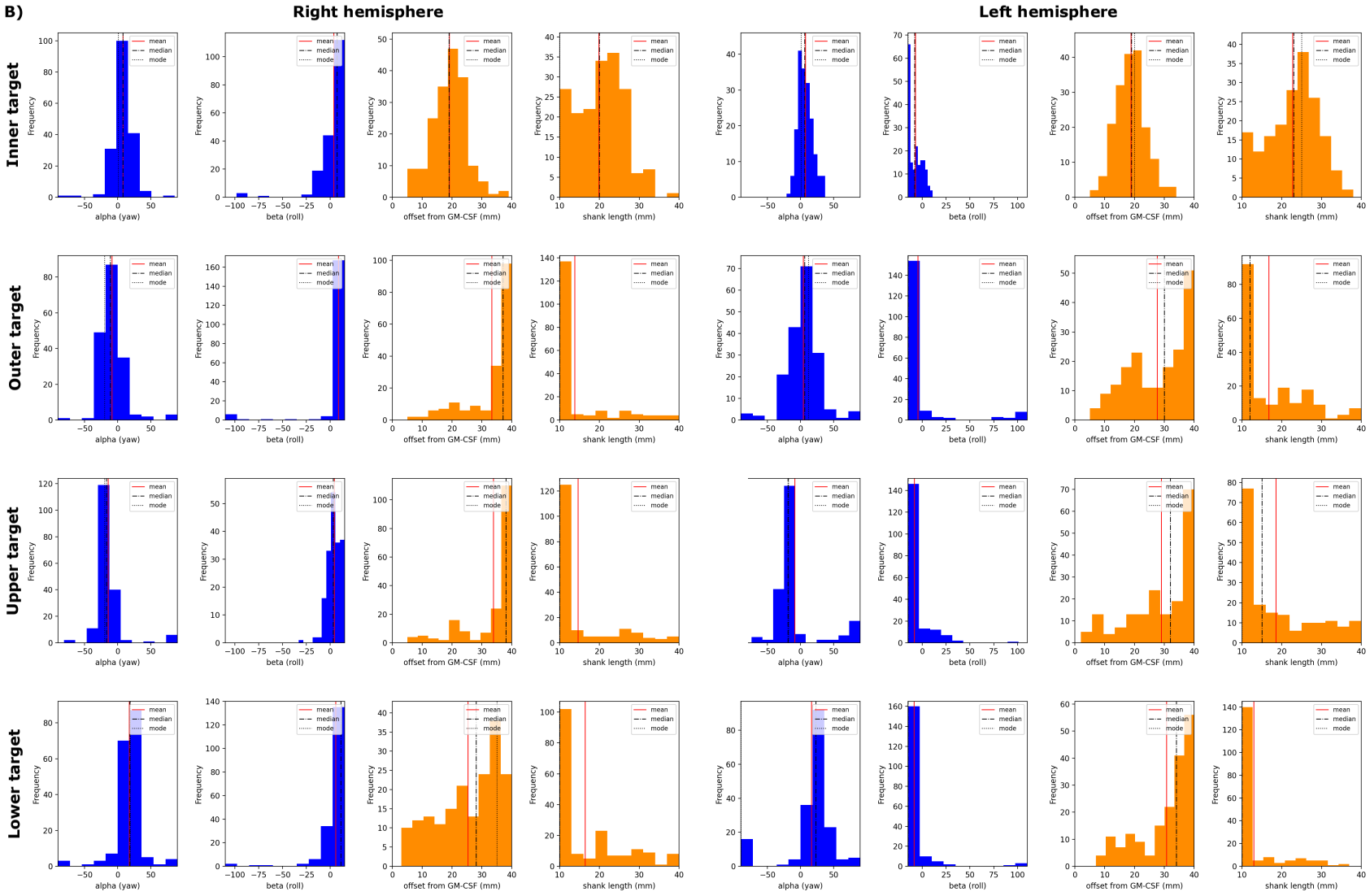



**Figure 3.3.6: Bayesian optimization parameter distributions.** B) Histograms of selected parameters across subjects. An inversion of selected beta is visible between left and right hemisphere. An inversion for alpha is seen between upper and lower PM targets.

### 3.3.5 pRF density distributions

An alternative way to illustrate between-subject variability is to evaluate group average probability density functions of phosphenes across the visual field. In figure 3.3.7, the PDFs are shown per target phosphene map across the regions of interest. For the inner and outer target most phosphenes were located in the upper quadrant of the visual field, perhaps due to the initial position of the Bayesian search. Phosphenes obtained for the upper and lower target maps were mostly located in the expected visual quadrants, however relatively many phosphenes were positioned in V2 and V3. This indicates that for these target regions, frequently no solution could be found where all 1000 electrodes are contained by the V1 label. Note that variability of the individual distributions is not well reflected in the group-average. Importantly, the individual PMs (example shown in figure 3.3.3) were highly variable and non-uniformly distributed, even for similar overlap, number of phosphenes and relative entropy.

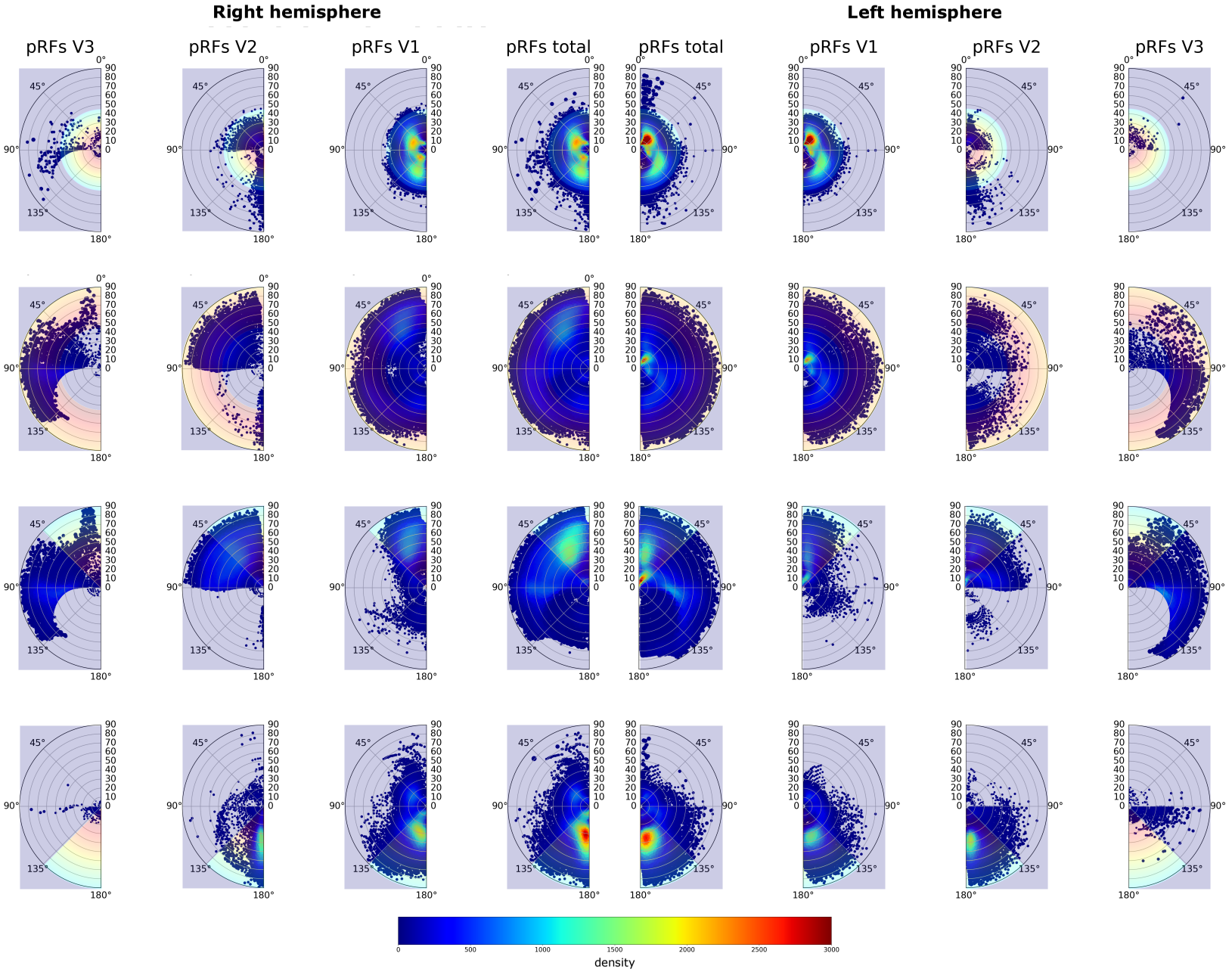



**Figure 3.3.7: group average pRF density polar plots (shown as polar angle vs. eccentricity).** The center plots show a summary of all pRF locations in either left or right hemisphere after optimizing electrode configurations for dice, yield and KL in all subjects. The color indicates the density (as determined by Gaussian KDE) of simulated phosphenes in visual space. As a reference, the target PM is shown on the background of each plot.

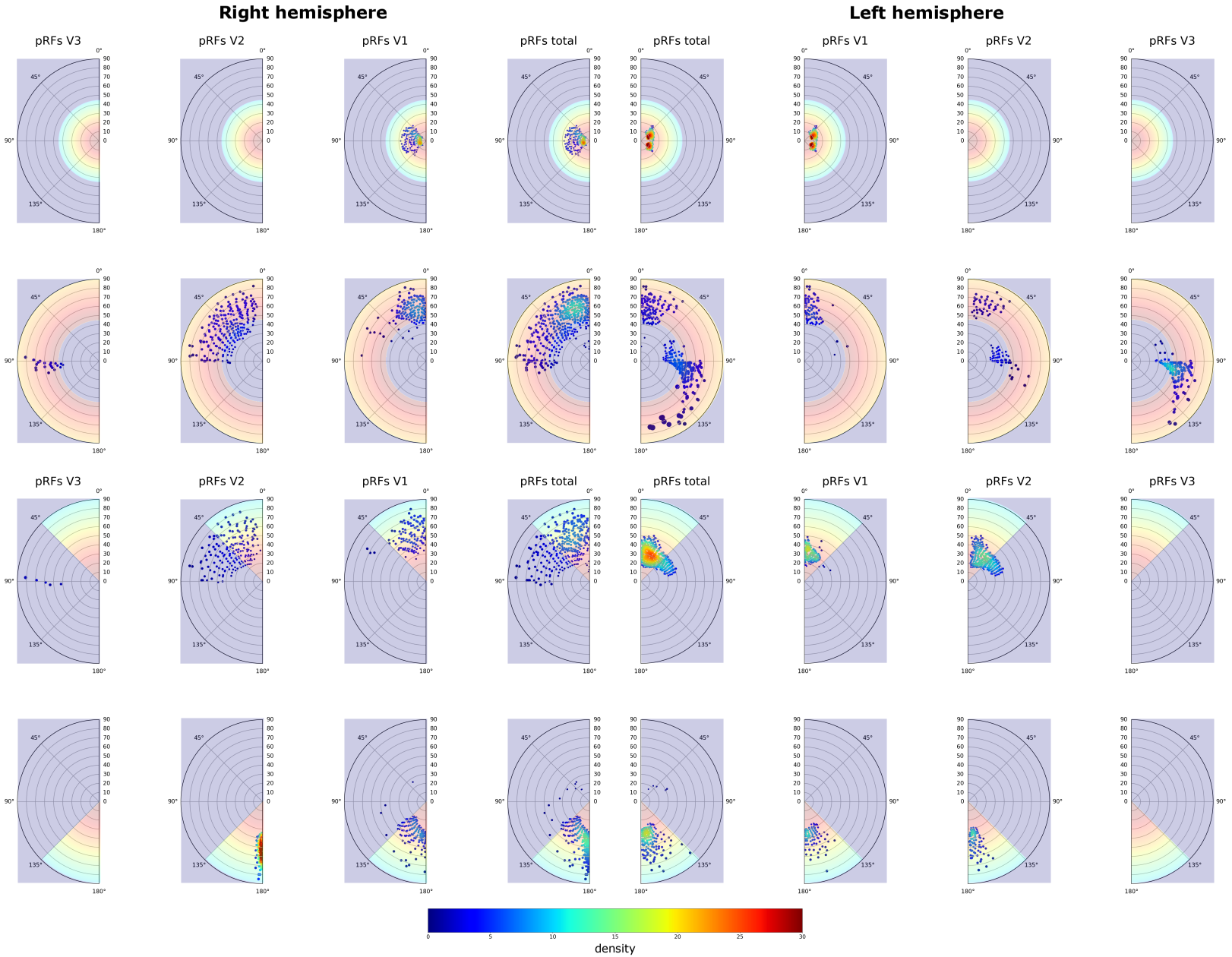



**Figure 3.3.8: pRF density polar plots in the fsaverage brain.** Results after running the pipeline on subject average pRF maps projected on the fsaverage brain. For reference, the target PM is shown on the background of each plot.

### 3.3.6 Group average estimation versus individual configuration estimation

A less computationally expensive way to design an electrode grid would be to take the ideal electrode configuration for the average brain and virtually implant this electrode grid in all individuals. In figure 3.3.9 the loss between PMs resulting from this group average approach and the results from the individualized approach are compared. There was no difference in simulated phosphene yield, however phosphene coverage (dice) and relative entropy (KL) was significantly better for the approach where the electrode configuration parameters were optimized for each individual.

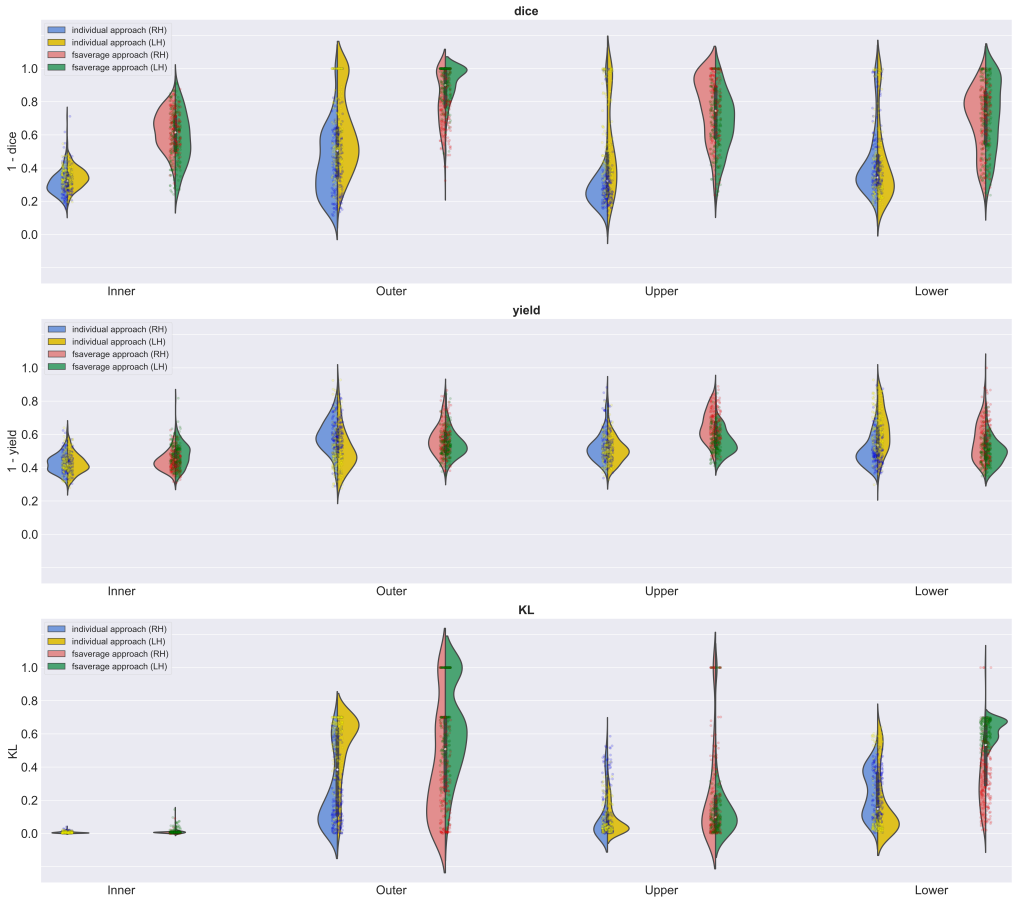



**Figure 3.3.9: loss comparison between fsaverage-based and individually optimized electrode configurations.** Each half of the violin plots shows the loss of dice (top graph), yield (center graph) and KL (bottom graph) for all individuals per right (RH) or left (LH) hemisphere. Each dot on top of the violins represents the loss for a single individual. Configurations based on individually optimized parameters result in significantly higher phosphene coverage (dice) and more accurate density distributions (KL).

## 3.4 Discussion

A visual prosthesis that interfaces directly with the brain may one day become a conventional clinical treatment for blindness when other treatments are not available. However, until that time comes, several (technical) challenges need to be overcome. Firstly, a prosthetic implant involves an invasive surgical procedure and therefore the potential benefits of having a prosthesis should outweigh the possible risks of getting one. Secondly, how can visual information be optimally transferred in phosphene patterns? Prostheses equipped with more electrodes allow for more complex phosphene patterns while modern image processing and deep learning techniques can convert camera feeds into efficient functional phosphene vision. However, these techniques are currently developed under the assumption of biologically unrealistic phosphene map configurations. Lastly, anatomical structures vary between individuals, which makes it even more difficult for scientists and clinicians to design the optimal electrode configuration for a cortical implant.

The approach we present here can partly solve these issues by simulating stimulation-evoked phosphene sizes and locations based on individual anatomy and optimize electrode configuration and placement to obtain phosphene patterns closer to idealized layouts. We found that the simulated phosphene distribution matches a desired map of phosphenes more closely when optimized for a specific individual's anatomical and functional retinotopic organization, compared to group-averaged data. When the electrodes are optimally placed in the desired cortical area, a prosthesis is more likely to yield functional phosphene vision for the blind patient. Furthermore, our simulation and optimization pipeline can provide realistic phosphene map predictions for any given set of electrode configuration constraints. The range of electrode configuration parameters can be chosen depending on surgical limitations or physical electrode constraints. As a result, phosphene simulator studies can make use of the more realistic phosphene maps to study behavioral performance on specific tasks involving phosphene vision. The knowledge from these studies will in turn lead to better assumptions of what type of phosphene map a prosthesis should ideally try to establish.

In particular, we provide an extensive overview of optimal realistically constrained electrode insertion trajectories, end-positions and their corresponding PMs based on a 1000-channel electrode grid in a large sample of sighted individuals. These findings could be used as a reference for designing safer and more efficient visual prostheses. On the individual level, we found that phosphene

maps rarely have evenly spaced distributions like in the fsaverage brain (see figure 3.3.3B and figure 3.3.8). The fsaverage phosphene distributions are likely more evenly spaced because the pRF data is smoothed due to averaging. The obtained loss distributions in figure 7A show that between-subject variability is high and average coverage of the target regions is 57%. A full target coverage can likely be achieved more closely by inserting multiple (smaller) electrode grids. However, new challenges arise as intersection of insertion angles of multiple grids may be problematic. Our approach serves as a proof of principle and can be extended with more sophisticated multi-grid trajectory planning that would take these constraints into account. Furthermore, about half of the electrodes in our simulations were located inside grey matter with most of the remaining electrodes likely located in white matter. Here, we assumed that white matter stimulation could not be used, but further research is required to understand and simulate phosphenes evoked by white matter stimulation.

