Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Challenges and opportunities of mesoscopic brain mapping with fMRI

Peter A Bandettini[1], Laurentius Huber[2] and Emily S Finn[3]

Layer fMRI, requiring high field, advanced pulse sequences, and sophisticated processing methods, has emerged in the last decade. The rate of layer fMRI papers published has grown sharply as more groups are overcoming the substantial technical challenges and the unique advantages are becoming clear. Rather than simple delineation of functional activation, layer fMRI promises to provide directional activity and connectivity measures as they are inferred by the layer-specific location. This short review highlights the methods used to achieve layer fMRI as well as the challenges in acquisition, processing and interpretation. It emphasizes the utility of extensive individual averaging as well as the extreme difficulty of cross subject averaging. It also highlights recent studies that have explored layer fMRI in cognitive tasks as well as resting state connectivity studies that have revealed cortical hierarchy.

Addresses
[1] Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA
[2] MR-Methods Group, Maastricht Brain Imaging Center, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[3] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH USA

Corresponding author: Bandettini, Peter A (bandettini@nih.gov)

Current Opinion in Behavioral Sciences 2021, **40**:189–200

This review comes from a themed issue on **Deep imaging – personalized neuroscience**

Edited by **Caterina Gratton** and **Rodrigo M Braga**

For a complete overview see the Issue and the Editorial

Available online 30th June 2021

https://doi.org/10.1016/j.cobeha.2021.06.002

2352-1546/Published by Elsevier Ltd.

Introduction

A complete understanding of the brain will likely integrate data and reveal principles that span all salient temporal and spatial scales. Temporal scales may range from milliseconds to years [1]. Spatial scales range from that of molecules to the entire brain. These scales embody genes, molecules, cells, microcircuits, regions, systems and behavior [2•]. The more effectively we can bridge scales the more likely it is that principles of brain function will be revealed as they likely span spatial and temporal scales. High resolution fMRI, having submillimeter voxel dimensions, promises to connect and integrate our understanding of whole brain systems level with circuit level scales, and has ushered in a new organizational dimension, that of cortical layers, for mapping and understanding human brain function. The ability to map functional activity across cortical depth may allow inferences that will shed light on human cortical hierarchy, connectivity, and computation.

Over the past several years, high resolution fMRI has shown rapid growth as high field MRI scanners, more sophisticated pulse sequences [3,4••,5–7], and advanced processing pipelines [8,9•] have become available. High resolution fMRI research has recently been used to explore cortical depth dependent organization that has provided evidence for feedforward and feedback activity and has efficiently discerned cortical hierarchy. Challenges that the field has grappled with for decades including low quality image acquisition, low sensitivity, poor image registration, and large vessel artifacts, all increase at submillimeter resolution. In this paper, we discuss these challenges and innovations that have been put forward for addressing them. We also address how high resolution challenges such as low sensitivity and poor image registration lend themselves to massive sampling of single subjects, also termed as 'extensive sampling' [10••]. We end with a discussion of recent studies and promising directions for layer fMRI.

Achieving submillimeter resolution fMRI

To complement the following section, a glossary has been included to define and provide context to the terms used. Submillimeter functional MRI resolution requires rapid and temporally stable acquisition strategies as well as functional contrast that is sensitized to small vessels proximal to neuronal activity. Below we describe the challenges involved.

Signal to noise ratio

MRI signal and therefore signal to noise ratio (SNR) increases with field strength and is inversely proportional to voxel volume, setting up a delicate tradeoff for layer fMRI where both SNR and resolution are at a premium and time for averaging signal is limited to about an hour per scan session. At voxel volumes of below 1 mm^3, SNR at 3T typically dips below a workable limit and at 7T, marginally rises above this limit [11], and thermal noise tends to outweigh physiologic noise. Other factors such as radiofrequency (RF) receive coil configuration — with

Glossary

Arterial Spin Labeling (ASL): This is a pulse sequence used to map brain perfusion. It involves an initial inversion pulse — typically outside the imaging plane of interest — to 'label' or 'tag' inflowing blood. After a waiting period to allow blood to flow into the imaging plane and influence the magnetization, images are collected. Image collection is alternated between applied label and no labeling applied. The pairs of images are averaged and subtracted to reveal perfusion.

B:

$_0$: The primary magnetic field in Magnetic Resonance Imaging. It's strength is typically measured in Tesla (T). Typical field strengths are 1.5T–3T, although scanners at 7T are becoming more popular. The highest field strength for human imaging is 11.7T at the National Institutes of Health. The MRI signal is directly proportional to the primary magnetic field.

BOLD contrast: Blood oxygen level dependent contrast. It was first coined by Seiji Ogawa and is the basis of most fMRI contrast. It is based on the fact that oxygenated blood has the same susceptibility as surrounding plasma and tissue, but deoxygenated blood has a lower susceptibility than surrounding plasma and tissue, causing magnetic field distortions and spin dephasing and thus a reduced signal. During activation, blood oxygenation locally increases, causing less spin dephasing and an increased signal.E

cho Planar Imaging (EPI): An MRI pulse sequence in which an entire plane or slice of data is collected following a single RF excitation. It requires rapid gradient switching, however allows collection of an entire volume of MRI data in less than 2 s. This is the most commonly used pulse sequence for fMRI, not only for its speed, but also for the added temporal stability that comes with single shot imaging.

Inversion recovery: A pulse sequence that involves applying an inversion pulse (180° pulse) before the 90° excitation pulse. This allows for T1-weighted scans since the time for the flipped magnetization to recover is determined by T1. It can be used for perfusion imaging using ASL or for blood volume imaging using VASO.

Motion artifact: In the context of fMRI, these artifacts arise from head motion as well as chest wall motion. They typically manifest themselves near large spatial gradients signal intensity such as near edges and regions of signal dropout. Chest wall motion causes field distortions which manifest themselves as changes in image ghosting or image warping. Many ways exist to correct motion artifacts, but these are far from perfect. Motion is still a problem in fMRI.

Physiologic noise: Also known as physiologic fluctuations, this noise currently sets the upper limit on the temporal signal to noise ratio in fMRI time series at about 120/1. Components of physiologic noise are respiration, cardiac pulsation, vasomotion not related to neuronal activity, and movement. After a quarter century of trying, the field is still unable to eliminate or even substantially reduce physiologic noise.

Pial veins: The large veins found at the surface of the cortex and spinal cord that receive blood from capillaries. These are a major confound in attempting to resolve functional columns and layers as they contribute significantly to BOLD contrast and can overwhelm the more subtle spatially specific changes in capillary and small vein oxygenation that occur locally to specific layers and columns.R

adiofrequency(RF) coil: This is the coil that provides a resonant radiofrequency pulse to excite spins. Often the same coil is used to detect the projection of the spin magnetization on the transverse plane. Most commonly, a large head coil is used to provide a uniform excitation and an array of up to 64 receive coils detect the signal as the sensitivity is increased with decreasing size.

Repetition time (TR): In the context of fMRI using single shot EPI, this is the time between each volumetric collection of data. In multi-shot pulse sequences, it is the time between each excitation pulse.

Sensitivity encoding (SENSE): This is an approach to MR imaging that uses the sensitivity profile of multiple receive RF coils in order to assist in spatial encoding. This approach allows either significantly shorter readout windows or higher in plane resolutions for a given readout window duration. The advantages of a shorter readout window in the context of fMRI include: the ability to collect more echoes during the free induction decay, and that images collected with a shorter readout window have less distortion. This technique also enables significantly higher in plane resolution than is possible with a standard single shot approach due to constraints of T2* decay and the biologic limits of gradient switching.

Simultaneous m:

ulti-slice (SMS) imaging: Also known as Multiband Imaging, this is a recent advancement in MRI in which different frequency RF pulses are multiplexed in one pulse to excite multiple parallel planes at the same time. This approach speeds up the acquisition rate by up to ×8 at which an imaging volume can be obtained.

Single-shot MRI: An MR imaging approach in which the raw data necessary to create a single plane, or in extreme cases, volume of data is collected with a single RF excitation pulse or set of RF pulses in the case of single-shot spin-echo imaging. Echo Planar Imaging (EPI) is the most common type of single-shot approach.S

pecific Absorption Rate (SAR): This is a measure of the heat deposition by RF into the body when performing MRI. RF-intensive pulse sequences can have high SAR, limiting their use in humans.

Spin-echo: A spin-echo is formed by the combination of typically a 90° excitation pulse and then a 180° 'refocusing pulse' that inverts the precession of spins and thus brings back into phase spins that have dephased following the 90° pulse. The moment in time that the spins are back in phase is called the spin-echo. The time at which the spin-echo occurs is double the time between the excitation pulse and the refocusing pulse.

Susceptibility: This is property of all materials that is the degree to which a material becomes magnetized in the presence of a magnetic field. Diamagnetic materials repel or diffuse applied magnetic fields and paramagnetic materials attract or concentrate applied magnetic fields.

T1: In MRI, this is the rate constant at which the magnetization exponentially returns to complete recovery of longitudinal magnetization. Different materials have different T1 values.

T2: In MRI, this is the rate constant by which the magnetization as measured using a spin-echo sequence, decays in the transverse plane. Different materials have different T2, and the T2 of blood increases with oxygenation.

T2*: In MRI, this is the rate constant by which the magnetization as measured using a gradient-echo sequence, decays in the transverse plane. T2* is smaller than T2. Different materials have different T2* values, and T2* of blood increases with oxygenation.

TI: This is the time between an inversion pulse and the excitation pulse of an inversion recovery pulse sequence. This time determines the degree of perfusion and/or longitudinal or T1 contrast.V

ascular Sensitivity Imaging (VASO): In the context of fMRI, this is an approach that selectively nulls the blood signal based on the difference between blood T1 and gray matter T1. This allows imaging of blood volume changes as with a blood volume increase the signal would decrease as more nulled blood fills each voxel. In this sequence, an inversion pulse is applied then as the blood longitudinal relaxation is passing through the longitudinal magnetization null point, an excitation pulse is applied, thus only exciting non-blood spins. The contrast used by VASO has been shown to be more specific to small vessel hemodynamic changes than BOLD contrast.

smaller more focal receive elements increasing SNR — may allow layer fMRI at 3T in some instances.

Dropout and warping

At high resolution, signal dropout is reduced with smaller voxel volumes; however, image warping is increased as the readout window duration (the amount of time needed to sample raw data or k-space) associated with high resolution is generally longer. Therefore, registration to standard, un-warped, high-resolution structural scans are

wrought with errors. Advanced nonlinear warping algorithms are considered inadequate for the precision required for layer fMRI. However, because functional time series scans are collected at sufficiently high resolution and anatomic contrast to resolve most fine anatomic structure, functional images may be superimposed directly on the structural images used for fMRI time series, thus obviating the need for image unwarping or registration to a separately acquired structural scan [12].

Acquisition strategies

In high resolution fMRI, several strategies have been used. The field is experiencing a rapid growth in this area. The standard fMRI acquisition of echo-planar imaging (EPI) is insufficient to achieve the resolutions required as T2* or T2 signal decay time is shorter than the required readout window duration. Three strategies have been used to address this limit. The first is the use of parallel imaging approaches [13]. Here, the sensitivity profile of multiple receive RF coils is used in order to assist in spatial encoding. This approach allows either significantly shorter readout windows or higher in plane resolutions for a given readout window duration. The second strategy is the use of simultaneous multi-slice (SMS) imaging [7,14], also known as multiband imaging, in which different frequency RF excitation pulses are multiplexed in one pulse to excite multiple parallel planes at the same time. This approach speeds up the volume and slice acquisition rate by up to a factor of eight. The third strategy is to use full 3D acquisition rather than separate encoding of each slice. For 3D acquisition, time-consuming slice selection is replaced by a third phase encoding direction, and the complete 3D image is then reconstructed. 3D acquisition therefore increases the volume-encoding time efficiency. 3D acquisition is a multi-shot (multiple RF excitations per image) approach and is generally considered more susceptible to physiologic noise; however, fMRI at these extremely high resolutions is dominated by thermal noise and not physiologic noise therefore bypassing to some degree this problem. In addition most 3D sequences use 'navigator' pulses to align k-space phase that is disrupted by physiological fluctuations.

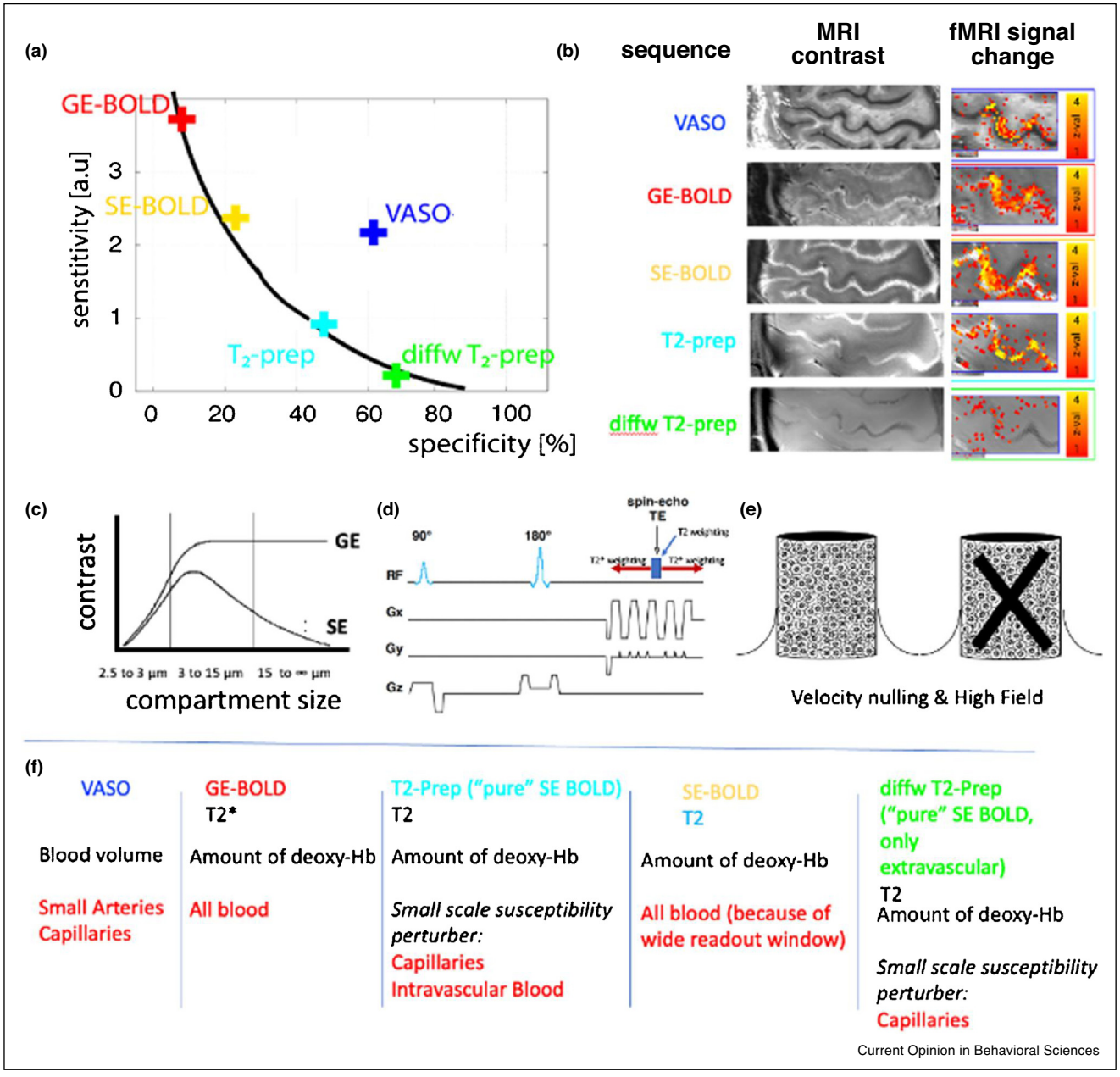
In general, the trend in acquisition strategies for layer fMRI is clearly towards segmented or multi-shot readout strategies, most involving 3D acquisition as mentioned. Each major imaging group appears to be developing their own particular segmented readout strategy. The field of 3D segmented acquisition in the context of layer fMRI is experiencing rapid expansion, and currently, an exhaustive comparison would be premature.