The main goal of our work here is not to define the best loss function -this can easily be customized by the implant designer- but to provide a flexible implant optimization framework. Nonetheless, we implemented and described specific combination sets of loss functions and describe model performance on generating optimal phosphene maps. We found that PMs varied greatly depending on the chosen loss terms. For example, when only optimizing for phosphene coverage and number of phosphenes, relative entropy is sacrificed at the cost of the number of phosphenes (Fig. 3.3.4B). In general, optimization results were more successful for the foveal (inner) target region, compared to the more peripheral (outer) target region. This makes sense, considering it takes a longer shank to reach locations along the calcarine with higher eccentricities, which increases the electrode spacing along the shank. In the same context, a relatively large area is dedicated to lower eccentricities because of cortical magnification. An alternative to cover more of the periphery is to target V2 or V3 sites with additional electrodes, since spatial representation is preserved and repeated multiple times across the visual cortex (Wandell et al., 2007). This alternative should be approached with care, as phosphene vision could be complicated by the increase in RF size in these regions compared to V1, and simultaneous stimulation across multiple areas potentially leads to conflicting signals.

Some limitations should be considered when using our optimization approach. The results presented here are based on evenly spaced electrode grids and show that corresponding phosphene maps are often non-uniform. Contiguous phosphene configurations would require custom-built electrode configurations. These custom configurations, however, can also be evaluated or

positioned with minimal changes to the framework. The accuracy of the simulations furthermore depends on the quality and the spatial resolution of the population receptive field mapping. A possible improvement could be to acquire pRF data at a higher resolution, or resample the pRF maps to a submillimeter resolution for more fine-grained PMs and to allow for more accurate submillimeter electrode spacing. In blind persons it is highly recommended to obtain the individual anatomy using an MRI scan whenever possible. Retinotopy can then be estimated based on anatomy and group-average (probabilistic) pRF maps. Our results showed that 'individual' parameter estimates are preferred, yet an average-based approach might be sufficient. Finally, the simulations do currently not take intracranial vasculature into account. The implant insertion trajectory can be determined not only based on the location of the desired cortical area, but also set to avoid larger arteries and veins. Hence, surgical risks can further be reduced and emphasizing the need for structural scans of individual patients.

In conclusion, the challenging task of electrode geometry design and surgical planning for prosthetic implantation can be aided using our simulation of optimization pipeline, which assists researchers, clinicians and neurosurgeons in finding suitable insertion angles and implant locations to assure maximal functional benefits while minimizing surgical risks. The software will be made publicly available via github soon.

## Chapter 4

# Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception

Visual mental imagery is the quasi-perceptual experience of "seeing in the mind's eye". While a tight correspondence between imagery and perception in terms of subjective experience is well established, their correspondence in terms of neural representations remains insufficiently understood. In the present study, we exploit the high spatial resolution of functional magnetic resonance imaging (fMRI) at 7T, the retinotopic organization of early visual cortex, and machine-learning techniques to investigate whether visual imagery of letter shapes preserves the topographic organization of perceived shapes. Sub-millimeter resolution fMRI images were obtained from early visual cortex in six subjects performing visual imagery of four different letter shapes. Predictions of imagery voxel activation patterns based on a population receptive field-encoding model and physical letter stimuli provided first evidence in favor of detailed topographic organization. Subsequent visual field reconstructions of imagery data based on the inversion of the encoding model further showed that visual imagery preserves the geometric profile of letter shapes. These results open new avenues for decoding, as we show that a denoising autoencoder can be used to pretrain a classifier purely based on perceptual data before fine-tuning it on imagery data. Finally, we show that the autoencoder can project imagery-related voxel activations onto their perceptual counterpart allowing for visually recognizable reconstructions even at the single-trial level. The latter may eventually be utilized for the development of content-based BCI letter-speller systems.

## 4.1 Introduction

Visual mental imagery refers to the fascinating phenomenon of quasi-perceptual experiences in the absence of external stimulation (Thomas 1999). The capacity to imagine has important cognitive implications and has been linked to working memory, problem solving, and creativity (Albers et al. 2013; Kozhevnikov et al. 2013). Yet, the nature of mental representations underlying imagery remains controversial. It has been argued that visual imagery is pictorial, with an intrinsic spatial organization resembling that of physical images (Kosslyn et al. 1997, 2006). Others have claimed that imagery resembles linguistic descriptions, lacking any inherent spatial properties (Pylyshyn 1973, 2003; Brogaard and Gatzia 2017). This debate has become increasingly informed by neuroimaging. For instance, several functional magnetic resonance imaging (fMRI) studies have indicated that imagery activates cortical networks that are also activated during corresponding perceptual tasks (Kosslyn et al. 1997; Goebel et al. 1998; Ishai et al. 2000; O'Craven and Kanwisher 2000; Ganis et al. 2004; Mechelli et al. 2004), lending credence to the notion that imagery resembles perception. Applying multi-voxel pattern analyses (MVPA), furthermore, enabled the decoding of feature-specific imagery content related to orientations (Harrison and Tong 2009; Albers et al. 2013), motion (Emmerling et al. 2016), objects (Reddy et al. 2010; Cichy et al. 2012; Lee et al. 2012), shapes (Stokes et al. 2009, 2011), and scenes (Johnson and Johnson 2014).

The MVPA approach has recently been criticized on the grounds that it does not rely on an explicit encoding model of low-level visual features, leaving open the possibility that classification may have resulted from confounding factors such as attention (Naselaris et al. 2015). To overcome this limitation, the authors developed an encoding model based on Gabor wavelets which they fit to voxel activations measured in response to perception of artworks. Subsequently, they used the estimated encoding model to identify an imagined artwork from a set of candidates by comparing voxel activations empirically observed in response to imagery with those predicted from encoding each candidate (Naselaris et al. 2015).

While this study constitutes a major methodological advancement and largely defuses the aforementioned confounds, a complex encoding model allows only for limited inferences regarding the similarity of perception and imagery with respect to any particular feature. It is, for instance, conceivable that the largest contributor to image identification stemmed from an unspecific top–down modulation of salient regions in the imaged artwork with crude retinotopic organization. That is, activations in response to mental imagery might have been co-localized to highly salient regions of the image (without otherwise resembling it) and this might have been sufficient for image identification.

Indeed, results from studies reconstructing the visual field from fMRI data leveraging the retinotopic organization of early visual cortex give the impression that the retinotopic organization of mental imagery is rather diffuse. For instance, while seminal work has been conducted detailing the ability to obtain straightforwardly recognizable reconstructions of perceived physical stimuli (Thirion et al. 2006; Miyawaki et al. 2008; Schoenmakers et al. 2013); similar successes have not been repeated for imagery. Retinotopy-based reconstructions of imagined shapes have so far merely been co-localized with the region of the visual field, where they were imagined but bore no visual resemblance to their geometry (Thirion et al. 2006).

However, unless imagery of an object preserves the object's geometry, it unlikely it would preserve any of its more fine-grained features. It is thus pivotal to empirically establish precise topographic correspondence between imagery and perception. Utilizing the high spatial resolution offered by 7T fMRI and the straightforwardly invertible population receptive field model (Dumoulin and Wandell 2008), we provide new evidence that imagery-based reconstructions of letter shapes are recognizable and preserve their physical geometry. This supports the notion of tight topographic correspondence in early visual cortex. Such correspondence opens new avenues for decoding. Specifically, we show that using a denoising autoencoder, it is possible to pretrain a classifier, intended to decode imagery content, purely based on easily obtainable perceptual data. Only fine-tuning of the classifier requires (a small amount of) additional imagery data. Finally, we show that an autoencoder can project imagery-related voxel activations onto their perceptual counterpart allowing for recognizable reconstructions even at a single-trial level. The latter could open new frontiers for brain–computer interfaces (BCIs).

## 4.2 Materials and methods

### 4.2.1 Participants

Six participants (2 female, age range=(21–49), mean age=30.7) with normal or corrected-to-normal visual acuity took part in this study. All participants were experienced in undergoing high-field fMRI experiments, gave written informed consent, and were paid for participation. All procedures were conducted with approval from the local Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University.

### 4.2.2 Stimuli and tasks

Each participant completed three training sessions to practice the controlled imagery of visual letters prior to a single scanning session which comprised four experimental (imagery) runs of 11 min and one control (perception) run of 9 min as well as one pRF mapping run of 16 min.

#### Training session and task

Training sessions lasted ca. 45 min and were scheduled 1 week prior to scanning. Before the first training session, participants filled in the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and the Object–Spatial Imagery and Verbal Questionnaire (Blazhenkova and Kozhevnikov 2009). These questionnaires measure the subjective clearness and vividness of imagined objects and cognitive styles during mental imagery, respectively. In each training trial, participants saw one of four white letters ('H', 'T', 'S', or 'C') enclosed in a white square guide box (8° by 8° visual angle) on grey background and a red fixation dot in the center of the screen (see Fig. 4.2.1). With the onset of the visual stimulation, participants heard a pattern of three low tones (note C5) and one high tone (note G5) that lasted 1000 ms. This tone pattern was associated with the visually presented letter with specific patterns randomly assigned for each participant. After 3000 ms, the letter started to fade out until it completely disappeared at 5000 ms after trial onset. The fixation dot then turned orange and participants were instructed to maintain a vivid image of the presented letter. After an 18 s imagery period, the fixation dot turned white and probing started. With an inter-probe-interval of 1500 ms (jittered by±200 ms), three white probe dots appeared within the guide box. These dots were located within the letter shape or outside of the letter shape (however, always within the guide box). Participants were instructed to indicate by button press whether

a probe was located inside or outside the imagined letter shape (Podgorny and Shepard 1978). Depending on the response, the fixation dot turned red (incorrect) or green (correct) before turning white again as soon as the next probe was shown. The positions of the probe dots were randomly chosen, such that they had a minimum distance of 0.16° and a maximum distance of 0.32° of visual angle from the edges of the letter (and the guide box), both for inside and outside probes. This ensured similar task difficulty across trials. A resting phase of 3000 ms or 6000 ms followed the three probes. At the beginning of a training run, all four letters were presented for 3000 ms each, alongside the associated tone pattern (reference phase). During one training run, each participant completed 16 pseudo-randomly presented trials. In each training session, participants completed two training runs during which reference letters were presented in each trial (described above) and two training runs without visual presentation (i.e., the tone pattern was the only cue for a letter). Participants were instructed to maintain central fixation throughout the entire run. After the training session, participants verbally reported the imagery strategies they used.

**Imagery runs**

Imagery runs were similar to the training task with changes to the probing phase and the timing of the trial phase. After the reference phase in the beginning of each run, there was no visual stimulation other than the fixation dot and the guide box. Imagery phases started when participants heard the tone pattern and the fixation dot turned orange. Imagery phases lasted 6 s. Participants were instructed to imagine the letter associated with the tone pattern as vividly and accurately as possible. The guide box aided the participant by acting as a reference for the physical dimensions of the letter. The resting phases that followed each imagery phase lasted 9 s or 12 s. There was no probing phase in normal trials. In each experimental run, there were 32 normal trials and two additional catch trials which entailed a probing phase of four probes. There was no visual feedback for the responses in the probing phase (the fixation dot remained white). Data from the catch trials were not included in the analyses.

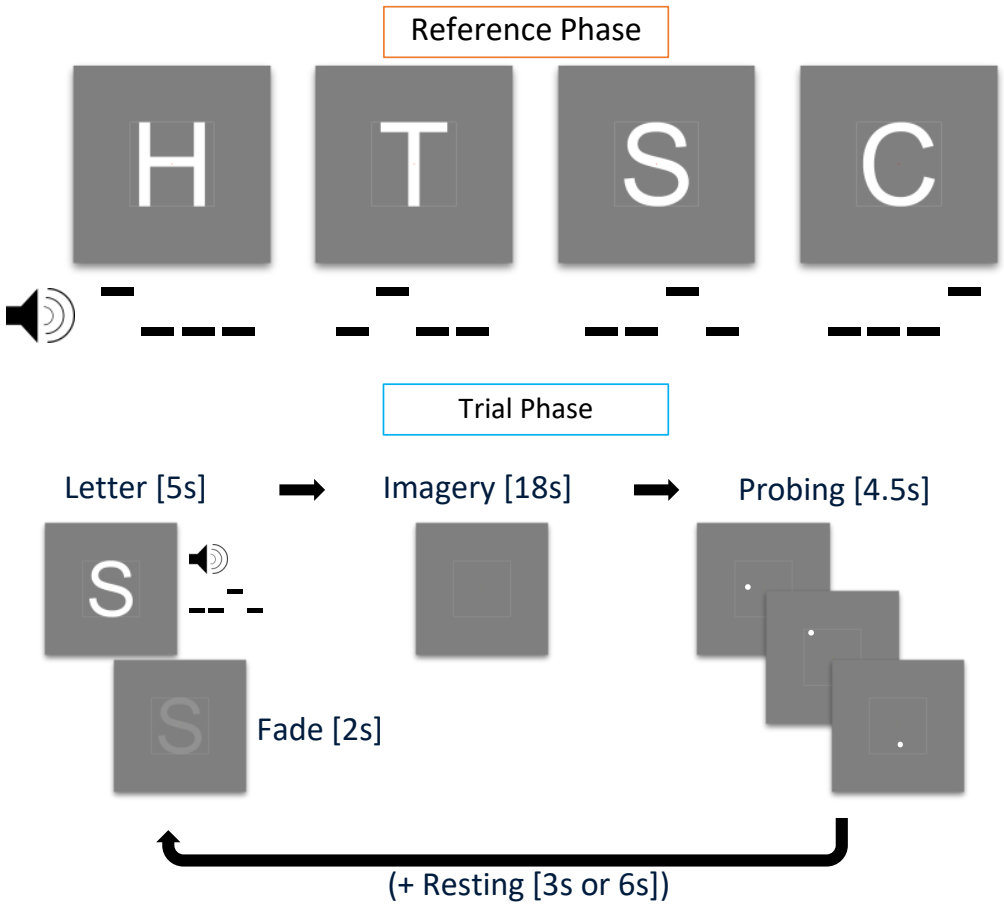



**Figure 4.2.1: training task.** In the reference phase (top), four letters H, T, 'S' 'C' were paired with a tone pattern. In the trial phase (bottom), the tone pattern was played and the letter shown for 5s (fading out after 3s) followed by an imagery period of 18s, a probing period of 4.5s, and a resting period of 3s or 6s.

### Perception run

To measure brain activation patterns in visual areas during the perception of the letters used in the imagery runs, we recorded one perception run during the scanning session. The four letters were visually presented using the same trial timing parameters as in the experimental runs. There were neither reference nor probing phases. Letters were presented for the duration of the imagery phase (6 s) and their shape was filled with a flickering checkerboard pattern (10 Hz). No tone patterns were played during the perception run. The recorded responses were also used to train a denoising autoencoder (see below).

### pRF mapping

A bar aperture (1.33° wide) revealing a flickering checkerboard pattern (10 Hz) was presented in four orientations. For each orientation, the bar covered the entire screen in 12 discrete steps (each step lasting 3 s). Within each orientation, the sequence of steps (and hence of the locations) was randomized (cf. Senden et al. 2014). Each orientation was presented six times.

### Stimulus presentation

The bar stimulus used for pRF mapping was created using the open source stimulus presentation tool BrainStim (http://svengijsen.github.io/BrainStim/). Visual and auditory stimulation in the imagery and perception runs were controlled with PsychoPy (version 1.83.03; Peirce 2007). Visual stimuli were projected on a frosted screen at the top end of the scanner table by means of an LCD projector (Panasonic, No PT- EZ57OEL; Newark, NJ, USA). Auditory stimulation was presented using MR-compatible insert earphones (Sensimetrics, Model S14; Malden, MA, USA). Responses to the probes were recorded with MR-compatible button boxes (Current Designs, 8-button response device, HHSC-2×4-C; Philadelphia, USA).

### Magnetic resonance imaging

We recorded anatomical and functional images with a Siemens Magnetom 7 T scanner (Siemens; Erlangen, Germany) and a 32-channel head-coil (Nova Medical Inc.; Wilmington, MA, USA). Prior to functional scans, we used a T1-weighted magnetization prepared rapid acquisition gradient echo (Marques et al. 2010) sequence [240 sagittal slices, matrix=320 320, voxel size=0.7 by 0.7 by 0.7 mm3, first inversion time TI1=900 ms, second inversion time TI2=2750 ms, echo time (TE)=2.46 ms, repetition time (TR)=5000 ms, first nominal flip

angle=5°, and second nominal flip angle=3°] to acquire anatomical data. For all functional runs, we acquired high-resolution gradient echo (T2* weighted) echo-planar imaging (Moeller et al. 2010) data (TE=26 ms, TR=3000 ms, generalized auto-calibrating partially parallel acquisitions (GRAPPA) factor=3, multiband factor=2, nominal flip angle=55°, number of slices=82, matrix=186 by 186, and voxel size=0.8 by 0.8 by 0.8 mm3). The field-of-view covered occipital, parietal, and temporal areas. In addition, before the first functional scan, we recorded five functional volumes with opposed phase encoding directions to correct for EPI distortions that occur at higher field strengths (Andersson et al. 2003).

**Processing of (f)MRI data**

We analyzed anatomical and functional images using BrainVoyager 20 (version 20.0; Brain Innovation; Maastricht, The Netherlands) and custom code in MATLAB (version 2017a; The Mathworks Inc.; Natick, MA, USA). We interpolated anatomical images to a nominal resolution of 0.8 mm isotropic to match the resolution of functional images. In the anatomical images, the grey/white matter boundary was detected and segmented using the advanced automatic segmentation tools of BrainVoyager 20 which are optimized for high-field MRI data. A region-growing approach analyzed local intensity histograms, corrected topological errors of the segmented grey/white matter border, and finally reconstructed meshes of the cortical surfaces (Kriegeskorte and Goebel 2001; Goebel et al. 2006). The functional images were corrected for motion artefacts using the 3D rigid body motion correction algorithm implemented in BrainVoyager 20 and all functional runs were aligned to the first volume of the first functional run. We corrected EPI distortions using the COPE ("Correction based on Opposite Phase Encoding") plugin of BrainVoyager that implements a method similar to that described in Andersson, Skare, and Ashburner (Andersson et al. 2003) and the 'topup' tool implemented in FSL (Smith et al. 2004). The pairs of reversed phase encoding images recorded in the beginning of the scanning session were used to estimate the susceptibility-induced off-resonance field and correct the distortions in the remaining functional runs. After this correction, functional data were high-pass filtered using a general linear model (GLM) Fourier basis set of three cycles sine/cosine, respectively. This filtering included a linear trend removal. Finally, functional runs were co-registered and aligned to the anatomical scan using an affine transformation (9 parameters) and z-normalized to eliminate signal offsets and inter-run variance.

### pRF mapping and region-of-interest definition

For each subject, we fit location and size parameters of an isotropic Gaussian population receptive field model (Dumoulin and Wandell 2008) by performing a grid search. In terms of pRF location, the visual field was split into a circular grid of 100 by 100 points, whose density decays exponentially with eccentricity. Receptive field size exhibits a linear relationship with eccentricity with the exact slope depending on the visual area (Freeman and Simoncelli 2011). For this reason, we explored slopes in the range from 0.1 to 1 (step=0.1), as this effectively allows for exploration of a greater range of receptive field sizes (10 for each unique eccentricity value). We used the pRF mapping tool from the publicly available Computational Neuroimaging Toolbox (https://github.com/MSenden/CNI_toolbox). Polar angle maps resulting from pRF fitting were projected onto inflated cortical surface reconstructions and used to define regions-of-interest (ROIs) for bilateral visual areas V1, V2, and V3. The resulting surface patches from the left and right hemisphere were projected back into volume space (from 1 mm to +3 mm from the segmented grey/white matter boundary). Volume ROIs were then defined for V1, V2, V3, and a combined ROI (V1V2V3).

### Voxel patterns

All our analyses and reconstructions are based on letter-specific spatial activation profiles of voxel co-activations; i.e., voxel patterns. Voxel patterns within each ROI were obtained for both perceptual and imagery runs. First, for each run, single-trial letter-specific voxel patterns were obtained by averaging BOLD activations in the range from +2 until +3 volumes following trial onset and z-normalizing the result. This lead to a total of eight (one per trial) perceptual and 32 (four runs with 8 trials each) imagery voxel patterns per letter. We, furthermore, computed perceptual and imagery average voxel patterns per letter by averaging over all single-trial patterns (and runs in case of imagery) of a letter and z-normalizing the result. Imagery average voxel patterns were used in an encoding analysis and for assessment of reconstruction quality, while perceptual average patterns were used for training a denoising autoencoder (Vincent et al. 2008).