Hemodynamic sensitivity

The central challenge in high functional resolution is sensitization to hemodynamic changes that are localized both spatially and temporally to neuronal activity. Certain pulse sequences are more sensitive to small vessel hemodynamic changes than others. However, this increased specificity comes at a cost to sensitivity to hemodynamic changes. Figure 1 shows four pulse sequences, their functional maps corresponding to motor cortex activation, and where the sequences fall in terms of sensitivity versus specificity.

Here, a brief summary of contrast weightings across five pulse sequences are shown for illustrative purposes. Please refer to Figure 1, which illustrates sensitivity versus specificity measures and functional maps associated with these sequences and describes the contrast mechanisms more in depth. **The first sequence**, *(red)* Blood Oxygenation Level Dependent (BOLD) Contrast using gradient-echo EPI (GE-BOLD), has the highest sensitivity yet is also the least specific to vessel type and therefore has limited spatial specificity. **The second sequence**, *(yellow)* is spin-echo BOLD (SE-BOLD). SE-BOLD involves the use of a 180-degree RF pulse to refocus all spins that become out of phase due to experiencing magnetic field distortions. Hemoglobin, and vessels containing hemoglobin create microscopic magnetic field distortions that cause spin dephasing in proportion to the blood oxygenation. Gradient-echo sequences do not rephase the spins. Spin-echoes rephase all spins except those that have diffused to a slightly different magnetic field in an echo time (TE). Only very small magnetic field distortions set up by red blood cells and capillaries are small enough such that a diffusing spin experiences different magnetic fields in an echo time. Please refer to Figure 1c for a graph depicting relative spin-echo (SE) versus gradient-echo (GE) contrast across susceptibility compartment size. Spin-echo sequences in themselves have three shortcomings. First, as shown in Figure 1c, Spin-echo are insensitive to extravascular gradients set up by large vessels, yet are highly sensitive to intravascular red blood cell (small compartment) effects. Intravascular signal is relatively diminished at 7T due to the shorter baseline T2 of blood, yet it still is present to a sufficient degree that it contributes to the functional contrast. Second, spin-echo EPI uses a long readout window that is mostly positioned outside of where the spin-echo occurs, creating similar sensitivity to large vessels as gradient-echo sequences as shown in Figure 1d. Third the overall sensitivity of SE contrast relative to GE contrast is lower by at a factor between two and four. **The third sequence** *(in teal)* is a 'pure' T2 weighted SE sequence, obtained with a non-EPI readout and a preparation period that mimics a spin-echo. This removes the T2* effect seen in Figure 1d, however at a cost in sensitivity. **The fourth sequence** *(in green)* is a pure spin echo sequence with the addition of diffusion weighting. As seen in the cartoon in Figure 1e, diffusion weighting (also known as velocity nulling), removes the rapidly flowing intravascular signal — thus removing all large vessel effects from 'pure' spin-echo sensitization. High field strength also reduces intravascular signal due to the

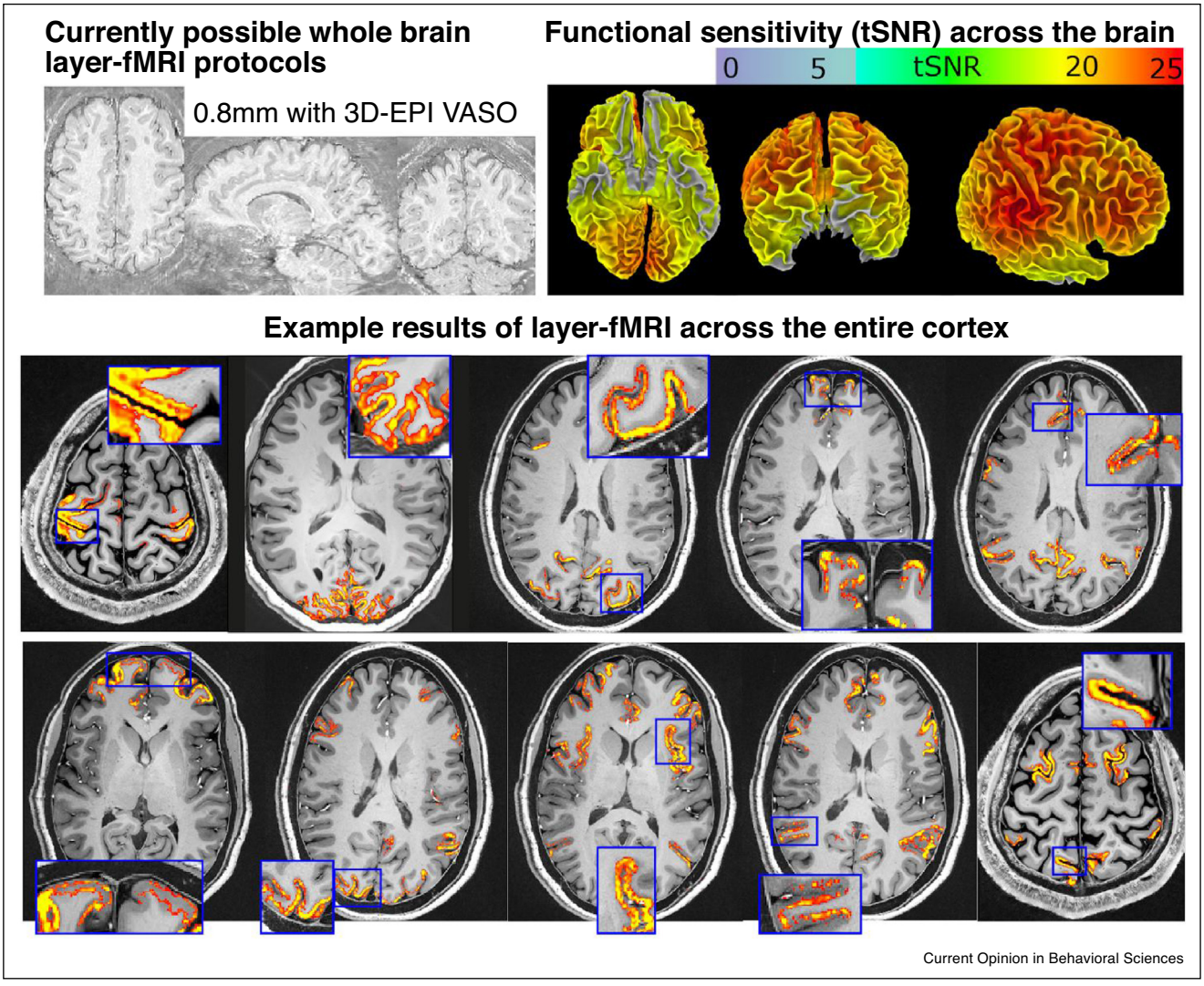
Figure 1



Comparison of basic pulse sequence functional sensitivity and specificity. **(a)** Shows that while GE-BOLD is most sensitive, it is the least specific. Most of the pulse sequences follow a single specificity-sensitivity tradeoff curve, but VASO appears to have the best ratio. **(b)** Shows the anatomic and functional contrast of each sequence in the motor cortex with a finger tapping paradigm. Comparison of contrast sensitivities of the various sequences: **(c)** Depicts the susceptibility compartment size sensitivity of spin-echo and gradient-echo pulse sequences. Note that while GE is sensitive to all compartment sizes above 3 microns, SE shows a selectivity to 3–15 µm sizes. **(d)** Shows the spin-echo sequence, illustrating that due to the long readout window typical of spin-echo, additional T2* (or GE) weighting is present. **(e)** Shows that large vessels also have small intravascular compartments (red blood cells). These are diminished either by shorter T2 of blood than tissue at high field or by the addition of diffusion or 'velocity nulling' gradients. **(f)** Is a depiction of the type of contrast and the vessel sensitivity of each pulse sequence. Black text shows the biophysical contrast, and the red text shows the corresponding vascular contrast.

shortening of blood T2* relative to gray matter at high field. The problem with 'pure' spin-echo sequences with diffusion weighting is that while they are now sensitive only to capillary BOLD effects, there is almost no functional contrast left, therefore they are generally unusable for fMRI. **The fifth sequence** *(in blue)* is known as vascular sensitivity weighted imaging or VASO [15•]. This sequence does not rely on susceptibility contrast,

Figure 2



Current trends of high-resolution fMRI to obtain whole brain datasets with sufficient resolution for obtaining laminar activation patterns. Upper left panel depicts modern sequences such as MAGEC VASO that can sample the entire brain at 0.8 mm resolutions every 6–8 s with relatively high functional sensitivities at 7T. Temporal SNR of VASO is shown in the upper right panel. The MAGEC VASO sequence uses variable flip angles across a long 3D readout window to prohibit a free exponential T1-relaxation. Instead, the variable flip angles allow the experimenter to maintain the T1-related blood volume weighting for longer readouts, ultimately allowing whole brain coverage. The bottom panel shows exemplary functional connectivity maps across almost all cortical areas with area-dependent laminar signatures (see double-stripe versus single stripe patterns).

but rather makes use of the difference in longitudinal relaxation (T1) rate between blood and tissue. In the pulse sequence, an inversion pulse is followed by an excitation pulse at the null point of blood longitudinal relaxation, thus creating a signal void for blood. With an activation-induced blood volume increase, the signal void becomes larger and the activation induced signal change decreases. As can be seen in Figure 1a and b, the sensitivity-specificity tradeoff is more favorable for VASO contast than for the other contrasts, as this has a specificity comparable to 'pure' T2 SE-BOLD contrast yet a sensitivity that is comparable to standard SE-BOLD. A major shortcoming of most versions of this approach is that considerable 'dead' time is spent after the inversion pulse to allow the longitudinal signal to recover. This time delay has limited standard implementations of VASO sequences to a single slab of acquisition and increases the TR or sampling time to the range of 3–5 s.

New pulse sequences to achieve either increased coverage or added sensitivity while maintaining high specificity have continued to emerge. 3D-GRASE [6] has shown

promise in having both sufficient specificity and sensitivity for layer fMRI as well as the ability to cover the whole brain within a TR. 3D-GRASE has more 'pure' spin-echo contrast as the readout window width in the 3D acquisitions are sufficiently short to minimize T2* contrast. The gain in sensitivity over pure spin-echo sequences has been hypothesized to be due to T1 signal enhancement (adding perfusion contrast) from stimulated echoes arising from multiple inversions of lower flip angle than 180°. For comparisons of VASO with more sequences such as 3D-GRASE, please see Ref. [6].

Whole brain VASO pulse sequences have also emerged, as shown in Figure 2. The VASO strategy was generalized to extract CBV changes at any inversion time [16]. As long as there is a different T1 weighting between the extravascular signal and intravascular signal, any volume redistribution between these pools of longitudinal (or 'z') magnetization will result in a VASO signal change. Thus, instead of using an inversion pulse, T1 weighting can also be introduced by variable flip angles that create a dynamic steady-state across k-space segments along the 3D-EPI trajectory. This approach has the advantage that the T1 weighting can be maintained in a dynamic equilibrium for as long as needed. The new sequence, called Multiple Acquisitions with Global Excitation Cycling (MAGEC) [17•] VASO uses multiple inversion pulses and a variable flip angle. Since MAGEC VASO does not rely on a given inversion time, the readout can be prolonged as much as needed (at the cost of TR). This allows for increased coverage with up to 72–104 slices at 0.8 mm isotropic resolution and with TR of 6.5–8 s. Since the blood z-magnetization is not completely nulled, the MAGEC approach may contain helpful cerebral blood flow (CBF) dependent VASO signal amplification. Since CBF is believed to be dominated by capillary water exchange only, this will not compromise the layer-specificity.

Perfusion contrasts using arterial spin-labeling have comparable specificity to VASO, however their sensitivity is lower, preventing their use for layer fMRI. A pulse sequence named VAPER (integrated VASO and PERfusion) has been developed which uses DANTE (Delay Alternating with Nutation for Tailored Excitation) [18] pulses for both nulling blood (blood volume contrast) and tagging blood (perfusion contrast) [5]. During DANTE pulses, blood signal in the microvasculature is nearly nulled to achieve a VASO contrast. After DANTE, fresh blood from outside of the coil coverage flows into the image microvasculature and replaces the nulled blood, generating a perfusion contrast. The signal difference between during-blood suppression and after-blood suppression conditions forms an integrated VASO and perfusion contrast. Both contrasts are sensitive to the microvasculature and add to increase sensitivity. Because no waiting period is needed for the blood to pass through the null point, this approach is more time efficient and therefore allows greater brain coverage per unit time and therefore shorter TR values.