### Encoding analysis

To test the hypothesis that spatial activation profiles of visual mental imagery are geometry-preserving, we tested whether voxel activations predicted from

the encoding model (one isotropic Gaussian per voxel) and a physical (binary) stimulus corresponding to the imagined letter provides a significantly better fit with measured voxel activations than predictions from the remaining binary letter stimuli. Specifically, for each participant and ROI, we predicted voxel activations for each of the four letters based on pRF estimates and physical letter stimuli.

### Autoencoder

We trained an autoencoder with a single hidden layer $k = \lfloor 0.1 \cdot N_{\text{voxels}} \rfloor$ to reproduce average perceptual voxel patterns from noise-corrupted versions per subject and ROI. Since the values of voxel patterns follow a Gaussian distribution with a mean of zero and unit standard deviation, we opted for zero-mean additive Gaussian noise with a standard deviation $\sigma = 12$ for input corruption. Note that the exact value of $\sigma$ is not important as long as it sufficiently corrupts the data. We achieved similar results with values in the range [8,14]. The hidden layer consisted of units with rectified linear activation functions. Output units activated linearly. Encoding weights (from input to hidden layer) and decoding weights (from hidden to output layer) were shared. Taken together, the input, hidden, and output layers were, respectively, given by:

$$\begin{aligned} \mathbf{y_c} &= \mathbf{y} + \epsilon, \text{ with } \epsilon \sim N(\mathbf{0}, \sigma) \\ \mathbf{h} &= \phi\left(\mathbf{W_e}\mathbf{y_c} + \mathbf{b_e}\right), \text{ with } \phi(x) = \frac{1}{1+e^{-x}} \\ \mathbf{y_r} &= \mathbf{W_d}\mathbf{h} + \mathbf{b_d}, \text{ with } \mathbf{W_d} = \mathbf{W_e}^T \end{aligned} \tag{4.1}$$

In Eq.4.1, y is a voxel pattern (of length $v$), $\mathbf{y_c}$ its noise corruption, and yr its restoration. $\mathbf{W_e}$ (k-by-v matrix) and $\mathbf{W_d}$ (v-by-k matrix) are the tied encoding and decoding weights, respectively. Finally, be (k-by-1 vector) and $\mathbf{b_d}$ (v-by-1 vector) are the biases of the hidden and output layers, respectively. We used mean squared distances to measure loss between the input and its restoration and implemented the autoencoder in the TensorFlow library (Abadi et al. 2016) for Python (version 2.7, Python Software Foundation, https://www.python.org/). The autoencoder was trained using the Adam optimizer (Kingma and Ba 2014) with a learning rate of $1 \times 10^{-5}$ and a batch size of 100 for 2000 iterations. In addition to the four average perceptual voxel patterns, we also included an equal amount of noise-corrupted mean luminance images to additionally force reconstructions to zero if the input contained no actual signal. No imagery data were used for training the autoencoder.

**Reconstruction**

For each subject and ROI, we reconstructed the visual field from average perceptual and imagery voxel patterns. We obtained weights mapping the cortex to the visual field by inverting the mapping from visual field to cortex given by the population receptive fields. Since $\mathbf{W_p RF}$, a v-by-p matrix (with v being the number of voxels and p the number of pixels) mapping a 150-by-150 pixel visual field to the cortex (i.e., p=22500 pixels; after vectorizing the visual field) is not invertible, we minimize the error function:

$$E = \left(\mathbf{y} - \mathbf{W}_{pRF}\mathbf{x}\right)^T \left(\mathbf{y} - \mathbf{W}_{pRF}\mathbf{x}\right) + \mathbf{D}\|\mathbf{x}\|_2^2 \tag{4.2}$$

with respect to the input image x (a vector of length p). The vector $\mathbf{y}$ is of length v and reflects a measured voxel pattern. Finally, $\mathbf{D}$ is a diagonal matrix of the outdegree of each pixel in the visual field which provides pixel-specific scaling of the L2 regularization term $\|\mathbf{x}\|_2^2$ and accounts for cortical magnification. Minimizing Eq.4.2 leads to the expression:

$$\mathbf{x} = \left(\mathbf{W}_{pRF}^T\mathbf{W}_{pRF} + \mathbf{D}\right)^{-1}\mathbf{W}_{pRF}^T\mathbf{y} \tag{4.3}$$

with which we can reconstruct the visual field from voxel patterns. To minimize computational cost, we compute the projection matrix $\mathbf{W}_{VF} = \left(\mathbf{W}_{pRF}^T\mathbf{W}_{pRF} + \mathbf{D}\right)^{-1}\mathbf{W}_{pRF}^T$ once per ROI and subject rather than performing costly matrix inversion for every reconstruction. Note that both raw voxel patterns (y) as well as restored voxels patterns (yr) obtained from passing y through the autoencoder, can be used for image reconstruction. In the former case, $\mathbf{x} = \mathbf{W}_{VF}\mathbf{y}$. In the latter case, $\mathbf{x} = \mathbf{W}_{VF}\mathbf{y_r} = \mathbf{W}_{VF}\left[\mathbf{W_d}\phi\left(\mathbf{W_e}\mathbf{y} + \mathbf{b_e}\right) + \mathbf{b_d}\right]$.

For each letter, we assessed the quality of its reconstruction by calculating the correlation between the reconstruction and the corresponding binary letter stimulus. This constitutes a first-level correlation metric. However, since the four letters bear different visual similarities with each other (e.g., 'S' and 'C' might resemble each other more closely than either resemble 'H'), we also defined a second-level correlation metric. Specifically, we obtained one vector

of all pairwise correlations between physical letter stimuli and a second vector of pairwise correlations between corresponding reconstructions and correlated these two vectors.

### Classification

We replaced the output layer of the pretrained autoencoder with a four-unit (one for each letter) softmax classifier. Weights from the hidden to the classification layer as well as the biases of output units were then trained to classify single-trial imagery voxel patterns using cross entropy as a measure of loss. Note that pretrained weights from input to hidden layer ($\mathbf{W_e}$ in Eq.4.1) as well as pretrained hidden unit biases ($\mathbf{b_e}$ in Eq.4.1) remained fixed throughout training of the classifier. These weights and biases were thus dependent purely on perceptual data. This procedure is equivalent to performing multinomial logistic regression on previously established hidden layer representations. Imagery runs were split into training and testing data sets in a leave-one-run-out procedure, such that the classifier was repeatedly trained on a total of 96 voxel patterns (8 trials per 4 letters for each of three runs) and tested on the remaining 32 voxel patterns. We again trained the network using the Adam optimizer. However, in this case, the learning rate was $1 \times 10^{-4}$, the batch size equal to 96, and training lasted merely 250 iterations.

### Statistical analysis

Statistical analyses were performed using MATLAB (version 2017a; The Mathworks Inc.; Natick, MA, USA). We used a significance level of $\alpha = 0.05$ (adjusted for multiple comparisons where appropriate) for all statistical analyses. Behavioral results were analyzed using repeated-measures ANOVA with task (visible or invisible runs) and time as within-subject factors.

For the encoding analysis, we performed a mixed-model regression for the average voxel activations of each imagined letter within each ROI with physical letter as fixed and participant as random factors, respectively. This was followed by a contrast analysis. For each imagined letter, the contrast was always between the corresponding physical stimulus and all remaining physical stimuli. For example, when considering voxel activations for the imagined letter 'H', a weight of 3 was placed on activations predicted from the physical letter 'H' and a weight of 1 was placed on activations predicted from each of the remaining three letters. Since we repeated the analysis for each imagined letter

(4) and single region ROI (3), we performed a total of 12 tests and considered results significant at a corrected cutoff of $\alpha_c = 0.05/12 = 0.0042$.

To evaluate which factors contribute most to first-level reconstruction quality, we performed mixed-model regression with the VVIQ and the OSIVQ spatial and OSIVQ object scores, ROI (using dummy coding, $V1 = reference$), letter (dummy coding, '$H' = reference$), and number of selected voxels (grouped by ROI). To assess second-level reconstruction quality, we use the same approach omitting letter as a predictor.

To assess the significance of classification results, we evaluated average classification accuracy across the four runs against a Null distribution obtained from 1000 permutations of a leave-one-run-out procedure with randomly scrambled labels. We performed this analysis separately for each subject and ROI and consider accuracy results significant if they exceed the 95th percentile of the Null distribution. To statistically evaluate which factors contribute most to classification accuracy, we performed mixed-model regression with the VVIQ and the OSIVQ spatial and OSIVQ object scores, ROI (using dummy coding, $V1 = reference$), letter (dummy coding, '$H' = reference$), and number of selected voxels (again grouped by ROI).

## 4.3 Results

### 4.3.1 Behavioral results

VVIQ and OSIVQ scores for each participant are shown in Fig. 4.3.1. The average score over participants for VVIQ was 4.07 (95% CI [3.71, 4.43]). For the object, spatial, and verbal sub-scales of OSIVQ, average scores were 2.88 (95% CI [2.48, 3.27]), 3.08 (95% CI [2.75, 3.41]), and 3.81 (95% CI [3.33, 4.29]), respectively. Participants reported that they tried to maintain the afterimage of the fading stimulus as a strategy to enforce vivid and accurate letter imagery. Furthermore, participants determined through button presses whether a probe was located inside or outside the letter shape, while the letter was either visible or imagined. A repeated-measures ANOVA with task (visible or invisible runs) and time as within-subject factors revealed a statistically significant effect of time on probing accuracy $\left(F_{(2,10)} = 19.84, p < 0.001\right)$ and no significant difference for task $\left(F_{(1,5)} = 1.10, p = 0.341\right)$ (Table 4.1).

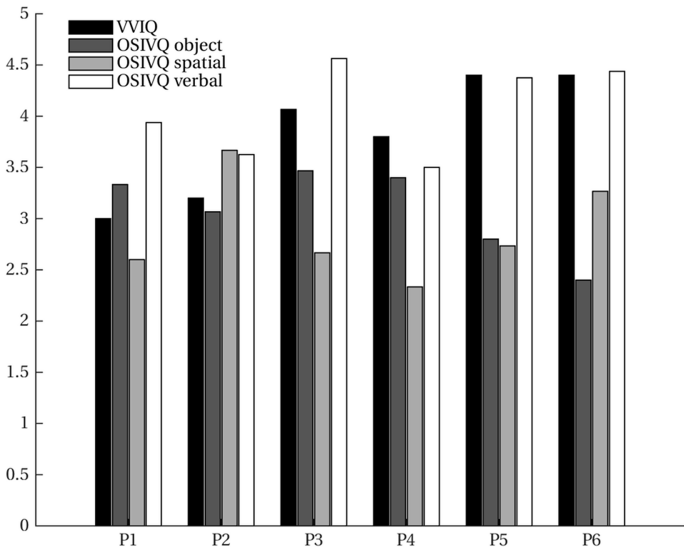



**Figure 4.3.1: vividness of visual imagery.** Vividness of Visual Imagery Questionnaire (VVIQ) and Object-Spatial Imagery and Verbal Questionnaire (OSIVQ) scores (with the subscales for "object", "spatial", and "verbal" imagery styles) are shown for all participants.

### 4.3.2 Encoding

For each imagined letter (H, T, S, C) in each single-area ROI (V1, V2, V3), we investigated whether spatial voxel activation profiles can be predicted from a Prf-encoding model and the corresponding physical stimulus. That is, for each imagined letter–ROI combination, we ran a mixed-model regression with observed imagery voxel activations (averaged over trials and runs) as outcome variable, predicted voxel activations for each physical letter stimulus as predictors and participants as grouping variable. Since we were specifically interested in testing our hypothesis that the retinotopic organization of imagery voxel activations is sufficiently geometrically specific to distinguish among different imagined letters, we performed contrast analyses between the physical letter corresponding to the imagery and all the remaining letters (see "Methods" for details). Contrasts were significant after applying Bonferroni correction ($\alpha = 0.0042$) for each of the twelve letter-ROI combinations. In other words, predictions based on a specific physical letter gave a better account of voxel activations observed for the imagery of that specific letter than those based on every other physical letter, as can be appreciated from Table 4.2. Figure 4.3.2 visualizes these results in the form of boxplots of first-level beta values (i.e., distribution over participants per physical letter) in each letter-ROI combination.

**Table 4.1:** contrast analysis comparing the physical letter corresponding to the imagery with all remaining letters. Each of the twelve letter-ROI combinations was significant after applying Bonferroni correction ($\alpha_c = 0.0042$).

| | T1 | T2 | T3 |
|---|---|---|---|
| Visible | 60.42 (95% CI [48.2, 72.64]) | 75.39 (95% CI [66.70, 84.08]) | 77.73 (95% CI [69.36, 86.10]) |
| Invisible | 62.02 (95% CI [44.57.36, 79.45]) | 73.18 (95% CI [65.98, 80.38]) | 81.57(95% CI [75.47, 87.67]) |

**Table 4.2:** contrast analysis comparing the physical letter corresponding to the imagery with all remaining letters. Each of the twelve letter-ROI combinations was significant after applying Bonferroni correction ($\alpha_c = 0.0042$).

| | H | T | S | C |
|---|---|---|---|---|
| V1 | $t_{(2)} = 32.11, p = 0.0004$ | $t_{(2)} = 48.00, p = 0.0002$ | $t_{(2)} = 14.10, p = 0.0025$ | $t_{(2)} = 29.84, p = 0.0006$ |
| V2 | $t_{(2)} = 25.21, p = 0.0008$ | $t_{(2)} = 67.63, p = 0.0001$ | $t_{(2)} = 19.64, p = 0.0013$ | $t_{(2)} = 47.48, p = 0.0002$ |
| V3 | $t_{(2)} = 47.90, p = 0.0006$ | $t_{(2)} = 27.60, p = 0.0007$ | $t_{(2)} = 11.48, p = 0.0038$ | $t_{(2)} = 32.83, p = 0.0005$ |

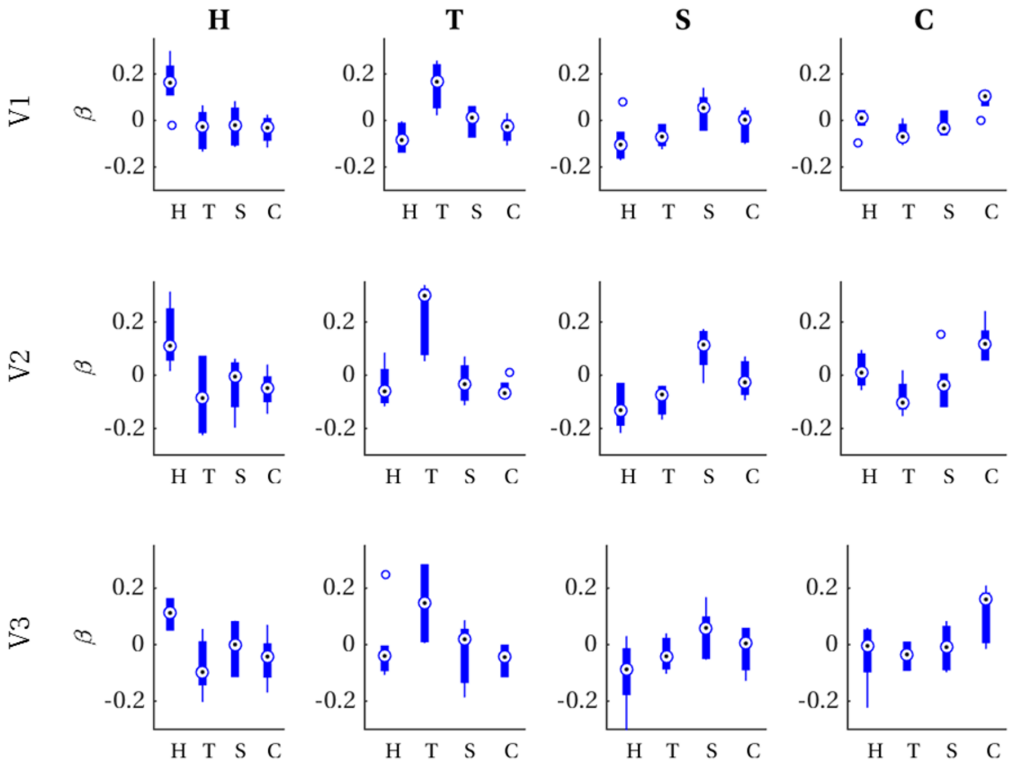



**Figure 4.3.2: first-level beta distributions.** Distribution of first-level beta values (across participants) for voxel patterns predicted from each physical letter (x-axis) for all combinations of ROI (rows) and imagined letters (columns).

### 4.3.3 Reconstruction

#### Raw imagery data

We reconstructed the visual field from average imagery voxel patterns in response to each letter (see Figs. 4.3.3, 4.3.4). Mean correlations between reconstructed imagery and physical letters are presented in Table 4.3 (for comparison, Table 4.4 shows correlations between reconstructed perception and physical letters). As can be appreciated from these results as well as the figures, first-level reconstruction quality varies across ROIs as well as across subjects. Differences between subjects might be due to differences in their ability to imagine shapes accurately and vividly as measured by the VVIQ and OSIVQ questionnaires. Differences between ROIs might be due to differences with respect to their retinotopy (most likely receptive field sizes) or due to different numbers of voxels included for analysis of each ROI. Only the former would be a true ROI effect. We investigate which factors account for observed correlations (transformed to Fisher z-scores for analyses) by performing a mixed-model regression with questionnaire scores, ROI (using dummy coding, V1=reference), letter (dummy coding, 'H'=reference), and number of selected voxels as predictors. A number of voxels were grouped by ROI. Furthermore, the regression model included the VVIQ and the OSIVQ spatial and object scores. However, the VVIQ score was not included since it correlated highly with the OSIVQ verbal score (leading to collinearity). To further prevent collinearity, we also only included single-area ROIs in this analysis and not the combined ROI. A number of voxels $[t_{(62)} = 2.59, p = 0.012]$ and the OSIVQ object score $[t_{(62)} = 2.64, p = 0.010]$ were significant quantitative predictors. Furthermore, letter was a significant categorical predictor. Specifically, letter 'T' $[t_{(62)} = 5.58, p0.001]$ presented with significantly improved correlation values over the reference letter 'H', whereas letters 'S' $[t_{(62)} = 3.88, p = 0.0003]$ and 'C' $[t_{(62)} = 2.25, p = 0.028]$ presented with significantly decreased correlation values with respect to the reference. Neither the OSIVQ verbal score $[t_{(62)} = 0.0278, p = 0.978]$ nor the ROI were significant predictors of reconstruction quality.