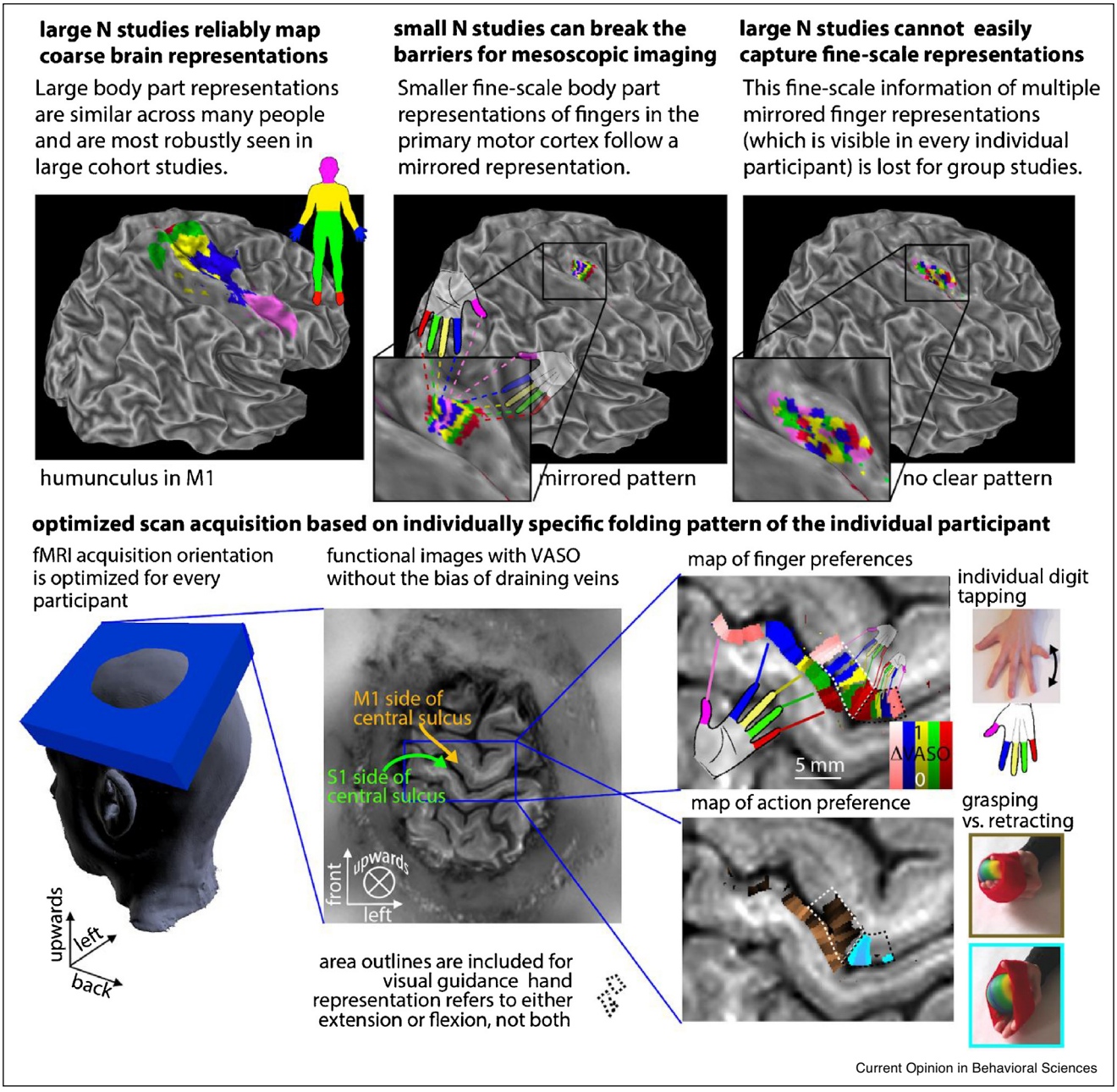
With GE-BOLD contrast, activation-induced signal changes across the layer depth show the largest changes near the pial surface which then ramp down with depth. While simple linear regression or hypercapnic calibration [19] shows promise in normalizing BOLD-based fMRI profiles along the cortical depth, it is challenging to clearly differentiate oxygenation changes within spatially distal 'draining veins' from more localized capillary and small vessels.

Overall, when considering all the above factors and strategies, the respective methods mentioned each have a different 'fingerprint' of variables discussed and those well outside this review. These include: (1) achievable coverage, (2) sampling efficiency, (3) RF power or Specific Absorption Rate (SAR) constraints, (4) required MR-physics expertise and user friendliness, (5) sensitivity-specificity compromise (6) sequence code availability, (7) quantifiability of activity in physically meaningful units, (8) point spread function of unwanted T2 and/or T2* decay. (9) structural contrast aiding subsequent alignment. (10) depth of understanding of the underlying contrast mechanism and signal origin. The field is developing rapidly, and it's still not apparent which sequence (s) will ultimately prevail and come into common practice.

Temporal resolution

Imaging approaches that use gradient-echo or spin-echo EPI, coupled with the time-efficient image sampling strategies mentioned above can achieve whole brain TR values on the order of 1 s, however when using whole brain VASO approaches, the minimum TR values are in the range of 6–8 s. Meaningful neural activation modulations happen across a wide range of temporal frequencies. While depth-dependent electrophysiology studies often focus on the modulation of neural activity, connectivity and phase amplitude coupling changes in the range of 50 ms–300 ms, optical imaging studies examine meaningful resting-state connectivity across from the regime of 100 ms up to the 10 s regime [20]. Because of the hemodynamic delay of the vascular response, conventional resting-state fMRI focuses on signal fluctuations in the time frame of 6−10 s, implying that fMRI is usually only sensitive to a small frequency window of a wide spectrum of neural fluctuations. The larger power of the fMRI frequency spectrum is in the regime of <0.01 Hz. Thus, the acquisition approaches for whole brain layer-dependent connectivity analyses are optimized for this temporal frequency window of ≥ 10 s. Since resting-state fMRI fluctuations follow the pattern of scale free dynamics [21], the focus on this frequency window is expected to be

Figure 3



Imaging fine-scale neural representations in the primary motor cortex with high-end fMRI protocols optimized on each individual's brain. While the fine-scale structures consistently seen in all five participants of the study, they are too variable across people to be seen in group analyses. Each participant had to undergo at least five two-hour scan sessions to reliably capture these fine-scale structures. While large body part representations along the central sulcus are robustly seen in group studies, in the millimeter and submillimeter regime, neuronally meaningful individual differences can be lost. Part of this figure was adopted from Huber *et al.* [23•].

largely representative of functional connections at any temporal scales.

Extensive sampling of individual subjects with ultra-high resolution

In the field of human neuroimaging, there is a strong and justified trend towards increasing sample-sizes in big data initiatives (UK Biobank, Human Connectome Project, Rhineland Study, ENIGMA, Rotterdam Scan Study, ADNI etc.), mainly motivated by the need for high statistical sensitivity in search for clinically actionable biomarkers derivation. As promising and successful as these initiatives are, they have specific shortcomings in addressing neuroscience and clinical questions. Even at

2–3 mm^3 voxel dimensions typical of these scans, fine-grained information that may be contained in the scans may be averaged away across subjects. Additionally, current methodology is not yet capable of precisely registering multi-subject data below a spatial certainty of five to ten millimeters, thus preventing effective meaningful cross-subject pooling and averaging of submillimeter voxel dimension data sets.

Increased functional organization variability at small scales also plays a role in preventing cross subject averaging at high resolution. On a macroscopic scale, the structure of the human brain is very consistent across most individuals. At a spatial scale at or below approximately five millimeters, the gyrification pattern of the brain becomes more variable [22]. Functional organization within macroscopic functional units also shows subject-specific, quasi-random organization at the millimeter scale. A clear example is cortical columns, which resemble a finger-print organization unique for each individual. With regard to layer activity, the precise location on the cortical ribbon of the salient layer function may vary several millimeters between individuals relative to cortical landmarks. Functional activation at this fine spatial scale defies current approaches for spatial normalization, making accurate cross-subject spatial averaging impossible.