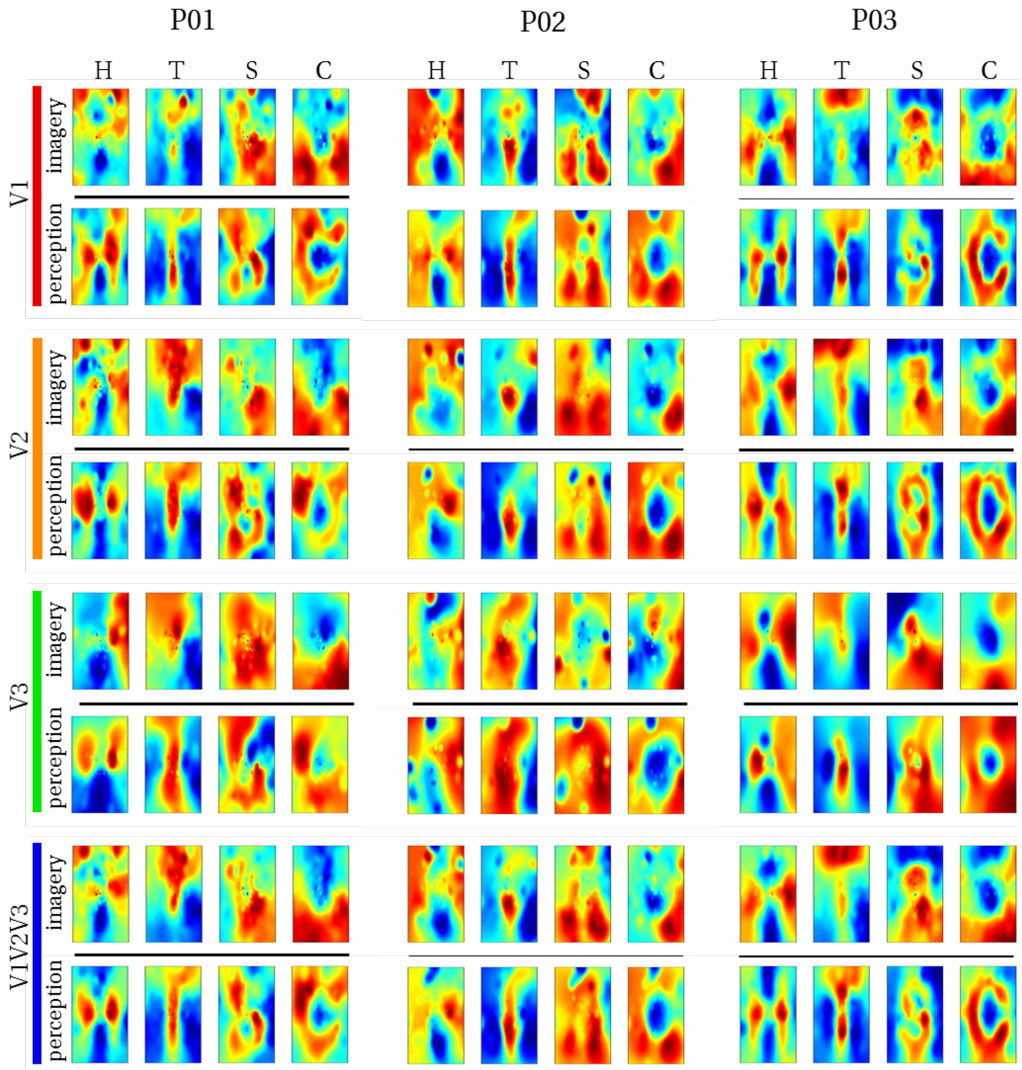



**Figure 4.3.3: reconstructed visual field images (participants 1-3).** Reconstructed average visual field images are visualized for each ROI of participants one, two, and three. Reconstructions of the remaining three subjects are shown in figure 4.3.4. Perceptual as well as imagery voxel patterns were obtained from raw BOLD time-series.

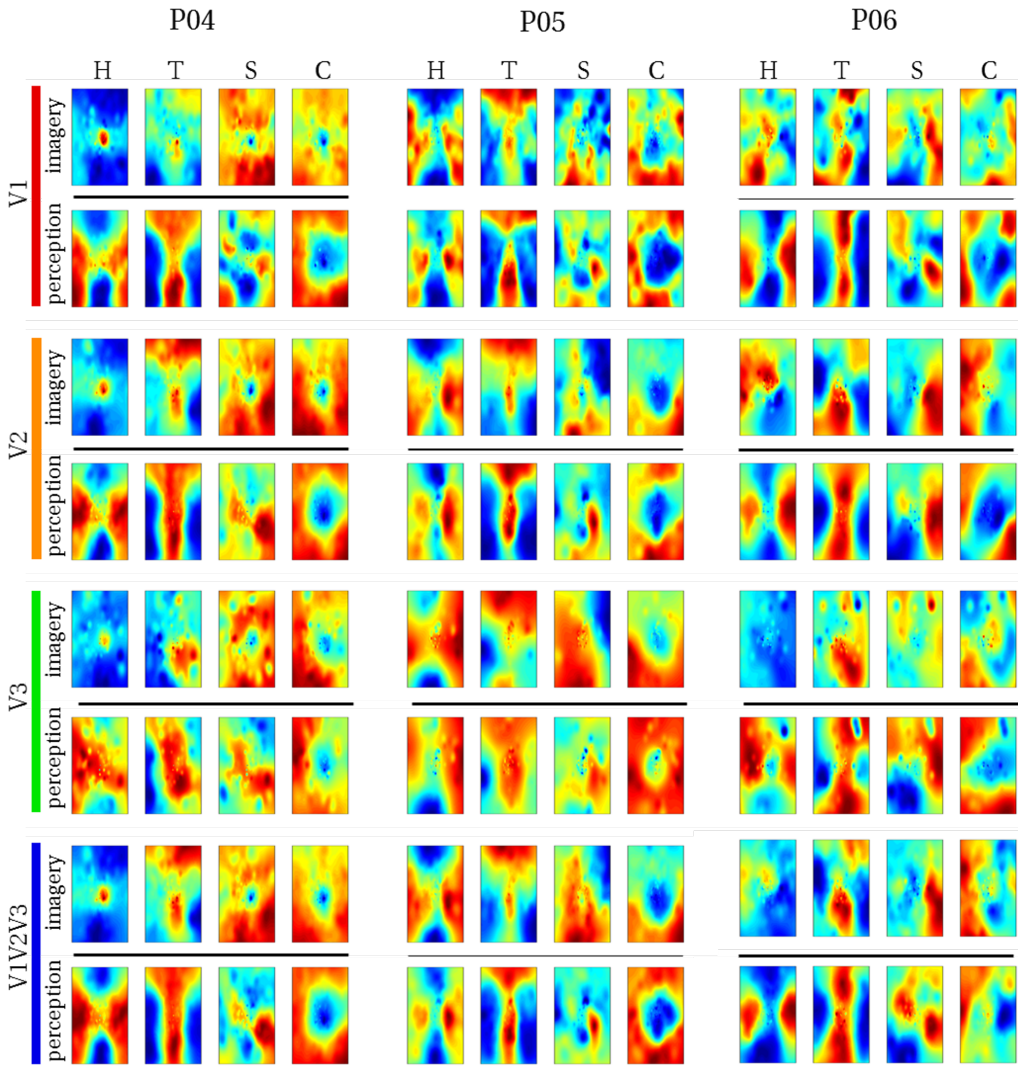



**Figure 4.3.4: reconstructed visual field images (participants 4-6).** Reconstructed average visual field images are visualized for each ROI of participants four, five, and six. Reconstructions of the remaining three subjects are shown in figure 4.3.3. Perceptual as well as imagery voxel patterns were obtained from raw BOLD time-series.

**Table 4.3:** first order correlations between reconstructed *imagined* letters and physical stimuli (averages over participants).

| | H | T | S | C |
|---|---|---|---|---|
| V1 | 0.24 (95% CI [0.09, 0.40]) | 0.49 (95% CI [0.38, 0.58]) | 0.08 (95% CI [0.03 0.19]) | 0.14 (95% CI [0.07, 0.21]) |
| V2 | 0.21 (95% CI [0.13, 0.30]) | 0.46 (95% CI [0.36, 0.55]) | 0.10 (95% CI [0.04, 0.17]) | 0.12 (95% CI [0.07, 0.18]) |
| V3 | 0.20 (95% CI [0.10, 0.30]) | 0.28 (95% CI [0.14, 0.46]) | 0.04 (95% CI [0.03, 0.12]) | 0.15 (95% CI [0.01, 0.29]) |
| V1V2V3 | 0.27 (95% CI [0.16, 0.37]) | 0.51 (95% CI [0.45, 0.56]) | 0.12 (95% CI [0.02, 0.21]) | 0.14 (95% CI [0.08, 0.20]) |

**Table 4.4:** first order correlations between reconstructed *perceived* letters and physical stimuli (averages over participants).

| | H | T | S | C |
|---|---|---|---|---|
| V1 | 0.40 (95% CI [0.35, 0.44]) | 0.65 (95% CI [0.60, 0.69]) | 0.27 (95% CI [0.15, 0.38]) | 0.32 (95% CI [0.22, 0.40]) |
| V2 | 0.37 (95% CI [0.31, 0.42]) | 0.58 (95% CI [0.50, 0.64]) | 0.19 (95% CI [0.09, 0.29]) | 0.31 (95% CI [0.23, 0.38]) |
| V3 | 0.25 (95% CI [0.19, 0.31]) | 0.41 (95% CI [0.30, 0.51]) | 0.06 (95% CI [-0.06, 0.18]) | 0.27 (95% CI [0.25, 0.30]) |
| V1V2V3 | 0.41 (95% CI [0.36, 0.46]) | 0.63 (95% CI [0.56, 0.68]) | 0.22 (95% CI [0.12, 0.32]) | 0.31 (95% CI [0.24, 0.38]) |

Next, we examined the second-level correlation metric of reconstruction quality. Correlations between physical and reconstruction pairwise first-level correlation vectors were 0.60 (95% CI [0.28, 0.80], p=0.103) for V1, 0.65 (95% CI [0.34, 0.83], p=0.082) for V2, 0.48 (95% CI [0.15, 0.71], p=0.167) for V3, and 0.64 (95% CI [0.34, 0.83], p=0.084) for V1V2V3, respectively. Finally, we performed a mixed regression to assess which factors account for the observed correlations (again transformed to Fisher z-scores). We included OSIVQ verbal, spatial, and object scores, ROI (dummy coding, V1=reference), and number of selected voxels (grouped by ROI) as predictors. The OSIVQ object score $[t_{(11)} = 3.26, p = 0.0076]$ and number of voxels $[t_{(11)} = 3.94, p = 0.0023]$ significantly predicted second-level correlations, while the spatial $[t_{(11)} = 0.71, p = 0.492]$ and verbal scores $[t_{(11)} = 0.81, p = 0.436]$ did not. Furthermore, there was significant effect of ROI, since neither V2 $[t_{(11)} = 1.56, p = 0.148]$ nor V3 $[t_{(11)} = 0.40, p = 0.697]$ significantly differed from V1.

### Processed imagery data

Our results confirm that visual mental imagery preserves perceptual topographic organization. This can be leveraged to obtain improved reconstructions of mental imagery. Specifically, an autoencoder trained to retrieve perceptual voxel patterns from their noise-corrupted version can be utilized to enhance imagery data. Figure 4.3.5 shows how the autoencoder affects first-level reconstruction quality on a single-trial basis for V1. As shown in the figure, reconstruction quality was best for 'T', followed by 'H', 'C', and 'S'. A subject effect is also clearly apparent with participants three and five generally displaying the best

results. Finally, imagery reconstruction quality was generally inferior to perception prior to using the autoencoder. However, using the autoencoder pushed imagery reconstruction quality towards perception levels. Indeed, the autoencoder maps imagery voxel patterns onto the corresponding perception voxel patterns it has learned previously. This explains two important observations. First, for some trials, using the autoencoder decreased resemblance to the physical letter. This is especially apparent for participants four and six, whose reconstructions were generally not particularly good. Such decrements in reconstruction quality result from imagery voxel patterns in response to one letter falling within the attraction domain of another letter (resembling the activation pattern of that letter slightly more) and hence get mapped onto the wrong pattern. Second, even the few imagery trials, whose reconstructions match the physical letter better than the perceptual data were mapped onto the perceptual pattern. A notable example is two trials for the letter 'S' by participant two. This implies that reconstruction quality of the perceptual data used to train the autoencoder constitutes an upper limit for imagery when using the autoencoder.

As a general effect, the autoencoder maps imagery voxel patterns onto their perceptual counterpart for most individual trials. Hence, reconstructions of average imagery voxel patterns as well as of individual trials more strongly resemble the corresponding physical letter. Figure 4.3.6 shows reconstructions from average imagery voxel patterns after feeding the data through the autoencoder. Figures 4.3.7 and 4.3.8 show reconstructions of individual trials in a single run of participants three and five, respectively.

Obviously, these participants are not representative of the population at large but provide an indication of what is possible for people with a strong ability to imagine visual shapes. Table 4.4 shows the mean correlation values across trials of participant three and five when the data of these participants were fed through the autoencoder and without using the autoencoder.

**Table 4.5:** effect of the autoencoder on mean correlation values across trials for two participants

| | | H | T | S | C |
|---|---|---|---|---|---|
| P03 | autoencoder | 0.39 (95% CI [0.32, 0.45]) | 0.55 (95% CI [0.46, 0.62]) | 0.10 (95% CI [0.04, 0.16]) | 0.09 (95% CI [0.06, 0.12]) |
| | raw | 0.19 (95% CI [0.15, 0.21]) | 0.33 (95% CI [0.28, 0.38]) | -0.02 (95% CI [-0.06, 0.02]) | 0.02 (95% CI [-0.02, 0.06]) |
| P05 | autoencoder | 0.28 (95% CI [0.20, 0.35]) | 0.53 (95% CI [0.43, 0.61]) | 0.08 (95% CI [0.00, 0.17]) | 0.21 (95% CI [0.12, 0.31]) |
| | raw | 0.12 (95% CI [0.08, 0.15]) | 0.32 (95% CI [0.25, 0.37]) | 0.02 (95% CI [-0.01, 0.06]) | 0.07 (95% CI [0.03, 0.10]) |

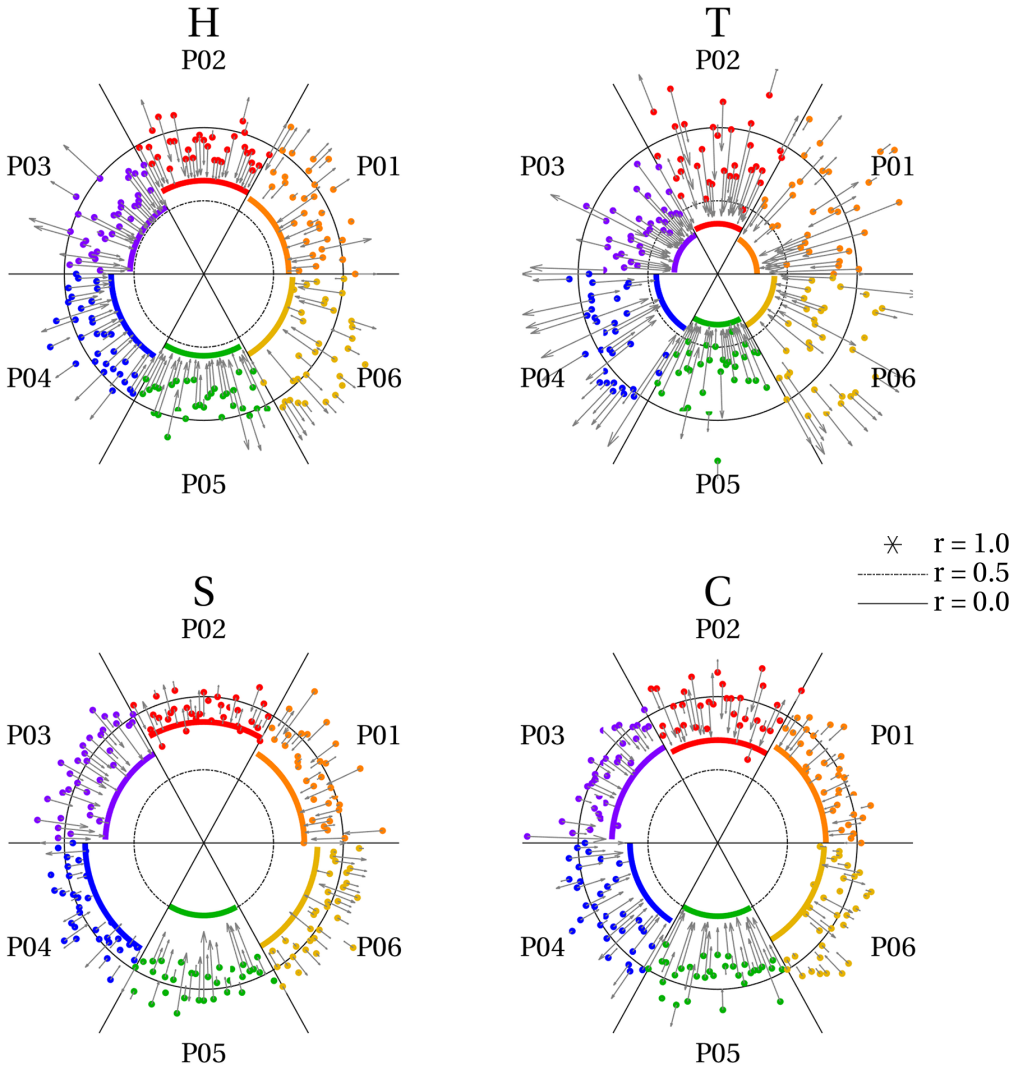



**Figure 4.3.5: effect of autoencoder on trial-specific reconstruction quality.** The radius of the circles represents reconstruction quality (correlations) with r=1 at the center, r = 0.5 at the inner ring (dash-dot) and r = 0 at the outer ring (solid). Each angle represents an imagery trial with 32 trials per letter and participant. Participants are color coded. Solid colored lines reflect reconstruction quality based on average perceptual voxel patterns of one participant. This constitutes a baseline against which to compare imagery reconstruction quality. Colored dots reflect imagery reconstruction quality for each individual trial of a participant. Finally, arrows reflect the displacement of each of these dots after feeding imagery data through the autoencoder. That is, the tip of the head reflects the new position of the dot after applying the autoencoder. Most points were projected onto the perception-level correlation value and hence approached the center. However, some moved further away from the center.

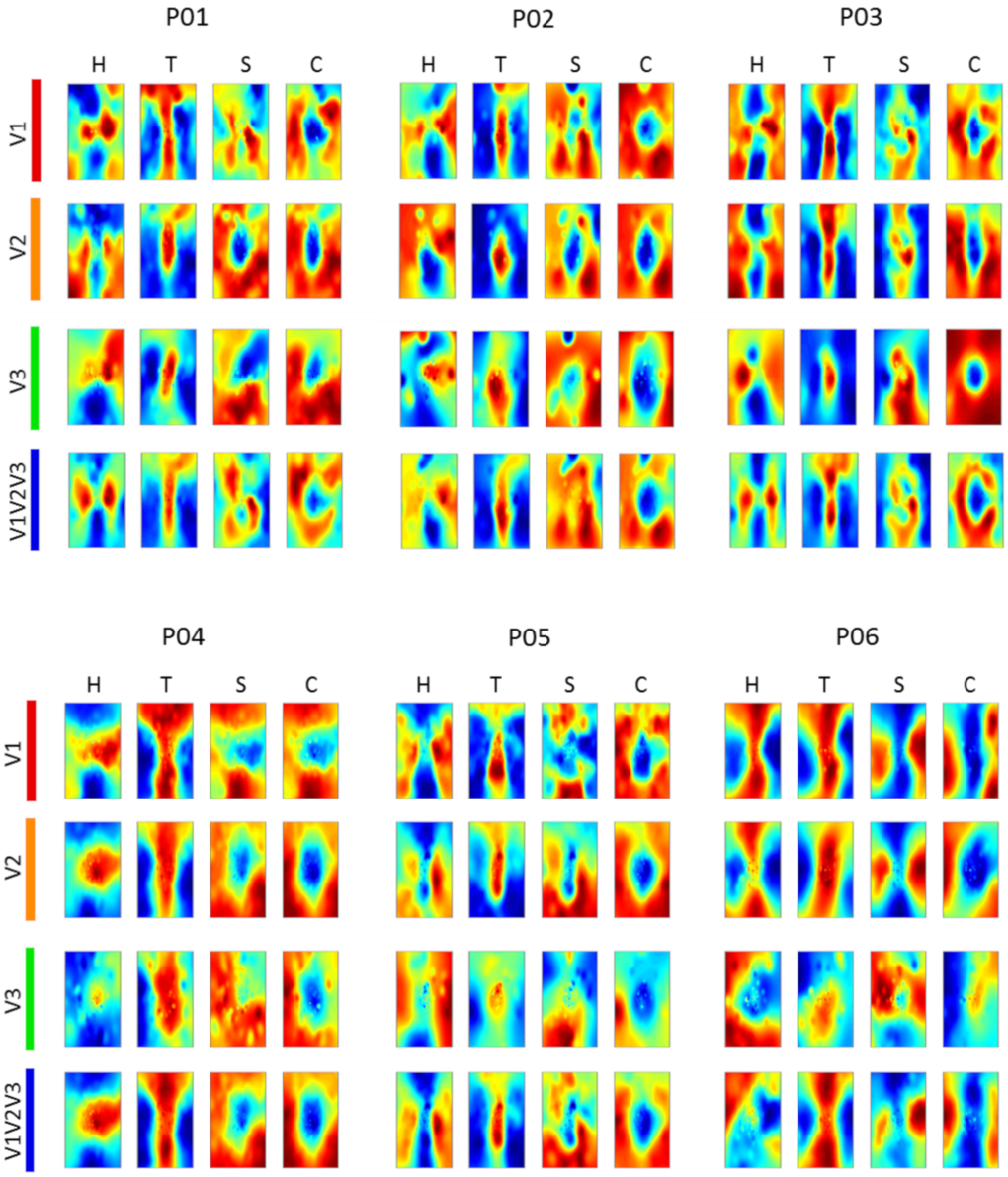



**Figure 4.3.6: reconstructed imagery.** Reconstructed average visual field images of mental imagery are visualized for each ROI of each participant. Imagery voxel patterns were obtained from cleaned BOLD time-series after feeding raw data through the autoencoder.