As an example, while cross-subject averaging can be easily performed to map the human homunculus, ultra-high resolution found a variable spatial mapping, on the millimeter scale, of individual digit activation [23•]. These representations are differentially engaged depending on the specific motor action such as grasping or releasing as illustrated in Figure 3.

Because of the quasi-random variability of functional organization across subjects at submillimeter resolution, 'extensive sampling' involving the repeated scanning of single subjects is currently a primary way forward for building stable representative functional maps of activity at this spatial scale. Low SNR at high resolution requires repeated averaging to achieve functional statistical significance. Extreme averaging of single subjects has been shown to reveal widespread subthreshold activation [24]. From these maps, obtained only by 'extensive sampling' of single individuals, functional information can be gleaned with sufficient sensitivity and then can be compared across subjects. The field of extensive sampling of single individuals has been growing rapidly due to the emergence of clear, stable, and relevant individual features that are overlooked with group averaging, thus opening up new areas in the potential impact of brain imaging on characterizing individual differences and associating these differences with individual behavior or other measures [25–28].

Currently, the best way to 'collapse' meaningful data within subjects allowing accurate cross-subject comparison or pooling is a source of ongoing research. The current practice involves semi-manual identification of the salient regions and averaging or comparing laminar profile plots. Maps can be qualitatively compared and summary statistics of networks or cortical activation profiles from semi-manually chosen locations can be used for cross-subject comparisons, however, cross subject averaging should not always be a scientific goal. It has been repeatedly shown in vision fMRI literature, information about individuals can be used to derive insights that generalize to populations even in the presence of individual variations.

The increased variability of individual subject activation at finer resolutions is currently only addressable with a single subject extensive sampling approach, however the problem of how to spatially average across subjects may not be entirely insurmountable. Future studies may be able to utilize functional approaches such as 'hyperalignment' [9•] to achieve layer specific alignment across subjects. Rather than aligning data across subjects based on anatomy, hyperalignment and other functional alignment approaches (e.g. shared response modeling) project subject data into a common space based on multivariate functional patterns associated with each voxel. Traditionally, these patterns were temporal activation responses to naturalistic stimulation (i.e. movie watching), but a newer approach involves using functional connectivity (i.e. a vector of correlation values between a given voxel and a set of other spatial locations) as the input to functional alignment algorithms. Such algorithms could, in theory, be adapted to work across voxels spanning both the cortical depth and cortical surface, to enable more precise alignment from different individuals across layers and columns

Recent layer fMRI progress

Precise interpretation of layer fMRI results depends on known laminar functional architecture of the cortex. To the degree that laminar level structural and functional connections are understood, fMRI activity may achieve inferential power to delineate directional connectivity, hierarchy, and perhaps computation, rather than simply location of activity. The current challenge is that layers, their groupings, and their corresponding connection types have been shown to vary across the brain, and in particular, in regions of association cortex [29–31,32••]. Laminar organization may also show variation across species, making translation from even primate organization to human speculative. However, recent studies have demonstrated similarities [33].

The first human layer fMRI studies were performed using gradient-echo contrast at 3T [34]. Layer-fMRI studies have mostly focused on primary visual [35,36,37•], auditory [38,39], motor [40••,41], and somatosensory [42]

cortex. These have been confirmatory in that they have demonstrated that the fMRI signal is robust and reflective of laminar-specific neuronal activity. Recent layer fMRI studies have explored beyond primary cortices into activation of hippocampus [43], dorsolateral prefrontal cortex [44••], and the occipito-temporal sulcus [45]. Review articles on layer fMRI [4••,46•,47,48] have outlined advances in methodology, interpretation, and applications over the past 13 years.

Rather than focusing simply on the areas that layer fMRI can probe, it is useful to highlight the types of questions that layer fMRI can address. It may provide mechanisms of predictive coding. A current construct is that the brain is a prediction engine with higher areas generating models that are relayed to lower areas via feedback connections. Lower areas receive sensory input and calculate prediction errors, then send this information via feedforward connections. Each of these connection terminations and calculations are hypothetically layer-specific [49]. Related to this prediction engine hypothesis, progress has been made elucidating the differential activativation from perception and action [41,50,51]. Laminar-dependent modulating effects of attention [50] and visual learning[52] have been probed. Lawrence *et al.* [46•] outline areas where layer fMRI might be brought to bear on questions in cognitive neuroscience, however a still open question is whether or not layer architecture holds throughout the association cortex.

A current technical hurdle central to the advancement of high-resolution fMRI is the aim to move from imaging protocols that are specialized on individual brain areas with reduced field of views towards imaging the entire cortex as well as subcortical regions at sufficiently high resolution to allow laminar analysis [17•,45]. Even with the latest imaging methodologies, these large coverage approaches still come with constraints of temporal sampling efficiency and thus, require multi-session experiments of several hours each per participant. Processing of this rich multi-dimensional data space is a daunting task that may lend itself to and perhaps even require automated approaches requiring precise yet flexible and adaptive models of cortical folding, laminar thickness, functional cortical landmarks, and layer profiles.

Layer fMRI is poised to uncover a rich tapestry of laminar information related to functional connectivity that promises to uncover detailed information about hierarchal relationships. Resting state fMRI as well as fMRI using naturalistic stimuli may be analyzed in a way that reveals feedforward and feedback activity depending on the seed voxel chosen. A recent study by Huber *et al.* [17•] described visual cortex data from a resting state study in which regions were classified as receiving feedforward or feedback input from a seed which was systematically shifted from the thalamus to increasingly higher regions in visual cortex, clearly delineating, based on the changing laminar connectivity profile with the chosen seed, the cortical areas that were receiving feedforward or feedback information from the seed. Figure 4 shows a summary of these results. It's easy to imagine that the next steps in this type of analysis is map hierarchical relationships through the entire brain, systematically shifting the seed voxel from region to region or adopting a more encompassing approach beyond seed voxel analysis that captures whole brain laminar connectivity and hierarchy in a few processing steps. Certainly, many unexpected complexities in implementation and interpretation will arise, however this example illustrates the novel types of possible analyses approaches exist to explore these rich data.