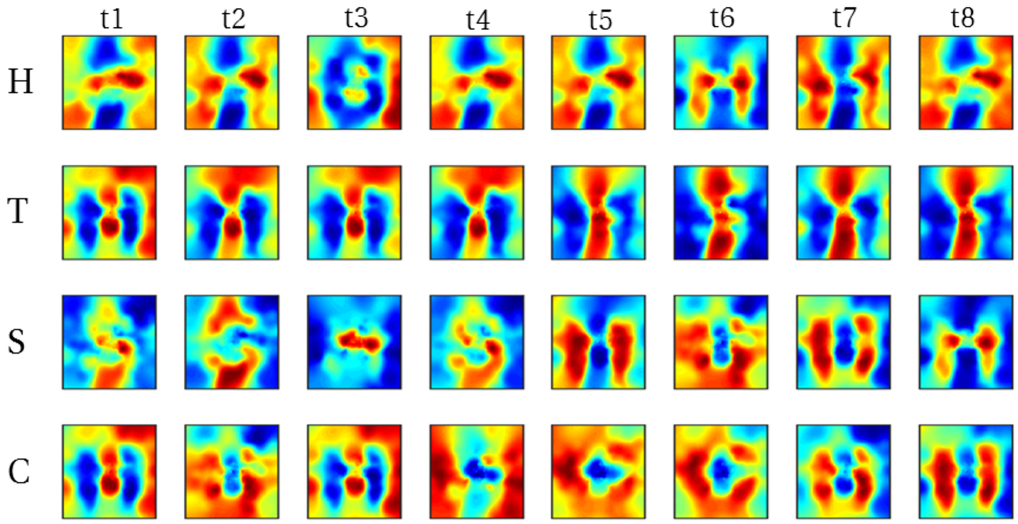



**Figure 4.3.7: reconstructed visual field images from denoised single trials in a single run of participant 3.** Each run comprised of 8 trials (columns) per letter (rows). Recognizable reconstructions can be obtained for a number of (though not all) individual trials.

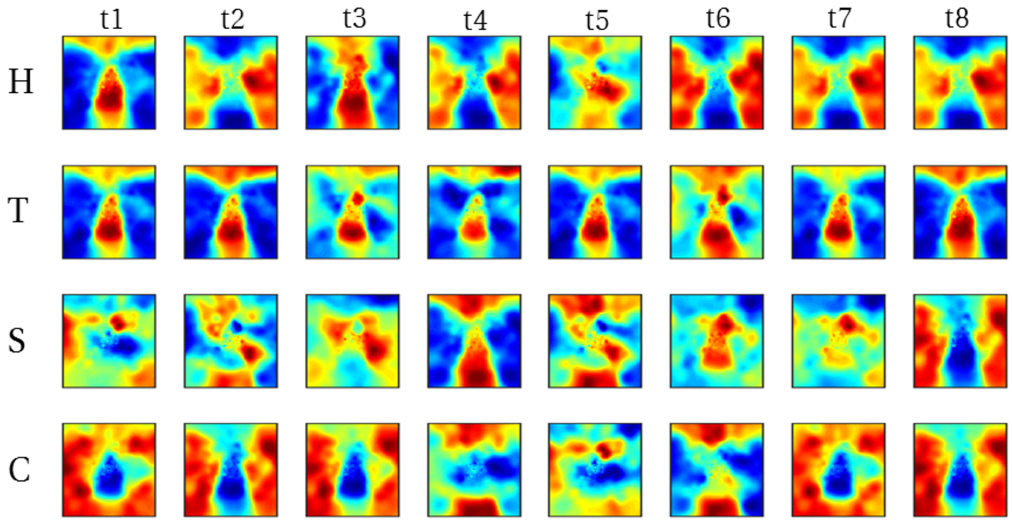



**Figure 4.3.8: reconstructed visual field images from denoised single trials in a single run of participant 5.** Each run comprised of 8 trials (columns) per letter (rows). Recognizable reconstructions can be obtained for a number of (though not all) individual trials.

### 4.3.4 Classification

Having established support for the hypothesis that activity in early visual cortex in response to imagery exhibits a similar topographical profile as perception, we proceeded to test whether it is possible to pretrain latent representations for an imagery classifier using purely perceptual data. The classifier consists of three layers with the output layer being a softmax classifier stacked onto the hidden layer of an autoencoder pretrained to denoise perceptual voxel patterns (see "Methods" for details). We trained the classifier on imagery data using a leave-one-run-out procedure; that is, we trained the classifier on three of the four imagery runs and tested classification accuracy on the left-out run. Figure 4.3.9 shows average classification accuracies per subject and ROI (including the combined ROI 'V1V2V3'). For five of the six participants, average classification accuracies exceeded theoretical chance levels (25% correct) as well as the 95th percentile of 1000 permutation runs (randomly scrambled labels) in all ROIs. For participant six, theoretical chance levels as well as the 95th percentile were (barely) exceeded for V2 only.

We performed a mixed-model regression with the OSIVQ object, spatial and verbal scores, ROI (using dummy coding, V1=reference), and number of selected voxels (again grouped by ROI) as predictors to assess which factors account for the observed accuracies. Number of voxels $[t_{(11)} = 4.80, p = 0.0006]$, the object sub-score of OSIVQ $[t_{(11)} = 4.83, p < 0.0005]$, and the spatial sub-score of OSIVQ $[t_{(11)} = 3.45, p = 0.006]$ were significant predictors of accuracy, whereas the verbal sub-score of OSIVQ $[t_{(11)} = 0.656, p = 0.525]$ was not. Furthermore, neither V2 $[t_{(11)} = 1.72, p = 0.113]norV3[\mathrm{t}_{(11)} = 1.85, p = 0.092]$ differed significantly from reference (V1).

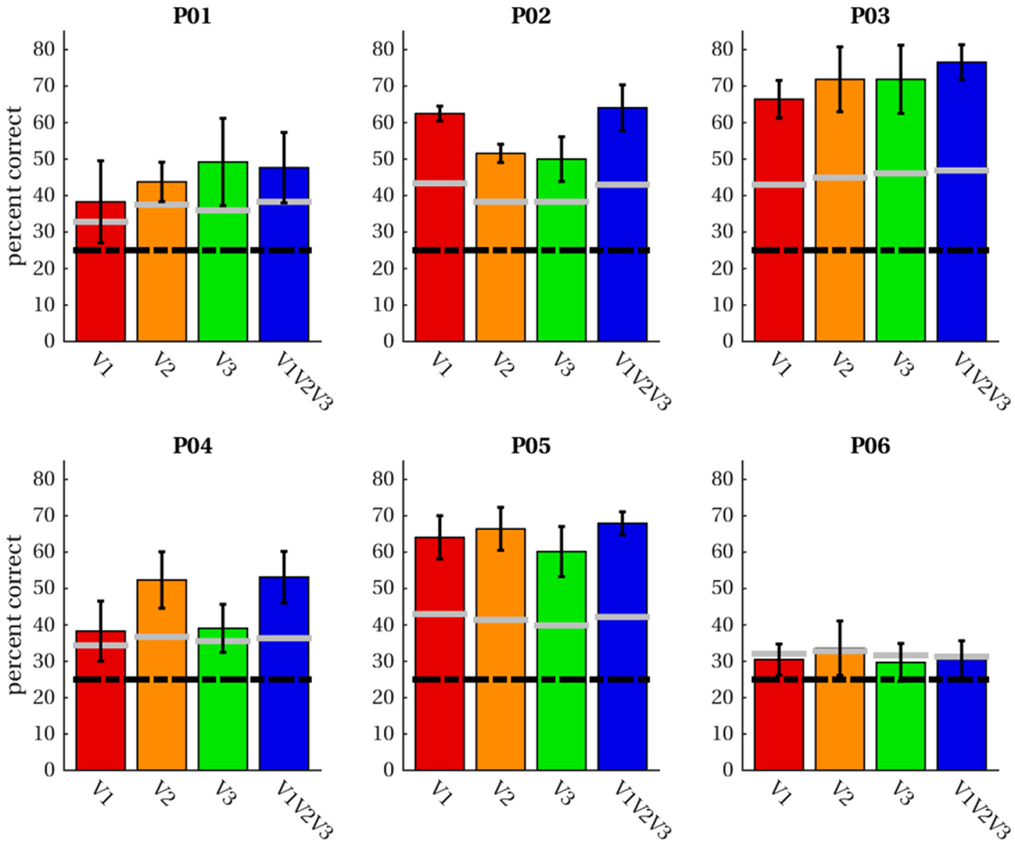



**Figure 4.3.9: classification accuracies.** Average classification accuracies across four leave-one-out runs of imagery data are given for four ROIs in each participant. Classification was performed for letter-specific voxel patterns averaged in the range from +2 until +3 volumes after trial onset. The black dashed line indicates accuracies expected by chance; grey lines demarcate the 95th percentile of permutation classification accuracies.

## 4.4 Discussion

The aim of the present study was to investigate whether visual imagery exhibits sufficient topographic organization to preserve the geometry of internally visualized objects. To that end, we trained participants to maintain a vivid mental image of four letter shapes. Subsequently, we obtained sub-millimeter resolution 7T fMRI measurements from early visual cortex, while participants viewed or imagined the same letter shapes. Finally, we conducted a series of encoding, reconstruction, and decoding analyses to establish the degree of similarity between imagined and perceived shapes. Our results reveal that an object's geometry is preserved during visual mental imagery.

Over training sessions, all participants reached a high probing accuracy for both imagery and perception trials, showing that they could reliably indicate the location of the invisible letter shape in visual space. The ability to imagine the borders of the letter in the absence of visual stimulation suggests that participants were able to generate a precise internal representation of the instructed letter. While providing explicit instructions to participants prohibited them from engaging in a more ecologically valid form of imagery, it is unlikely to fundamentally alter the neural processes underlying imagery. Instead, it allowed us to have a reasonable degree of confidence in the ground truth of imagined shapes. Next, we showed that patterns of voxel activations predicted by a pRF-encoding model and a physical (binary) letter stimulus can account for observed activation patterns in response to mental imagery of the letter corresponding to the physical stimulus. Given that pRF mapping has been shown to accurately predict fMRI responses to visual stimuli (Wandell and Winawer 2015), our results suggest that intrinsic geometric organization of visual experiences is also maintained during visual mental imagery. Our encoding analysis is somewhat reminiscent of that employed by Naselaris et al. (Naselaris et al. 2015) who used a more computational complex encoding model to identify an imagined artwork from a set of candidates. Given that our encoding model is limited to retinotopy, our approach is more restricted in its applications than that detailed in (Naselaris et al. 2015). However, a more restricted approach has the advantage of affording tighter experimental control providing a stronger basis for drawing conclusions. By focusing on a single feature (retinotopic organization), using stimuli differing solely with respect to their geometric properties, and directly comparing the predictions based on each stimulus regarding the activation profiles in early visual cortex, allowed us to draw specific conclusions regarding the topographic organization of mental imagery.

With respect to reconstructions, we found significant overlap between reconstructed imagery and the physical stimulus in terms of object geometry. While we anticipated this given findings that visual mental imagery exhibits retinotopic organization in early visual cortex (Slotnick et al. 2005; Albers et al. 2013; Pearson et al. 2015), these results were, nonetheless, exciting, because the previous reconstructions of mental imagery based on retinotopy did not preserve object geometry (Thirion et al. 2006). Indeed, to the best of our knowledge, we present the first visually recognizable reconstructions of mental imagery, even at the single-trial level. Our first-level correlation metric of reconstruction quality revealed that reconstruction quality of letter 'S' was significantly reduced, while that of letter 'T' was significantly improved with respect to that of letter 'H'. This fits with the notion that stimuli exhibiting finer (coarser) spatial layouts would be harder (easier) to reconstruct. Furthermore, the OSIVQ object score was a significant predictor of first-level reconstruction quality, whereas the OSIVQ verbal score was not. This indicates that participants relying on an object-based imagery strategy were generally more successful at imagery of the letter shapes than participants relying on verbal strategies. Our findings are further in line with recent observations that neural overlap between imagery and perception in the visual system depends on experienced imagery vividness (Dijkstra et al. 2017). Interestingly, while inspection of Figs. 4.3.3 and 4.3.4 would suggest that reconstruction quality is ROI-specific, ROIs do not constitute a significant predictor of first-level reconstruction quality. Rather, the number of voxels included for any given ROI determined the quality. However, this does not imply that uncritically adding more voxels will definitely lead to higher classification accuracies. We included only those voxels for which pRF mapping yielded a high fit. It is likely that reconstructions benefit from a large number of voxels, whose pRF can be estimated to a high degree of precision (i.e., which show a strong spatially selective visual response, especially with high-resolution 7T fMRI) rather than a large number of voxels per se. The OSIVQ object score and number of voxels were also significant predictors of second-level reconstruction quality for reasons similar to those just mentioned.

Both our encoding and reconstruction results show that it is possible to extract similar information from perceived and imagined shapes. This possibility has previously been suggested to be strongly indicative of the pictorial nature of vividly experienced mental images (Brogaard and Gatzia 2017). In conjunction with the observation that the object but not the verbal score of the OSIVQ significantly predicts reconstruction quality, these results support the view that

mental imagery is represented pictorially, at least within early visual cortex. At later stages of (visual) processing, mental imagery may become increasingly symbolic. As such, we do not wish to imply that our results settle the imagery debate, as they do not rule out the existence of (additional) symbolic representations.

We further show that the tight topographic correspondence between imagery and perception in early visual cortex allows for improved reconstruction and opens new avenues for classification. Specifically, our results indicate that training a denoising autoencoder on perceptual data creates an attractor landscape with one attractor per perceived letter. Importantly, the resemblance of imagery activation profiles in early visual cortex is sufficiently similar to its perceptual pendant to ensure that activation patterns of a large proportion of imagery trials fall within the attraction domain of the correct letter. The autoencoder then projects these imagery activation patterns onto the corresponding perceptual activation patterns. Though this is not the case for every trial with some being projected onto the wrong perceptual activation pattern pointing to intra-individual fluctuations regarding successful imagery (Dijkstra et al. 2017). Nevertheless, these projections allow for perception-level reconstruction quality even for individual imagery trials for those participants with good imagery ability. It may be an interesting avenue for future research to study the attractor landscape formed through training the autoencoder and investigate under which conditions imagery trials fall inside or outside the attraction domain of each letter. Our current observations regarding the autoencoder imply that perception provides an upper limit on the achievable reconstruction quality. That is, any improvements of perceptual reconstructions, for instance, obtaining a more accurate encoding model by correcting for eye movements during pRF mapping (Hummer et al. 2016) should improve imagery reconstructions as well.

Furthermore, the autoencoder can be utilized to pretrain a classifier purely based on perceptual data before fine-tuning it on imagery data. We showed the feasibility of this approach using it for classifying imagined letters with a high degree of accuracy from at least one region of interest (between 50 and 70p correct) in five out of six participants. Statistical analyses revealed that both the OSIVQ object and OSIVQ spatial scores are significant predictors of classification accuracy. The finding that the OSIVQ spatial score constitutes a significant predictor here indicates that for classification, a cruder retinotopic organization of mental imagery might already be sufficient. Indeed, classification may rather benefit from an increased signal-to-noise ratio (SNR) which could be achieved

by lowering the spatial resolution. Here, we opted for high spatial resolution to obtain precise receptive field estimates (additionally trading temporal resolution for SNR). In any case, successful classification may not be sufficient to draw conclusions regarding the precise geometry of imagined objects. As before, the number of voxels also constitutes a significant predictor of classification accuracy.

The autoencoder enables leveraging perceptual data to improve reconstructions of imagined letters and pretrain classifiers. This may eventually be utilized for the development of content-based BCI letter-speller systems. So far, fMRI-based BCI communication systems have mostly focused on coding schemes arbitrarily mapping brain activity in response to diverse mental imagery tasks (e.g., mental spatial navigation, mental calculation, mental drawing, or inner speech), and hence originating from distinct neural substrates, onto letters of the alphabet (Birbaumer et al. 1999; Sorger et al. 2012). As such, current BCI speller systems do not offer a meaningful connection between the intended letter and the specific content of mental imagery. This is demanding for users, as it requires them to memorize the mapping in addition to performing imagery tasks unrelated to intended letters and words. Our results constitute a proof-of-concept that it may be possible to achieve a more natural, content-based, BCI speller system immediately decoding internally visualized letters from their associated brain activity.

In conclusion, our letter encoding, reconstruction, and classification results indicate that the topographic organization of mental imagery closely resembles that of perception. This lends support to the idea that mental imagery is quasi-perceptual not only in terms of its subjective experience but also in terms of its neural representation and constitutes an important first step towards the development of content-based letter-speller systems.

## Chapter 5

# General Discussion

This thesis started with depictions of how the brain processes visual information and how this information travels via electrical signals to brain regions in a topographically organized way. In addition, the cerebral cortex can be subdivided into parcellations that show functional specialization to specific visual categories. This subdivision is important as accurate parcellation provides a map of where we are in the brain, enabling efficient comparison of results across studies and communication among investigators. It also allows more accurate predictions of functional organization in the case when no functional mapping can be performed, such as retinotopy in blind individuals. Next to processing sensory information, our brains also have the ability to generate visual percepts without any sensory input. We showed that mental imagery shares a tight topographic correspondence with visual perception by decoding imaginary shapes using machine learning and fMRI-based decoding techniques. Now, we discuss the implications of our findings and how the methods described in this thesis can be used to generate artificial images designed to enhance or restore brain functioning.

## 5.1 visfAtlas

Extensive physiological evidence exists for the specialization of visual processing by delineated brain regions. These regions often provide a map of external space, yet the maps represent different types of information. Neurons within a region not only code for the location of an object in visual space, but also provide information about the object's attributes. Each visual area provides a limited analysis, and this information is integrated across dimensions to form recognizable percepts (Ochsner et al. 2013). Hypotheses about the role(s) or functionality of a specific region-of-interest (ROI), or interactions between ROIs, in relation to human behavior are at the center of a substantial body of neuroimaging research (Poldrack 2007; Etzel et al. 2009; Hutchison et al. 2014; Tong et al. 2016), since an ROI-based approach has several advantages: 1) it allows hypothesis-driven comparisons of signals within independently defined ROIs across many different conditions, 2) it increases statistical sensitivity in multisubject analyses, 3) it reduces the number of multiple comparisons present in whole-brain analyses, and 4) it identifies ROIs in each participant's native brain space. Often ROIs are defined functionally using localizers to more accurately target functionally distinct regions that are difficult to delineate using anatomical markers alone. Localizers also allow successive scans to depend on a parcellation that was defined based on independent data. However, in some cases (e.g., in certain clinical populations) it is impossible to delineate regions-of-interest without the use of a brain atlas.