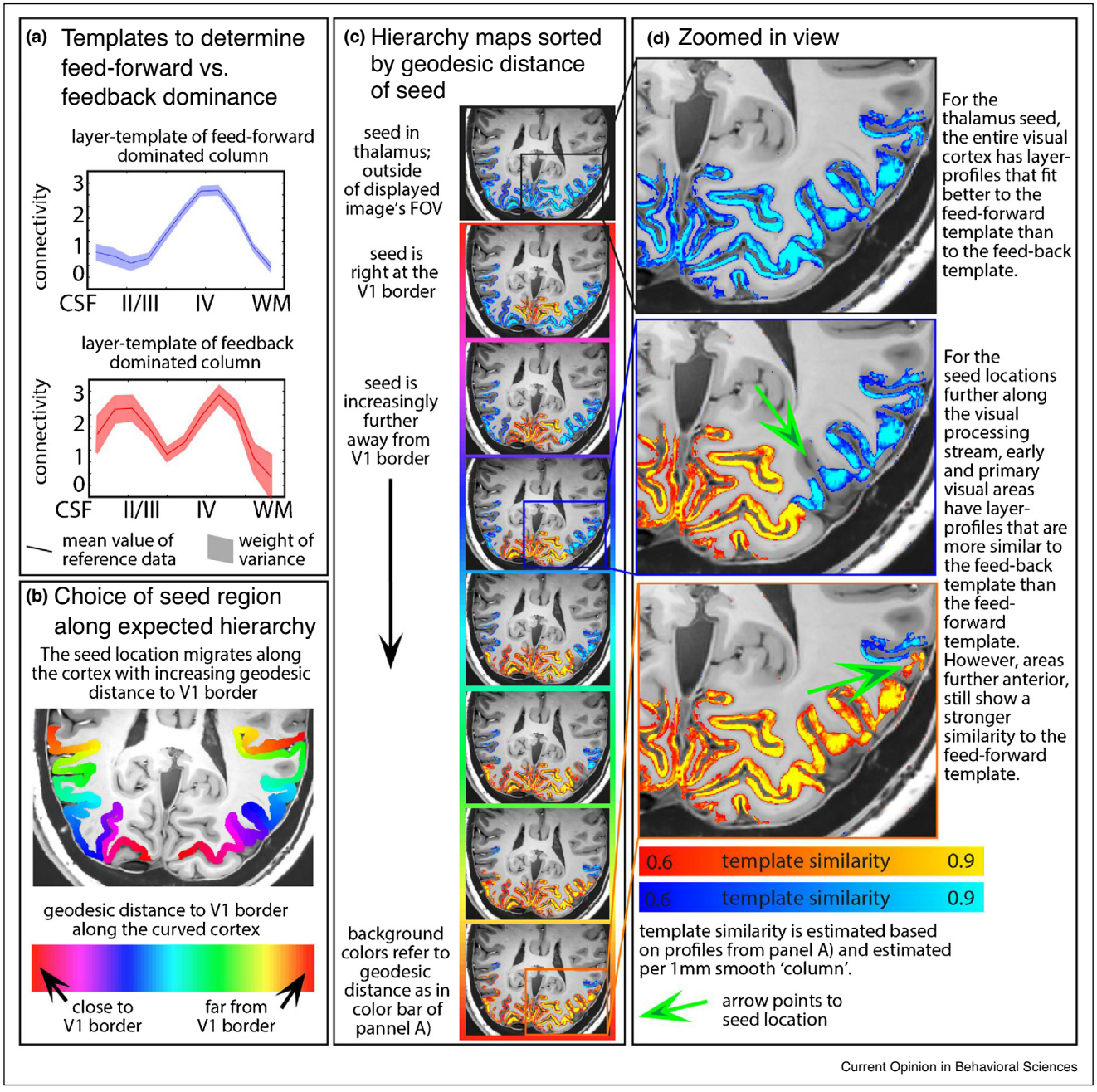
Future challenges and prospects

Layer fMRI is still in its infancy. Early results are promising as clear task modulation and resting state connectivity across cortical depth has been demonstrated. Challenges that face the field are those related to methodology including further increasing functional resolution, mitigating motion, removing large draining vein effects that skew laminar profiles, spatially registering functional activity accurately onto underlying structural images, and developing and refining whole brain pulse sequences that also have sufficient sensitivity and specificity to resolve laminar activity.

Entirely new classes of paradigm designs, created to systematically modulate hypothesized feedforward and feedback activity, are being developed. New analysis approaches are being developed to handle the wealth of multi-dimensional data that is generated. Experiments comparing measures from modalities such as EEG with layer fMRI results are being advanced with intriguing insights suggesting predominant frequencies being signatures of feedforward and feedback activity [37•]. Extensive sampling of 'canonical' individuals is a necessary first step before differences between individuals are determined. The challenge of how to compress and average data across subjects looms large. It is hopeful that relevant 'biomarkers', based on individual differences in laminar activity and connectivity, may be derived.

Lastly, our knowledge of human laminar cortical organization is limited by what is known from the prior invasive electrophysiologic literature to make sense of present and future layer fMRI studies. Convergent studies will have to be carried out to deduce the full implications of laminar activity and functional connectivity in the future. It is speculated that in the next decade with the dissemination of high field scanners and methods more advanced and standardized than those outlined here, groups around the world will be regularly reporting whole brain layer fMRI findings that directly complement and meaningfully add to growing models of brain organization and computation.

Figure 4



Mapping of visual hierarchy using seed-based clustering analysis, from Huber *et al.* [17•]. Panel **(a)** Canonical depth-resolved functional profiles obtained from seed voxels from regions lower in hierarchy (blue) and higher in hierarchy (red). The blue profiles indicate functional connectivity from feed-forward activity from lower regions and the red profiles indicate functional connectivity from feedback activity from higher regions. At 0.8 mm^2 resolution, layer IV cannot be separated from layer V/VI. Even though a given layer profile contains the superposition of feed-forward and feedback peaks, typically one profile dominates for any given seed. Panel **(b)** This indicates the manner in which the seed voxel was systematically migrated along the cortex following geodesic distance from the V1 border. Panel **(c)** Clusters are formed by systematically shifting the seed voxel for correlation analysis. Columns are considered 1 mm smooth patches of cortex. Each column's layer profile can be color coded either red or blue on the map, depending on which layer profile it is most correlated with. Top image in this panel shows clearly that when the seed is chosen in the thalamus, the entire visual cortex demonstrates feed forward activity from the thalamus. As the seed voxel shifts from V1 along the cortical ribbon to just before MT, it can be clearly seen that the regions' correspondence to higher or lower in the hierarchy flips. At V1, only a small patch of visual cortex receives feedback from the seed but most of the cortex receives feedforward information. At the level of the just below MT, most of visual cortex is now receiving feedback from the seed and only MT remains as receiving feedforward information. Panel **(d)** Zoomed image of three images in panel (c).

At this moment in time, it's clear that as a field, layer fMRI is, as was fMRI in its early years, at the very beginning of an increasingly upward trend in utility for addressing unknowns of the human brain.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work was supported by the National Institute of Mental Health Intramural Program ProjectZIA MH002783.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kiebel SJ, Daunizeau J, Friston KJ: **A hierarchy of time-scales and the brain**. *PLoS Comput Biol* 2008, **4**:1-12 http://dx.doi.org/10.1371/journal.pcbi.1000209.

2. • Betzel RF, Bassett DS: **Multi-scale brain networks**. *Neuroimage* 2017, **160**:73-83
This paper discusses multi-scale topological, temporal, and spatial structure of the brain and describes network-based approaches to understand these structures.

3. Wang F, Dong Z, Reese TG, Bilgic B, Katherine Manhard M, Chen J, Polimeni JR, Wald LL, Setsompop K: **Echo planar time-resolved imaging (EPTI)**. *Magn Reson Med* 2019, **81**:3599-3615.

4. •• Norris DG, Polimeni JR: **Laminar (f)MRI: a short history and future prospects**. *Neuroimage* 2019, **197**:643-649
This review article discusses the history and the current state of structural and functional imaging of cortical layer organization.

5. Chai Y, Li L, Huber L, Poser BA, Bandettini PA: **Integrated VASO and perfusion contrast: a new tool for laminar functional MRI**. *Neuroimage* 2020, **207**:116358.

6. Beckett AJS, Dadakova T, Townsend J, Huber L, Park S, Feinberg DA: **Comparison of BOLD and CBV using 3D EPI and 3D GRASE for cortical layer functional MRI at 7T**. *Magn Reson Med* 2020, **84**:3128-3145 http://dx.doi.org/10.1002/mrm.28347.

7. Feinberg DA, Cvetesic N, Beckett A: **Pushing the limits of ultra-high resolution human brain imaging with SMS-EPI demonstrated for columnar level fMRI**. *Neuroimage* 2018, **164**:155-163.