In chapter 2 we showed that a probabilistic atlas including category-selective regions can be a reliable source to filter or select ROIs common to most fMRI subjects, and we validated the utility of the visfAtlas in the following ways:

1. We assessed prediction accuracy of category-specific, motion selective and retinotopic locations in a new subject using a leave-one-out cross-validation procedure. While there was substantial variability in prediction accuracy between ROIs, category-selective voxels in some probabilistic category-selective ROIs could be identified with up to 94% accuracy in left-out subjects. In addition, spatial consistency in both retinotopic and category-selective regions was generally higher after cortex-based alignment as compared to nonlinear volumetric alignment.

2. For category-specific regions we performed a leave-one-subject-out responsivity analysis in volume space and found that voxels assigned to a ROI indeed showed a maximum voxel responsivity to the correct category.

3. Comparing the visfAtlas with an independent dataset acquired with similar localizers, on different subjects and at another facility revealed similar inter-subject spatial consistency of category-specific fROIs.
4. We found that the ROIs in visfAtlas are similarly quantifiable in comparison to previously published atlases regardless of differences in experimental design, stimuli type, and sample size.
5. We reported how inter-subject consistency can be affected by fROI size and intra-session reproducibility.

Overall, most of the category-selective regions can be reliably predicted in new subjects. Inter-subject variability might be dependent on the coupling between macroanatomical landmarks and functional regions, as indicated by differences in cortex-based versus nonlinear volume alignment. Especially pFus-faces, LOS-bodies, ITG-bodies, CoS-places, and motion-selective hMT+ showed striking functional-marcoanatomical coupling when opting for cortex-based alignment. visfAtlas is the first probabilistic atlas to contain such an extensive set of functional regions in occipitotemporal cortex and is now available to the neuroimaging community.

## 5.2 Bayesian optimization of phosphene locations for a cortical prosthesis

Adaption to a life without visual cues can be incredibly difficult for the late blind. Visual cortical prostheses are now a viable therapeutic consideration for restoring a rudimentary form of vision, especially when parts of the retina, optic nerve or LGN are damaged (Gabel, 2016; Farnum & Pelled, 2020). However, current prosthetic designs are still far away from producing optimal phosphene vision. Several scientific and technical challenges need to be addressed, such as inter-individual variability in functional neuroanatomy, the trade-off between implant functionality and invasiveness, and the complex neurosurgical procedure inherent to a cortical implant that includes the positioning of intracortical electrodes. In chapter 3 we presented an extensive exploration and optimization procedure of electrode placements that uses the individual brain anatomy to predict stimulation-evoked phosphene sizes and locations. In a large sample of 362 human hemispheres our pipeline automatically found the electrode configuration and implant insertion trajectory that optimally matches a preset 'ideal' phosphene distribution within predetermined practical constraints. The optimal location and insertion angles of the electrodes for a visual cortical prosthesis are based on Bayesian optimization that efficiently minimizes a cost function which considers the electrode yield in grey matter, visual field coverage of the phosphene distribution, and the relative entropy between the desired phosphene distribution and the predicted phosphene map.

Our results show realistic phosphene map predictions for a thousand-electrode array implanted in primary visual cortex. Central receptive fields located near the surface of the occipital lobe could be comfortably accessed, while regions corresponding to peripheral fields were difficult to target. On the individual level, phosphene maps showed unevenly spaced distributions, likely due to the uniform distribution of the simulated electrode array. This poses a problem for functional phosphene vision, as next-generation pattern-recognition and deep-learning algorithms can possibly only partly compensate for non-uniform phosphene grids. The development of custom electrode layouts with electrode positions following brain curvature might produce more evenly spaced phosphene distributions. In a follow-up study, the pipeline can be applied to optimize virtual implantation of any custom electrode design. On the group-level, phosphene map distributions showed high between-subject variability and an average coverage of the target regions of 57%. Admittingly, the size of each visual field

target map was rather large for a 1000 electrode channel array of which the interspacing position was restricted. On the other hand, the limitations in terms of visual field coverage using a single electrode grid have become clear. The optimized phosphene map distributions suggest that distant parts of the visual field might be more efficiently reached by a cortical prosthesis using a modular approach involving multiple electrode grids. Importantly, our publicly available pipeline can in the future be extended to solve the problem of multiple insertion trajectories and positioning of multiple grids. By exploring the functional-anatomical search-space in which a cortical prosthetic implant can be placed, we provide a better understanding of inter-individual variability in functional neuroanatomy in early visual cortex. In addition, four distinct examples of task-dependent phosphene configurations were probed to get a better idea of the upper and lower limits of possible functional phosphene distributions, using more realistic constraints of electrode parameters. Crucially, our cortical implant simulations can help to minimize surgical risks, and it can be applied in blind subjects for whom anatomical brain scans are available, since the pipeline uses individual brain anatomy as a starting point.

## 5.3 Reading letter shapes from the mind's eye

The futuristic concept of mind reading has been featured in science fiction and fantasy tales. For instance, the popular Netflix show "Black Mirror" envisioned an AI-driven world where people's memories could be retrieved by insurance companies using a brain-computer-interface chip. Life events (such as an accident or car crash) could be replayed like a movie from the user's perspective. In reality, brain-computer interfaces (BCIs) are not nearly that successful at decoding memories, yet. The machines used in functional magnetic resonance imaging (fMRI) are much larger than the aforementioned brain chip, yet currently constitute one of the best tools for decoding brain activity in vivo. The workings of the brain remain an immensely complex puzzle and it will continue to be a great challenge to decode neural content, until we have gathered sufficient pieces. However, with the increasing number of neuroscientific empirical findings and recent improvements of machine learning algorithms, researchers are finding new ways to decode mental imagery from fMRI data.

Initial perceptual decoding successes built on multi-voxel pattern analysis revealed categorical representations of faces and objects (Haxby et al. 2001; Carlson et al. 2003; Kamitani & Tong 2005). Later on, fMRI brain responses to stimuli that were visually very similar could also be decoded (Kay et al. 2008). Nowadays, more sophisticated models such as principal-component-analysis (PCA) (Cowen et al. 2014) and deep generative neural networks trained on large fMRI datasets, even allow for reconstruction of unseen individual faces (Axelrod et al. 2015; van Rullen & Reddy 2019). Importantly, perceptual encoding of these concepts might share the same neural substrates as those required for generation of mental images (Dijkstra et al., 2019). Since neural activity in sensory cortex that was first engaged when an item was initially perceived seems to be re-instantiated during mental imagery (Petrides, 1994; Kosslyn et al., 2001; Curtis and D'Esposito, 2003; Ruchkin et al., 2003; Pasternak and Greenlee, 2005; Ranganath and D'Esposito, 2005), several groups have tried to decode visual mental imagery as well (Thirion et al. 2006; Miyawaki et al. 2008; Schoenmakers et al. 2013). Indeed, Horikawa et al. (2013) were able to decode dreams, to some extent, by monitoring patterns of brain activity during sleep and comparing these to neural responses during awake perception. They suggested that visual imagery during sleep shares the same neural substrate as waking responses to visual stimuli. In addition, Naselaris et al. (2015) identified imagined artworks from a set of candidates using an encoding model based on Gabor wavelets which they fit to voxel activations measured in response to

perception of said artworks. Even though these two examples represent important technological advancements, they rely on complex encoding models in which the nature of the neural underpinnings is obscured. More intuitive encoding models are required to provide insight into the mechanisms with which the brain operates.

In chapter 4 we investigated whether internally visualized letter shapes can be reconstructed from the mind's eye by an inverse retinotopy model. Our decoding process inverts the established relationship between visual field and cortex such that activation of voxels can be projected back into the visual field. Importantly, the approach relies on the assumption that geometric properties of visual stimuli are preserved in their mental representations. In addition, the encoding model differs from previous attempts to decode mental imagery in three ways. Firstly, our method is more restricted in its applications, as the encoding is limited to retinotopic organization. Consequently, we have a stronger basis for drawing conclusions due to tighter experimental control and the fact that stimuli differ solely with respect to their geometric properties. It should only be possible to selectively predict activation response patterns to each imagined stimulus if fine-grained geometric properties are present in the data. Hence, confounding factors such as attention for classification are not likely to occur. Secondly, whereas multivariate pattern classification (e.g. Horikawa et al. 2013) demonstrates reliable distinctions among brain states, the inverse retinotopy approach offers a more direct characterization of how brain states are organized for a specific picture or internally generated image. Finally, the mental image reconstructions presented in chapter 3 are far closer to the ground-truth (i.e. visual stimuli) than those previously reported (Thirion et al. 2006; Naselaris 2015).

While the decoding of visual responses to images or videos is often referred to as 'mind reading', retrieving mental images that have already been processed by the brain and stored in memory is even more impressive, as the richness of the visual representation of reactivated images is diminished (Klein et al. 2000; Cui et al. 2007; Amedi et al 2005; Reddy et al. 2010; Dijkstra et al. 2017). According to the reverse hierarchy theory (Dentico et al. 2014; Dijkstra et al. 2017; Pearson 2019), mental imagery is a voluntary top-down process. Mental images are generated by combining content stored in our memory that our senses have previously been exposed to, and resulting signals are propagated along feedback pathways in the reverse direction along regions in the visual hierarchy. Consequently, one would expect imagery representations to become increasingly similar to their perceptual counterparts moving from V3

to V1. The preservation of geometric features in our imagery decoding results are in line with the reverse hierarchy theory, however we did not find differences in decoding accuracy or reconstruction quality between areas V1 to V3. However, we did not account for differences in signal-to-noise ratio (SNR) or ROI size. Accounting for these possible confounds would be worth pursuing. Indeed, Breedlove et al. (2020) modeled the hierarchy of processing levels using a deep generative network and found that activity in low-level visual areas encode variation in mental images with less precision than seen images. In addition, their fMRI results showed that the decrease in imagery prediction accuracy from high to low level areas matched the corresponding attenuation of SNR.

Finally, the relatively low SNR during imagery in striate cortex, and therefore noisy voxel pattern reconstructions, could partly be overcome in our experiment by reducing the amount of noise. We created an attractor landscape with one attractor per perceived letter by training a denoising autoencoder on perceptual data. Since imagery activation profiles in early visual cortex closely resemble those of perceptual activation profiles, the majority of imagery trials were projected to the attraction domain of the correct letter, effectively allowing perception-level reconstruction quality. Interestingly, this process proved viable even on the single trial level in subjects with good imagery ability. The autoencoder enables leveraging perceptual data to improve reconstructions of imagined letters and pretrain classifiers. Our work constitutes an important first step towards the development of content-based letter-speller systems.

## 5.4 Future direction and recommendations

The chapters presented in this thesis individually help to understand how the brain processes visual information. However, together they can also serve as groundwork for brain computer interfacing. A probabilistic functional atlas like the one presented in chapter 2 can be used for BCIs in the following ways;

1. regions described by the probabilistic atlas are conceivable target candidates for a vision-restoring BCI in blind individuals (chapter 3). All that is required to obtain an estimation of the most probable location of a certain region in a new subject, is to project the atlas onto the new brain (in this case the brain of a blind person), by aligning the anatomy of the individual with the group average (BVaverage) space. The precision of the atlas prediction will depend on the proximity of the new subject's anatomy to the average brain and the inter-individual variability of the functional region. As mentioned before, V1 is the first visual processing unit in the cerebral cortex and therefore an interesting target for a visual cortical implant. In addition, functional atlas predictions of V1 have been shown to be relatively accurate (Benson et al., 2014; O. Hinds et al., 2009; O. P. Hinds et al., 2008; Rosenke et al., 2021). However, the interhemispheric fissure presents an anatomical barrier and consequently long intracortical electrodes are required for stimulating V1 electrically (Trobe, 2001). Future research could explore candidacy of higher visual areas, especially those more easily accessible from the outer surface of the brain. Category- or motion-selective regions determined in chapter 2 might facilitate conscious perception of (artificial) motion and objects.

2. the definition of regions in early visual cortex are usually defined manually, which would be costly in a real-time setting. Building on the letter decoding techniques described in chapter 4, the atlas can be utilized to mask voxels selected in a real-time imagery decoding paradigm, such as a letter-speller BCI. Specifically, ROI size decreases and receptive field size increases from V1 to V3. Knowing which voxels correspond to a certain area a priori, could boost BCI accuracy and save computational resources by reducing the number of voxels (with overlapping receptive fields) on which the decoding process must be performed.

# Impact

Modern neuroscientific research is advancing toward a more profound comprehension of the anatomo-functional coupling underlying brain processing. In line with these developments, we used ultra-high field fMRI, a cutting-edge biomedical imaging tool, to non-invasively study human cognitive functioning with great detail in-vivo. The work presented in this thesis interfaces with three important aspects of human vision, namely visual perception (chapter 2), restoring vision loss (chapter 3), and visual mental imagery (chapter 4).

Chapter 2 describes a probabilistic map of the visual brain called visfAtlas; detailing the functional location and variability of category-specific regions in occipito-temporal cortex. While the traditional localizationist concept of "one region–one function" is shifting to a more dynamic vision of the central nervous system as a complex network (Borner et al. 2007, Reijneveld et al. 2007), functional brain atlases remain crucial to assess connections between parcels and to extract network properties. visfAtlas adds a unique parcellation to the existing pool of functional brain atlases and, in addition, supports reproducibility of empirical studies by adding to the standardization of the selection of voxels (i.e. region-of-interest) related to certain cognitive functions. Furthermore, functional brain atlases potentially reduce the number of resources required for empirical research (i.e., obviate the need for functional localizers), as scanning time is a costly parameter of experimental setups.

Chapter 3 presents a novel methodology using Bayesian optimization to explore high-channel cortical implant locations for functional phosphene vision in a large retinotopy fMRI dataset. This software tool addresses some of the scientific and technical challenges that need to be overcome before a visual prosthesis -that interfaces directly with the brain- may one day become a conventional clinical treatment for blindness. Cortical visual prostheses are especially relevant when treatments like retinal prostheses, stem cell transplants, and gene therapy are not available. This work can benefit clinicians and engineers with new insights into neurosurgical restrictions and manufacturing requirements

for a high-channel-count, biocompatible, chronically implantable neuronal interface for the visual cortex. Importantly, vision loss is paired with devastating reductions in quality of life, autonomy and economic losses to society due to reduced workforce participation. The next generation of visual cortical prostheses could generate artificial assistive vision that is useful in everyday life, using video footage from a camera that is worn by the patient.

In Chapter 4 we revealed a tight topographic correspondence between visual mental imagery and perception, by exploiting the high spatial resolution of fMRI at 7 Tesla, uncovering the retinotopic organization of early visual cortex and combining it with machine-learning techniques to retrieve imagined letter shapes from the mind's eye. Next to providing new insights into the neural underpinnings of mental imagery, our work can act as a foundation for experimental applications in a clinical setting. Our letter imagery encoding scheme offers a direct and natural way for decoding letters without a need for voluntary muscle control or eye-sight. Therefore, our results might constitute an important first step for the development of a letter-speller BCI that can benefit those with paralysis of voluntary muscles, such as people suffering from locked-in syndrome. In addition, visual imagery ability is retained (or exceptionally vivid; Hahamy et al. 2021) in blind individuals and allows for a novel way to investigate functional reorganization associated with vision loss.

Next to tangible scientific impact in the form of publications, a part of the work has been done in collaboration with NESTOR (Neuronal Stimulation for Recovery of Function), a research project that brought together world-leading academic and industry partners that contribute unique, essential knowledge and expertise. Together, NESTOR collaborators exchanged valuable formation aimed to restore everyday-life abilities after vision was lost, and aided the development of a clinically approved cortical visual prosthesis for the late blind.

# Bibliography

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Brain, G. (2016). TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, 265–284. https://doi.org/10.1038/nn.3331

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An Area within Human Ventral Cortex Sensitive to "Building" Stimuli. *Neuron*, *21*(2), 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/J.CUB.2013.05.065

Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., & Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic pace - where and how variable? . *NeuroImage*, *11*, 66–84.

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *J Neurosci*, *29*(34), 10638–10652. https://doi.org/29/34/10638 [pii]\r10.1523/JNEUROSCI.2807-09.2009

Avraham, D., Jung, J., Yitzhaky, Y., & Peli, E. (2021). Retinal prosthetic vision simulation: temporal aspects. *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2552/AC1B6C

Barriga-Rivera, A., Guo, T., Yang, C.-Y., Abed, A. Al, Dokos, S., Lovell, N. H., Morley, J. W., & Suaning, G. J. (2017). High-amplitude electrical stimulation can reduce elicited neuronal activity in visual prosthesis. *Scientific Reports 2017 7:1*, *7*(1), 1–13. https://doi.org/10.1038/srep42682

Barton, J. J. S. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, *2*(Pt 1), 197. https://doi.org/10.1348/174866407X214172

Beauchamp, M. S., Haxby, J. V, Jennings, J. E., & Deyoe, E. A. (1999). An fMRI Version of the Farnsworth – Munsell 100-Hue Test Reveals Multiple Color-selective Areas in Human Ventral Occipitotemporal Cortex. *Cereb Cortex*, *9*(May), 257–263.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI Responses to Video and Point-Light Displays of Moving Humans and Manipulable Objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001. https://doi.org/10.1162/089892903770007380

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4429–4434. https://doi.org/10.1073/pnas.1014818108

Belkacem, A. N., Jamil, N., Palmer, J. A., Ouhbi, S., & Chen, C. (2020). Brain Computer Interfaces for Improving the Quality of Life of Older Adults and Elderly Patients. *Frontiers in Neuroscience*, *0*, 692. https://doi.org/10.3389/FNINS.2020.00692

Benson, N. C., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2014). Correction of Distortion in Flattened Representations of the Cortical Surface Allows Prediction of V1-V3 Functional Organization from Anatomy. *PLOS Computational Biology*, *10*(3), e1003538. https://doi.org/10.1371/JOURNAL.PCBI.1003538

Benson, N. C., Butt, O. H., Datta, R., Radoeva, P. D., Brainard, D. H., & Aguirre, G. K. (2012). The retinotopic organization of striate cortex is well predicted by surface topology. *Current Biology*, *22*(21), 2081–2085. https://doi.org/10.1016/j.cub.2012.09.014

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., Essen, D. C. Van, Yacoub, E., Ugurbil, K., Winawer, J., & Kay, K.

(2018). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision*, *18*(13), 23–23. https://doi.org/10.1167/18.13.23

Benson, N. C., & Winawer, J. (2018). Bayesian analysis of retinotopic maps. *ELife*, *7*, 1–29.

Benson, N. C., Yoon, J. M. D., Forenzo, D., Engel, S. A., Kay, K. N., & Winawer, J. (2021). Variability of the Surface Area of the V1, V2, and V3 Maps in a Large Sample of Human Observers. *BioRxiv*, 2020.12.30.424856. https://doi.org/10.1101/2020.12.30.424856

Bhat, S., Lührs, M., Goebel, R., & Senden, M. (2021). Extremely Fast pRF Mapping for Real-Time Applications. *BioRxiv*, 2021.03.24.436795. https://doi.org/10.1101/2021.03.24.436795

Bihan, D. Le, Turner, R., Zeffiro, T. A., Cuénod, C. A., Jezzard, P., & Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, *90*(24), 11802–11805. https://doi.org/10.1073/PNAS.90.24.11802

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., & Flor, H. (1999). A spelling device for the paralysed. *Nature*, *398*(6725), 297–298. https://doi.org/10.1038/18581

Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, *23*(5), 638–663. https://doi.org/10.1002/acp.1473

Bollen, C. J. M., Guclu, U., Van Wezel, R. J. A., Van Gerven, M. A. J., & Gucluturk, Y. (2019). Simulating neuroprosthetic vision for emotion recognition. *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2019*, 85–87. https://doi.org/10.1109/ACIIW.2019.8925229

Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative Feedback Explains Distinct Brain Activity Codes for Seen and Mental Images. *Current Biology*, *30*(12), 2211-2224.e6. https://doi.org/10.1016/J.CUB.2020.04.014

Brewer., A., & Barton, B. (2012). Visual Field Map Organization in Human Visual Cortex. In *Visual Cortex - Current Status and Perspectives*. InTech. https://doi.org/10.5772/51914

Brindley, G. S., & Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, *196*(2), 479. https://doi.org/10.1113/JPHYSIOL.1968.SP008519

Britten, K. H., & Heuer, H. W. (1999). Spatial Summation in the Receptive Fields of MT Neurons. *Journal of Neuroscience*, *19*(12), 5074–5084. https://doi.org/10.1523/JNEUROSCI.19-12-05074.1999

Brogaard, B., & Gatzia, D. E. (2017). Unconscious Imagination and the Mental Imagery Debate. *Frontiers in Psychology*, *8*, 799. https://doi.org/10.3389/fpsyg.2017.00799

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *NeuroImage*, *155*(April), 437–449. https://doi.org/10.1016/j.neuroimage.2017.03.062

Caspers, J., Zilles, K., Eickhoff, S. B., Schleicher, A., Mohlberg, H., & Amunts, K. (2013). Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct Funct*, *218*(2), 511–526. https://doi.org/10.1007/s00429-012-0411-8

Chen, S. C., Suaning, G. J., Morley, J. W., & Lovell, N. H. (2009). Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research*, *49*(12), 1493–1506. https://doi.org/10.1016/J.VISRES.2009.02.003

Chen, X., Wang, F., Fernandez, E., & Roelfsema, P. R. (2020). Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science*, *370*(6521), 1191–1196. https://doi.org/10.1126/SCIENCE.ABD7435

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2012). Imagery and Perception Share Cortical Representations of Content and Location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Coalson, T. S., Van Essen, D. C., & Glasser, M. F. (2018). The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6356–E6365. https://doi.org/10.1073/pnas.1801582115

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*, 291–307.

de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *ELife*, *4*, 1–14. https://doi.org/10.7554/elife.06564

Dentico, D., Cheung, B. L., Chang, J. Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, *100*, 237–243. https://doi.org/10.1016/J.NEUROIMAGE.2014.05.081

DeWitt, D. A. (2013). *Visual Perception: More Than Meets the Eye | Answers in Genesis*. https://answersingenesis.org/human-body/eyes/visual-perception-more-than-meets-the-eye/

DeYoe, E. a, Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–2386. https://doi.org/10.1073/pnas.93.6.2382

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *The Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dobelle, W. H., Mladejovsky, M. G., & Girvin, J. P. (1974). Artificial vision for the blind: Electrical stimulation of visual cortex offers hope for a functional prosthesis. *Science*, *183*(4123), 440–444. https://doi.org/10.1126/science.183.4123.440

Dobelle, Wm H., Mladejovsky, M. G., Evans, J. R., Roberts, T. S., & Girvin, J. P. (1976). Braille reading by a blind volunteer by visual cortex stimulation. *Nature*, *259*(5539), 111–112. https://doi.org/10.1038/259111a0

Downing, P. E., Downing, P. E., Jiang, Y., Jiang, Y., Shuman, M., Shuman, M., Kanwisher, N., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science (New York, N.Y.)*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Dumoulin, S. O., Bittar, R. G., Kabani, N. J., Baker, C. L., Le Goualher, G., Bruce Pike, G., & Evans, a C. (2000). A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cerebral Cortex*, *10*(5), 454–463. https://doi.org/10.1093/cercor/10.5.454

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Emmerling, T. C., Zimmermann, J., Sorger, B., Frost, M. A., & Goebel, R. (2016). Decoding the direction of imagined visual motion using 7T ultra-high field fMRI. *NeuroImage*, *125*, 61–73. https://doi.org/10.1016/j.neuroimage.2015.10.022

Engel, S A, Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex*, *7*(2), 181–192.

Engel, Stephen A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. N. (1994). fMRI of human visual cortex. In *Nature* (Vol. 369, Issue 6481, p. 525). https://doi.org/10.1038/369525a0

Engell, A. D., & McCarthy, G. (2013). Probabilistic atlases for face and biological motion perception: An analysis of their reliability and overlap. *NeuroImage*, *74*, 140–151. https://doi.org/10.1016/j.neuroimage.2013.02.025

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Farnum, A., & Pelled, G. (2020). New Vision for Visual Prostheses. *Frontiers in Neuroscience*, *0*, 36. https://doi.org/10.3389/FNINS.2020.00036

Flaxman, S. R., Bourne, R. R. A., Resnikoff, S., Ackland, P., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J., Leasher, J., Limburg, H., Naidoo, K., Pesudovs, K., Silvester, A., Stevens, G. A., Tahhan, N., Wong, T., Taylor, H., … Zheng, Y. (2017). Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, *5*(12), e1221–e1234. https://doi.org/10.1016/S2214-109X(17)30393-5

Formisano, E., Linden, D. E. J., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., Grossi, D., Zanella, F. E., & Goebel, R. (2002). Tracking the mind's image in the brain I: Time-resolved fMRI during visuospatial mental imagery. *Neuron*, *35*(1), 185–194. https://doi.org/10.1016/S0896-6273(02)00747-X

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Frost, M. A., & Goebel, R. (2012). Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369–1381. https://doi.org/10.1016/j.neuroimage.2011.08.035

Frost, M. A., & Goebel, R. (2013). Functionally informed cortex based alignment: an integrated approach for whole-cortex macro-anatomical and ROI-based functional alignment. *NeuroImage*, *83*, 1002–1010. https://doi.org/10.1016/j.neuroimage.2013.07.056

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2018). *Cognitive Neuroscience; The biology of the mind* (5th ed.).

Gilaie-Dotan, S., Perry, A., Bonneh, Y., Malach, R., & Bentin, S. (2009). Seeing with profoundly deactivated mid-level visual areas: Non-hierarchical functioning in the human visual cortex. *Cerebral Cortex*, *19*(7), 1687–1703. https://doi.org/10.1093/cercor/bhn205

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., & Yacoub, E. (2016). A multi-modal parcellation of human cerebral cortex. *Nature Publishing Group*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Glasser, M. F., & Van Essen, D. C. (2011). Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI. *The Journal of Neuroscience*, *31*(32), 11597 LP – 11616. https://doi.org/10.1523/JNEUROSCI.2180-11.2011

Goebel, R, Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, *27*(5), 392–401. https://doi.org/10.1002/hbm.20249

Goebel, R, Khorram-Sefat, D., & Muckli, L. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal Of*.

Goebel, Rainer, Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401. https://doi.org/10.1002/HBM.20249

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*(1), 191–202. https://doi.org/10.1016/S0896-6273(00)80526-7

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*(6), 1167–1175. https://doi.org/10.1016/S0896-6273(02)00897-8

Haak, K. V., Cornelissen, F. W., & Morland, A. B. (2012). Population receptive field dynamics in human visual cortex. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037686

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*.

Hashemi, R. H., Bradley, W. G., & Lisanti, C. J. (2012). *MRI: The Basics - Ray Hashman Hashemi, William G. Bradley, Christopher J. Lisanti - Google Books*.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*(6), 1027–1041. https://doi.org/10.1016/S0896-6273(03)00144-2

Haxby, J. V, Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. https://doi.org/10.1016/j.neuron.2011.08.026

Hinds, O. P., Rajendran, N., Polimeni, J. R., Augustinack, J. C., Wiggins, G., Wald, L. L., Diana Rosas, H., Potthast, A., Schwartz, E. L., & Fischl, B. (2008). Accurate prediction of V1 location from cortical folds in a surface coordinate system. *NeuroImage*, *39*(4), 1585–1599. https://doi.org/10.1016/j.neuroimage.2007.10.033

Hinds, O., Polimeni, J. R., Rajendran, N., Balasubramanian, M., Amunts, K., Zilles, K., Schwartz, E. L., Fischl, B., & Triantafyllou, C. (2009). Locating the functional and anatomical boundaries of human primary visual cortex. *NeuroImage*, *46*(4), 915–922. https://doi.org/10.1016/J.NEUROIMAGE.2009.03.036

Huang, T., Chen, X., Jiang, J., Zhen, Z., & Liu, J. (2019). A probabilistic atlas of the human motion complex built from large-scale functional localizer data. *Human Brain Mapping*, *40*(12), hbm.24610. https://doi.org/10.1002/hbm.24610

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/JPHYSIOL.1959.SP006308

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/JPHYSIOL.1968.SP008455

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience[202F?]: The Official Journal of the Society for Neuroscience*, *22*(16), 7195–7205. https://doi.org/20026661

Hummer, A., Ritter, M., Tik, M., Ledolter, A. A., Woletz, M., Holder, G. E., Dumoulin, Schmidt-Erfurth, U., & Windischberger, C. (2016). Eyetracker-based gaze correction for robust mapping of population receptive fields. *NeuroImage*, *142*, 211–224. https://doi.org/10.1016/J.NEUROIMAGE.2016.07.003

Ishai, A., Ungerleider, L., & Haxby, J. (2000). Distributed neural systems for the generation of visual images. *Neuron*.

Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*, 59. https://doi.org/10.3389/fnhum.2014.00059

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kanwisher, N. G., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Https://Doi.Org/10.1152/Jn.00105.2013*, *110*(2), 481–494. https://doi.org/10.1152/JN.00105.2013

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53–60. https://doi.org/10.1016/J.CORTEX.2017.10.012

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Knauff, M., Kassubek, J., Mulack, T., & Greenlee, M. W. (2000). Cortical activation evoked by visual mental imagery as measured by fMRI. *NeuroReport*, *11*(18), 3957–3962. https://doi.org/10.1097/00001756-200012180-00011

Kniestedt, C., & Stamper, R. L. (2003). Visual acuity and its measurement. *Ophthalmology Clinics of North America*, *16*(2), 155–170. https://doi.org/10.1016/S0896-1549(03)00013-0

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*.

Kosslyn, S., Thompson, W., & Alpert, N. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *Neuroimage*.

Kozhevnikov, M., Kozhevnikov, M., Yu, C. J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *British Journal of Educational Psychology*, *83*(2), 196–209. https://doi.org/10.1111/bjep.12013

Kriegeskorte, N., & Goebel, R. (2001). An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage*, *14*(2), 329–346.

Kujovic, M., Zilles, K., Malikovic, A., Schleicher, A., Mohlberg, H., Rottschy, C., Eickhoff, S. B., & Amunts, K. (2013). Cytoarchitectonic mapping of the human dorsal extrastriate cortex. *Brain Struct Funct*, *218*(1), 157–172. https://doi.org/10.1007/s00429-012-0390-9

Kuperman, V. (2000). *Magnetic resonance imaging: physical principles and applications*.

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. G. (2016). Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques. *The Journal of Neuroscience[202F?]: The Official Journal of the Society for Neuroscience*, *36*(5), 1682–1697. https://doi.org/10.1523/JNEUROSCI.3164-15.2016

Lamb, T. D., Collin, S. P., & Pugh, E. N. (2007). Evolution of the vertebrate eye: Opsins, photoreceptors, retina and eye cup. In *Nature Reviews Neuroscience* (Vol. 8, Issue 12, pp. 960–976). Nature Publishing Group. https://doi.org/10.1038/nrn2283

Lee, Sangkyun, Papanikolaou, A., Logothetis, N. K., Smirnakis, S. M., & Keliris, G. A. (2013). A new method for estimating population receptive field topography in visual cortex. *NeuroImage*, *81*, 144–157. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.026

Lee, SH, Kravitz, D., & Baker, C. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*.

Leroy, C. (1755). Mémoire où l'on rend compte de quelques tentatives que l'on a faites pour guérir plusieurs maladies par l'Électricité. In *Histoire de l'Académie Royale des Sciences* (pp. 60–98). https://ci.nii.ac.jp/naid/10018834739/

Li, W. H. (2013). Wearable computer vision systems for a cortical visual prosthesis. *Proceedings of the IEEE International Conference on Computer Vision*, 428–435. https://doi.org/10.1109/ICCVW.2013.63

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature 2008 453:7197*, *453*(7197), 869–878. https://doi.org/10.1038/nature06976

Lorenz, S., Weiner, K. S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau, S., Eickhoff, S. B., Grill-Spector, K., Zilles, K., & Amunts, K. (2015). Two New Cytoarchitectonic Areas on the Human Mid-Fusiform Gyrus. *Cereb Cortex*, 1–13. https://doi.org/10.1093/cercor/bhv225

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron*, *63*(3), 397–405. https://doi.org/10.1016/j.neuron.2009.07.012

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. a, Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(18), 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T 1-mapping at high field. *Neuroimage*, *49*(2), 1271–1281.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. *Cerebral Cortex*. http://cercor.oxfordjournals.org/content/14/11/1256.short

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., & Morito, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*.

Moeller, S., Yacoub, E., & Olman, C. A. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic*.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. https://doi.org/10.1111/j.1751-9004.2009.00170.x.Experience

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *Journal of Neuroscience*, *31*(39), 13771–13785. https://doi.org/10.1523/JNEUROSCI.2792-11.2011

Naumann, J. (2012). *Search for Paradise: A Patient's Account of the Artificial Vision Experiment*.

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. https://doi.org/10.1016/j.neuroimage.2012.06.065

Niketeghad, S., & Pouratian, N. (2019). Brain Machine Interfaces for Vision Restoration: The Current State of Cortical Visual Prosthetics. *Neurotherapeutics*, *16*(1), 134. https://doi.org/10.1007/S13311-018-0660-1

Nowik, K., Langwińska-Wośko, E., Skopiński, P., Nowik, K. E., & Szaflik, J. P. (2020). Bionic eye review – An update. *Journal of Clinical Neuroscience*, *78*, 8–19. https://doi.org/10.1016/J.JOCN.2020.05.041

O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*.

Oleksiak, A., Klink, P. C., Postma, A., Ham, I. J. M. van der, Lankheet, M. J., & Wezel, R. J. A. van. (2011). Spatial summation in macaque parietal area 7a follows a winner-take-all rule. *Https://Doi.Org/10.1152/Jn.00907.2010*, *105*(3), 1150–1158. https://doi.org/10.1152/JN.00907.2010

Orban, G. A., Zhu, Q., & Vanduffel, W. (2014). The transition in the ventral stream from feature to real-world entity representations. *Frontiers in Psychology*, *5*(JUL), 1–9. https://doi.org/10.3389/fpsyg.2014.00695

Osher, D. E., Saxe, R. R., Koldewyn, K., Gabrieli, J. D. E., Kanwisher, N., & Saygin, Z. M. (2015). Structural Connectivity Fingerprints Predict Cortical Selectivity for Multiple Visual Categories across Cortex. *Cerebral Cortex*, 1–16. https://doi.org/10.1093/cercor/bhu303

Pearson, J, Naselaris, T., & Holmes, E. A. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive*.

Pearson, Joel. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience 2019 20:10*, *20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Peelen, M. V., Glaser, B., Vuilleumier, P., & Eliez, S. (2009). Differential development of selectivity for faces and bodies in the fusiform gyrus. *Developmental Science*, *12*(6), 16–25. https://doi.org/10.1111/j.1467-7687.2009.00916.x

Peelen, M. V, & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004

Peirce, J. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*.

Pichiorri, F., & Mattia, D. (2020). Brain-computer interfaces in neurologic rehabilitation practice. *Handbook of Clinical Neurology*, *168*, 101–116. https://doi.org/10.1016/B978-0-444-63934-9.00009-3

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Podgorny, P., & Shepard, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 21–35. https://doi.org/10.1037/0096-1523.4.1.21

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *16*(16), 5205–5215. https://doi.org/10.1523/jneurosci.16-16-05205.1996

Purves, D., Cabeza, R., Huettel, S., LaBar, K., & Platt, M. (2008). *Cognitive neuroscience*.

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences*, *7*(3), 113–118. https://doi.org/10.1016/S1364-6613(03)00003-2

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*(1), 1–24. https://doi.org/10.1037/h0034650

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *NeuroImage*.

Roelfsema, P. R. (2020). Writing to the Mind's Eye of the Blind. *Cell*, *181*(4), 758–759. https://doi.org/10.1016/J.CELL.2020.03.014

Rosenke, M., van Hoof, R., van den Hurk, J., Grill-Spector, K., & Goebel, R. (2021). A Probabilistic Functional Atlas of Human Occipito-Temporal Visual Cortex. *Cerebral Cortex*, *31*(1), 603–619. https://doi.org/10.1093/CERCOR/BHAA246

Rosenke, M., Weiner, K. S., Barnett, M. A., Zilles, K., Amunts, K., Goebel, R., & Grill-Spector, K. (2017a). Data on a cytoarchitectonic brain atlas: effects of brain template and a comparison to a multimodal atlas. *Data in Brief*, *12*, 327–332. https://doi.org/10.1016/j.dib.2017.04.007

Rosenke, M., Weiner, K. S., Barnett, M., Zilles, K., Amunts, K., Goebel, R., & Grill-Spector, K. (2017b). A cross-validated cytoarchitectonic atlas of the human ventral visual stream. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.02.040

Rottschy, C., Eickhoff, S. B., Schleicher, A., Mohlberg, H., Kujovic, M., Zilles, K., & Amunts, K. (2007). Ventral visual cortex in humans: Cytoarchitectonic mapping of two extrastriate areas. *Human Brain Mapping*, *28*(10), 1045–1059.

Sack, A. T., Sperling, J. M., Prvulovic, D., Formisano, E., Goebel, R., Di Salle, F., Dierks, T., & Linden, D. E. J. (2002). Tracking the mind's image in the brain II: Transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. *Neuron*, *35*(1), 195–204. https://doi.org/10.1016/S0896-6273(02)00745-6

Sanchez-Garcia, M., Martinez-Cantin, R., & Guerrero, J. J. (2020). Semantic and structural image segmentation for prosthetic vision. *PLOS ONE*, *15*(1), e0227677. https://doi.org/10.1371/JOURNAL.PONE.0227677

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: a defense of functional localizers. *NeuroImage*, *30*(4), 1088–1089. https://doi.org/10.1016/j.neuroimage.2005.12.062

Schiller, P. H., Slocum, W. M., Kwak, M. C., Kendall, G. L., & Tehovnik, E. J. (2011). New methods devised specify the size and color of the spots monkeys see when striate cortex (area V1) is electrically stimulated. *Proceedings of the National Academy of Sciences*, *108*(43), 17809–17814. https://doi.org/10.1073/PNAS.1108337108

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage*, *32*(3), 1385–1394. https://doi.org/10.1016/j.neuroimage.2006.05.037

Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K., & Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain*, *119*(2), 507–522. https://doi.org/10.1093/brain/119.2.507

Schoenmakers, S., Barth, M., Heskes, T., & van Gerven, M. (2013). Linear reconstruction of perceived images from human brain activity. *NeuroImage*. http://www.sciencedirect.com/science/article/pii/S1053811913007994

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate Face and Body Selectivity on the Fusiform Gyrus. *Journal of Neuroscience*, *25*(47), 11055–11059. https://doi.org/10.1523/JNEUROSCI.2621-05.2005

Senden, M., Reithler, J., Gijsen, S., & Goebel, R. (2014). Evaluating Population Receptive Field Estimation Frameworks in Terms of Robustness and Reproducibility. *PLoS ONE*, *9*(12), e114054. https://doi.org/10.1371/journal.pone.0114054

Sereno, M I, Dale, a M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., Tootell, R. B. H., Series, N., & May, N. (1995). Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging. *268*(5212), 889–893.

Sereno, Martin I., Mcdonald, C. T., & Allman, J. M. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex*, *4*(6), 601–620. https://doi.org/10.1093/cercor/4.6.601

Slotnick, S., Thompson, W., & Kosslyn, S. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*.

Smith, S. M., Jenkinson, M., Woolrich, M. W., & Beckmann, C. F. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. http://www.sciencedirect.com/science/article/pii/S1053811904003933

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of acquired prosopagnosia. *NeuroImage*, *35*(2), 836–852. https://doi.org/10.1016/j.neuroimage.2006.09.051

Sorger, B., Reithler, J., Dahmen, B., & Goebel, R. (2012). A real-time fMRI-based spelling device immediately enabling robust motor-independent communication. *Current Biology*, *22*(14), 1333–1338. https://doi.org/10.1016/j.cub.2012.05.022

Spillmann, L. (2014). Receptive fields of visual neurons: The early years. *Perception*, *43*, 1145–1176. https://doi.org/10.1068/p7721

Steeves, J. K. E., Culham, J. C., Duchaine, B. C., Pratesi, C. C., Valyear, K. F., Schindler, I., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2006). The fusiform face area is not sufficient for face recognition: Evidence from a patient with dense prosopagnosia and no occipital face area. *Neuropsychologia*, *44*(4), 594–609. https://doi.org/10.1016/j.neuropsychologia.2005.06.013

Steveninck, J. de R. van, Güçlü, U., Wezel, R. van, & Gerven, M. van. (2020). End-to-end optimization of prosthetic vision. *BioRxiv*, 2020.12.19.423601. https://doi.org/10.1101/2020.12.19.423601

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *Journal of Neuroscience*, *35*(36), 12412–12424. https://doi.org/10.1523/JNEUROSCI.4822-14.2015

Stokes, M., Saraiva, A., Rohenkohl, G., & Nobre, A. C. (2011). Imagery for shapes activates position-invariant representations in human visual cortex. *Neuroimage*. http://www.sciencedirect.com/science/article/pii/S1053811911002485

Stokes, M., Thompson, R., & Cusack, R. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal Of*. http://www.jneurosci.org/content/29/5/1565.short

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron*, *76*(3), 640–652. https://doi.org/10.1016/j.neuron.2012.08.026

Striem-Amit, E., Dakwar, O., Reich, L., & Amedi, A. (2012). The large-scale organization of "visual" streams emerges without visual experience. *Cerebral Cortex*, *22*(7), 1698–1709. https://doi.org/10.1093/cercor/bhr253

Susilo, T., Yang, H., Potter, Z., Robbins, R., & Duchaine, B. (2015). Normal Body Perception despite the Loss of Right Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *27*(3), 614–622.

Talairach, J., & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*.

Tangermann, M., M??ller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K. J., M??ller-Putz, G. R., Nolte, G., Pfurtscheller, G., Preissl, H., Schalk, G., Schl??gl, A., Vidaurre, C., Waldert, S., & Blankertz, B. (2012). Review of the BCI competition IV. *Frontiers in Neuroscience*, *6*(JULY), 1–31. https://doi.org/10.3389/fnins.2012.00055

Tehovnik, E. J., Slocum, W. M., Carvey, C. E., & Schiller, P. H. (2005). Phosphene Induction and the Generation of Saccadic Eye Movements by Striate Cortex. *Https://Doi.Org/10.1152/Jn.00736.2004*, *93*(1), 1–19. https://doi.org/10.1152/JN.00736.2004

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.-B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, *33*(4), 1104–1116.

Thomas, N. J. T. (1999). Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content. *Cognitive Science*, *23*(2), 207–245. https://doi.org/10.1207/s15516709cog2302_3

Thomas, N. J. T. (2020). *"Mental Imagery", The Stanford Encyclopedia of Philosophy (Fall 2021 Edition)*. https://plato.stanford.edu/entries/mental-imagery/ AnaProDeb

Tootell, R. B., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I., & Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience[202F?]: The Official Journal of the Society for Neuroscience*, *17*(18), 7060–7078.

Trojano, L., Grossi, D., Linden, D. E. J., Formisano, E., Hacker, H., Zanella, F. E., Goebel, R., & Di Salle, F. D. (2000). Matching two imagined clocks: The functional anatomy of spatial analysis in the absence of visual stimulation. *Cerebral Cortex*, *10*(5), 473–481. https://doi.org/10.1093/cercor/10.5.473

van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, *114*(22), E4501–E4510. https://doi.org/10.1073/pnas.1612862114

Vinberg, J., & Grill-Spector, K. (2008). Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. In *Journal of Neurophysiology* (Vol. 99, Issue 3, pp. 1380–1393). American Physiological Society. https://doi.org/10.1152/jn.01223.2007

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning - ICML '08*, 1096–1103. https://doi.org/10.1145/1390156.1390294

Virtanen, P., Gommers, R., & Oliphant, T. E. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/S41592-019-0686-2

Volta, Alexander, J. B. (1832). On the electricity excited by the mere contact of conducting substances of different kinds. In a letter from Mr. Alexander Volta, F. R. S. Professor of Natural Philosophy in the University of Pavia, to the Rt. Hon. Sir Joseph Banks, Bart. K. B. P. R. S. *Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London*, *1*, 27–29. https://doi.org/10.1098/rspl.1800.0016

Wandell, B. A., Brewer, A. A., & Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 693–707. https://doi.org/10.1098/rstb.2005.1628

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, *56*(2), 366–383. https://doi.org/10.1016/J.NEURON.2007.10.012

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research*, *51*(7), 718–737.

Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*, *19*(6), 349–357. https://doi.org/10.1016/J.TICS.2015.03.009

Wang, J., Zhu, H., Liu, J., Li, H., Han, Y., Zhou, R., & Zhang, Y. (2021). The application of computer vision to visual prosthesis. *Artificial Organs*. https://doi.org/10.1111/AOR.14022

Wang, L, Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2014). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex*, 1–21.

Wang, Lin, Yang, L., & Dagnelie, G. (2008). Virtual Wayfinding Using Simulated Prosthetic Vision in Gaze-locked Viewing. *Optometry and Vision Science[202F?]: Official Publication of the American Academy of Optometry*, *85*(11), E1057. https://doi.org/10.1097/OPX.0B013E31818B9F36

Weiner, K. S., Barnett, M. A., Witthoft, N., Golarai, G., Stigliani, A., Kay, K. N., Gomez, J., Natu, V. S., Amunts, K., Zilles, K., & Grill-Spector, K. (2018). Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *NeuroImage*, *170*(April 2017), 373–384. https://doi.org/10.1016/j.neuroimage.2017.04.040

Weiner, K. S., Barnett, M., Lorenz, S., Caspers, J., Stigliani, A., Amunts, K., Zilles, K., Fischl, B., & Grill-Spector, K. (2017). The Cytoarchitecture of Domain-specific Regions in Human High-level Visual Cortex. *Cereb Cortex*, 1–16. https://doi.org/10.1093/cercor/bhw361

Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., Amunts, K., & Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage*, *84*, 453–465. https://doi.org/10.1016/j.neuroimage.2013.08.068

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage*, *52*(4), 1559–1573. https://doi.org/10.1016/j.neuroimage.2010.04.262

Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *NeuroImage*, *56*(4), 2183–2199. https://doi.org/10.1016/j.neuroimage.2011.03.041

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97. https://doi.org/10.1007/s00426-011-0392-x

Weiner, K. S., Jonas, J., Gomez, J., Maillard, L., Brissart, H., Hossu, G., Jacques, C., Loftus, D., Colnat-Coulbois, S., Stigliani, A., Barnett, M. A., Grill-Spector, K., & Rossion, B. (2016). The face-processing network is resilient to focal resection of human visual cortex. *Journal of Neuroscience*, *36*(32), 8425–8440. https://doi.org/10.1523/JNEUROSCI.4509-15.2016

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Winawer, J., Kay, K. N., Foster, B. L., Rauschecker, A. M., Parvizi, J., & Wandell, B. A. (2013). Asynchronous Broadband Signals Are the Principal Source of the BOLD Response in Human Visual Cortex. *Current Biology*, *23*(13), 1145–1153. https://doi.org/10.1016/J.CUB.2013.05.001

Witthoft, N., Nguyen, M., Golarai, G., LaRocque, K. F., Liberman, A., Smith, M. E., & Grill-Spector, K. (2014). Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cerebral Cortex*, *24*(9), 2401–2408. https://doi.org/10.1093/cercor/bht092

Yang, G., Zhou, S., Bozek, J., Dong, H. M., Han, M., Zuo, X. N., Liu, H., & Gao, J. H. (2020). Sample sizes and population differences in brain template construction. *NeuroImage*, *206*, 116318. https://doi.org/10.1016/J.NEUROIMAGE.2019.116318

Zeki, S., Kennard, C., Watson, J. D. G., Lueck, C. J., & Frackowiak, R. S. J. (1991). *Cortex of Functional Specialization in Human Visual*. *17*(March).

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery - Congenital aphantasia. *Cortex*, *73*, 378–380. https://doi.org/10.1016/J.CORTEX.2015.05.019

Zhen, Z., Kong, X.-Z., Huang, L., Yang, Z., Wang, X., Hao, X., Huang, T., Song, Y., & Liu, J. (2017). Quantifying the variability of scene-selective regions: Interindividual, interhemispheric, and sex differences. *Human Brain Mapping*, *38*(4), 2260–2275. https://doi.org/10.1002/hbm.23519

Zimmermann, J., Goebel, R., de Martino, F., van de Moortele, P. F., Feinberg, D., Adriany, G., Chaimow, D., Shmuel, A., U[1E7?]urbil, K., & Yacoub, E. (2011). Mapping the organization of axis of motion selective features in human area mt using high-field fmri. *PLoS ONE*, *6*(12), 1–10. https://doi.org/10.1371/journal.pone.0028716

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center–surround configurations in population receptive fields using fMRI. *Journal of Vision*, *12*(3), 10–10. https://doi.org/10.1167/12.3.10

# Acknowledgements

I would like to take this opportunity to thank all the people in my life during and before my promotion. In particular:

my parents, you are the most caring and selfless people I know. Thank you for your unconditional love and encouragement, without you I would have never enjoyed so many opportunities.

Stacey and James, knowing I can come home to you and seeing your happy faces makes me feel I can do anything.

Rainer, from the very start you gave me the freedom to explore, learn and pursue my interests under your own roof at the Oxfordlaan. I watched and learned how you maintain and expand your network during the many business and conference trips that we shared. I am very grateful to you for being an inspirational supervisor and a wonderful person, with your infectious enthusiasm, vast knowledge and your endless optimism. Thank you for always believing in me.

Mario, for some miraculous reason, you always instantly find the time when I have a question and you come up with a brilliant answer, even though you are involved in what seems like a million projects. It still amazes me how you can have that attention for detail and at the same time a bird's-eye view. Thank you for your kind and uplifting spirit even when things didn't go as planned. I have no doubt you will be a great professor one day.

Joel, even though your main focus was (very understandably) elsewhere, I very much appreciated your input and perspective. I want to thank you for checking in on me every now and then. It means a lot.

My friends at Brain Innovation, Michael, Judith, Florian, Quentin, Armin, Hester, Caroline, Makis, Claudia, it was a pleasure sharing an office with you and I especially value the great memories we made during conferences, Braintrain trips and Christmas dinners.

A special thanks to my brothers Mario and Luigi (aka Leon and Carmine) for the many political and nerdy lunch conversations, pizza and game nights. I really miss those.

The CN research master class of 2015, we shared two hectic years in which I made wonderful friends and we prepared ourselves for PhD-life (if such a thing is possible..). Unfortunately, only me and the girls (Hannah, Laurien, Amaia, Anita) stayed for a PhD in Maastricht. Now, I am happy to see you all succeeded to find happiness in life and your careers. I am grateful we experienced PhD-life together and am looking forward to the reunion.

My paranymphs and dear friends, Lars and Jacob, for keeping me healthy and fit by forcing me to lift heavy weights and chase a ball.

The Cognitive Computational Neuroscience (CCN) group, who welcomed me into their pack, even though I am not part of the Human Brain Project. Many thanks for the weekly fruitful discussions and occasional good laugh.

A special thanks to Minye, Frank, Dimo and Michael, Chris and Esther who were always around to help out with scanning and everything that comes with it.

A shoutout to my ROI painting gang/NESTOR colleagues (Antonio, Feng, Chris and Jaap), thank you for the insightful discussions and incredibly fun programming sessions. We should finally grab those beers now that lockdown is over.

Finally, to everyone that contributed to the conducted research and my development in the field of neuroscience, my sincere gratitude.

# About the author

Henricus Johannes (known as Rick) van Hoof was born in Stiphout (Helmond), the Netherlands, on 25 August 1988. He finished his secondary school at Dr. Knippenbergcollege, Helmond, in 2005. He then obtained his first Bachelor degree in 2010 at Fontys University of Applied Sciences, Eindhoven, in Computer Science and Engineering, specializing in industrial automation. After working at Philips Lighting (Signify) as a software developer for about 8 months, he decided on a career-switch.

In November 2010 he started a second Bachelor programme, Health Sciences, specializing in Human Biology, at Maastricht University. During this time, he gained his first research experience in the field of Neuroscience, with Dr. Henry Otgaar as a research assistant, and during his graduation project on emotional working memory in times of stress, supervised by Dr. Tom Smeets. Inspired by these experiences, he successfully applied to the Clinical and Cognitive Neuroscience Research Master programme and obtained his Master of Research degree in 2015. His Master thesis titled 'Exploring the potential of fMRI Frontal Eye Field neurofeedback to modulate visuospatial attention networks' was completed at Oxford University (Wellcome Centre for Integrative Neuroimaging) under supervision of Dr. Jacinta O'Shea.

Prof. Dr. Rainer Goebel kindly offered Rick a position in his company, Brain Innovation, until a fitting PhD project would come to light. At Brain Innovation, Rick worked as a researcher and software engineer on BrainVoyager-related software and fMRI/EEG research projects. On the 1st of May 2017, he started his PhD in the department of Cognitive Neuroscience at the Faculty of Psychology and Neuroscience, Maastricht University, with Rainer, and daily supervisors Dr. Mario Senden and Dr. Joel Reithler. Since December 2021 he has been continuing the work as a postdoctoral researcher in the same department.

# Publication List

*Scientific articles*

Otgaar, H., Howe, M. L., van Beers, J., **van Hoof, H. J.**, Bronzwaer, N., & Smeets, T. (2014). The Positive Ramifications of False Memories using a Perceptual Closure Task. *Journal of Applied Research in Memory and Cognition*.

Florian Krause, Caroline Benjamins, Judith Eck, Michael Lührs, **Rick van Hoof,**, & Rainer Goebel (2019). Active head motion reduction in magnetic resonance imaging using tactile feedback. *Human brain mapping*, 40(14), 4026-4037.

Mario Senden*, Thomas Emmerling*, **Rick van Hoof***, Martin Frost & Rainer Goebel (2019). Reconstructing imagined letters from early visual cortex reveals tight topographic correspondence between visual mental imagery and perception. *Brain Structure  Function*.

Mona Rosenke*, **Rick van Hoof***, Job van den Hurk, Kalanit Grill-Spector, Rainer Goebel (2020). A probabilistic functional atlas of human occipito-temporal visual cortex. *Cerebral Cortex*.

**Rick van Hoof**, Antonio Lozano, Feng Wang, Chris Klink, Pieter Roelfsema & Rainer Goebel. "Bayesian optimization of high-channel cortical implant locations for functional phosphene vision.". *Manuscript in preparation*.

*Conference contributions*

**Rick van Hoof**[*], Mario Senden[*], Thomas Emmerling, Martin Frost, Rainer Goebel (2017). Single trial letter imagery decoding in early visual cortex with high-field MRI. rtFin2017 (Nara). Poster.

Alexander Kroner[*], **Rick van Hoof**[*], Andreas Bressler, Mario Senden, Rainer Goebel (2019). A large-scale fMRI data set at 7T towards the reconstruction of letters and natural images. OHBM 2019 (Rome). Poster.

**Rick van Hoof**, Salil Bhat, Michael Luhrs, Mario Senden, Rainer Goebel (2020). Real-time reconstruction of letter shapes in the mind's eye. OHBM 2020 (Vancouver). Oral session.

[*]contributed equally

# SUPPLEMENTAL MATERIALS - Chapter 2

*Methodological approach*

To evaluate the effect of number of subjects on the Dice coefficient for a given fROI, we calculated the Dice coefficient with an iterative number of subjects comprising the predicting group maps for each fROI in the visfAtlas. Details for the number of subjects that each fROI was defined in can be found in Table 1 of the main article. For each fROI and hemisphere, respectively, we started with N = 2 subjects where 1 subject was used to predict the other subject. Then we randomly, without replacement, drew N = 3 subjects and used two to predict the third subject. The prediction was quantitively evaluated with the Dice coefficient. For any predicting number of subjects and each fROI, we used the same threshold that was best across all alignment methods (see main article, Methods and Materials), which was 0.2. Within each iteration of a given number of subjects, we cross-validated the dice coefficient for each left-out subject so that there were three different cross-validation iterations for N = 3, since each subject was left out once. The number of subjects was increased until the total N for the respective hemisphere ROI was reached (see x-axis of Suppl. Fig. 1). Next, we repeated this procedure for each N 1000 times (where the same subjects could not be drawn within the same sample, but could for any of the 1000 times) and computed the standard deviation across those iterations. We chose to draw samples 1000 times to control for the fact that there are more possible combinations of lower N than higher N.

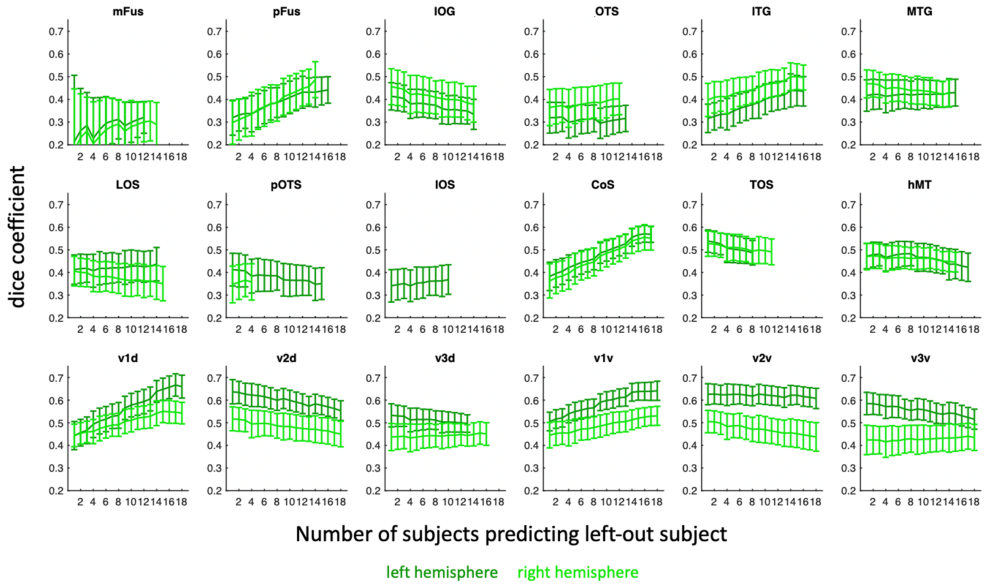



**Figure 5.4.1: Effect of number of subjects in a group map on the Dice coefficient for predicting left-out subjects' fROIs.** For each iterative number of subjects comprising an atlas we tested how well it predicts a left-out data using the Dice coefficient metric. x-axis: number of subjects predicting a left-out subject; y-axis: resulting Dice coefficients. Errorbars: standard deviation across 1000 sample computations.

*Propositions of the PhD-thesis*

**The brain as image processor and generator**
*towards function-restoring brain-computer-interfaces*

Rick van Hoof

1. Human visual perception needs to be studied *in-vivo* before we can understand human vision.
2. Ultra-high field fMRI is an invalueable neuroimaging tool for analyzing human cognitive functions in-vivo at the meso- and macroscopic level.
3. Functional parcellation of cortical brain regions can assist in understanding more complex brain dynamics.
4. Neural responses to specific categories of visual stimuli can be reliably mapped to a flattened model of the cerebral cortex.
5. Retinotopic organization of the visual cortex can partly be derived from anatomical structures, and can therefore also be estimated in blind individuals.
6. Virtual simulations of a visual cortical prosthesis using estimated retinotopic maps can improve visual field coverage of a cortical implant.
7. Studying the process of generating imagined percepts can aid the understanding of visual processing, and vice versa.
8. Visual perception and mental imagery share common circuitry.
9. To understand the mind, we should attempt to grasp the underlying mechanisms.
10. Scientists have the responsibility to evaluate both the risks and potential gains involved with restorative brain-computer-interfaces.
11. The academic publishing system is built on capitalism and is in need of reform.