8. Huber L, Poser BA, Bandettini PA, Arora K, Wagstyl K, Cho S, Goense J, Nothnagel N, Morgan AT, Van Den Hurk J *et al.*: **LAYNII: a software suite for layer-fMRI**. *bioRxiv* 2021:1-23 http://dx.doi.org/10.1101/2020.06.12.148080.

9. • Haxby JV, Guntupalli JS, Nastase SA, Feilong M: **Hyperalignment: modeling shared information encoded in idiosyncratic cortical topographies**. *eLife* 2020:1-26 http://dx.doi.org/10.7554/eLife.56601
This article describes a method for functional alignment that avoids the confounds of spatial alignment.

10. •• Naselaris T, Allen E, Kay K: **Extensive sampling for complete models of individual brains**. *Curr Opin Behav Sci* 2021, **40**:45-51
This article summarizes the unique advantages and the need for extensive sampling of single subjects rather than averaging across subjects.

11. Murphy K, Bodurka J, Bandettini PA: **How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration**. *Neuroimage* 2007, **34**:565-574.

12. Kashyap S, Ivanov D, Havlíček M, Poser BA, Uludağ K: **High-resolution T1-mapping using inversion-recovery EPI and application to cortical depth-dependent fMRI at 7 Tesla**. *Proc Int Soc Mag Reson Med* 2016.

13. Blaimer M, Breuer F, Mueller M, Heidemann RM, Griswold MA, Jakob PM: **SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method**. *Top Magn Reson Imaging* 2004, **15**:223-236.

14. Huber L, Ivanov D, Handwerker DA, Marrett S, Guidi M, Uludağ K, Bandettini PA, Poser BA: **Techniques for blood volume fMRI with VASO: from low-resolution mapping towards sub-millimeter layer-dependent applications**. *Neuroimage* 2018, **164**:131-143.

15. • Lu H, Golay X, Pekar JJ, Van Zijl PCM: **Functional magnetic resonance imaging based on changes in vascular space occupancy**. *Magn Reson Med* 2003, **50**:263-274
This is the seminal paper introducing VASO and its basic principles.

16. Ciris PA, Qiu M, Constable RT: **Noninvasive MRI measurement of the absolute cerebral blood volume-cerebral blood flow relationship during visual stimulation in healthy humans**. *Magn Reson Med* 2014, **72**:864-875.

17. • Huber L, Finn ES, Chai Y, Goebel R, Stirnberg R, Stöcker T, Marrett S, Uludag K, Kim SG, Han SH *et al.*: **Layer-dependent functional connectivity methods**. *Prog Neurobiol* 2020:1-19 http://dx.doi.org/10.1016/j.pneurobio.2020.101835
This is a review and demonstration of latest approaches to using layer fMRI methods for probing functional conectivity and heirarchy using resting state and movie viewing tasks.

18. Li L, Miller KL, Jezzard P: **DANTE-prepared pulse trains: a novel approach to motion-sensitized and motion-suppressed quantitative magnetic resonance imaging**. *Magn Reson Med* 2012, **68**:1423-1438.

19. Guidi M, Huber L, Lampe L, Gauthier CJ, Möller HE: **Lamina-dependent calibrated BOLD response in human primary motor cortex**. *Neuroimage* 2016, **141**:250-261.

20. Ma Y, Shaik MA, Kozberg MG, Kim SH, Portes JP, Timerman D, Hillman EMC: **Resting-state hemodynamics are spatiotemporally coupled to synchronized and symmetric neural activity in excitatory neurons**. *Proc Natl Acad Sci U S A* 2016, **113**:E8463-E8471.

21. He BJ: **Scale-free properties of the functional magnetic resonance imaging signal during rest and task**. *J Neurosci* 2011, **31**:13786-13795.

22. Zilles K, Schleicher A, Langemann C, Amunts K, Morosan P, Palomero-Gallagher N, Schormann T, Mohlberg H, Bürgel U, Steinmetz H *et al.*: **Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture**. *Human Brain Mapping*. 1997.

23. • Huber L, Finn ES, Handwerker DA, Bönstrup M, Glen DR, Kashyap S, Ivanov D, Petridou N, Marrett S, Goense J *et al.*: **Sub-millimeter fMRI reveals multiple topographical digit representations that form action maps in human motor cortex**. *Neuroimage* 2020, **208**:116463
This article demonstrates a novel mirror image digit representation in the hand. The specific digit sets show selective activation to contraction or expansion of the hand.

24. Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA, Ana-Lyzed NB: **Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis**. *Proc Natl Acad Sci U S A* 2012, **109**:5487-5492.

25. Poldrack RA, Laumann TO, Koyejo O, Gregory B, Hover A, Chen MY, Gorgolewski KJ, Luci J, Joo SJ, Boyd RL *et al.*: **Long-term neural and physiological phenotyping of a single human**. *Nat Commun* 2015, **6**.

26. Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H *et al.*: **Precision functional mapping of individual human brains**. *Neuron* 2017, **95**:791-807.e7.

27. Gratton C, Laumann TO, Nielsen AN, Greene DJ, Gordon EM, Gilmore AW, Nelson SM, Coalson RS, Snyder AZ, Schlaggar BL *et al.*: **Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation**. *Neuron* 2018, **98**:439-452.e5.

28. Braga RM, Buckner RL: **Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity**. *Neuron* 2017, **95**:457-471.e5.

29. Barbas H: **General cortical and special prefrontal connections: principles from structure to function**. *Annu Rev Neurosci* 2015, **38**:269-289 http://dx.doi.org/10.1146/annurev-neuro-071714-033936.

30. Godlove DC, Maier A, Woodman GF, Schall JD: **Microcircuitry of agranular frontal cortex: testing the generality of the canonical cortical microcircuit**. *J Neurosci* 2014, **34**:5355-5369 http://dx.doi.org/10.1523/JNEUROSCI.5127-13.2014.

31. Markov NT, Vezoli J, Chameau P, Falchier A, Quilodran R, Huissoud C, Lamy C, Misery P, Giroud P, Ullman S *et al.*: **Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex**. *J Comp Neurol* 2014, **522**:225-259.

32. •• Rockland KS: **What do we know about laminar connectivity?** *Neuroimage* 2019, **197**:772-784
This is a review of the anatomic laminar-level connectivity including extrinsic outputs, excitatotory and inhibitory intrinsic connectivity, and intrinsic inputs. This is put in the context of feedforward and feedback connections as well as corticothalamic connections.

33. Huber L, Poser BA, Kaas AL, Fear EJ, Desbach S, Berwick J, Goebel R, Turner R, Kennerley AJ: **Validating layer-specific VASO across species**. *bioRxiv* 2020:1-17 http://dx.doi.org/10.1101/2020.07.24.219378.

34. Ress D, Glover GH, Liu J, Wandell B: **Laminar profiles of functional activity in the human brain**. *Neuroimage* 2007, **34**:74-84 http://dx.doi.org/10.1016/j.neuroimage.2006.08.020.

35. Kok P, Bains LJ, Van Mourik T, Norris DG, De Lange FP: **Selective activation of the deep layers of the human primary visual cortex by top-down feedback**. *Curr Biol* 2016, **26**:371-376.

36. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E: **Contextual feedback to superficial layers of V1**. *Curr Biol* 2015, **25**:2690-2695 http://dx.doi.org/10.1016/j.cub.2015.08.057.

37. • Scheering R, Koopmans PJ, Van Mourik T, Jensen O, Norris DG: **The relationship between oscillatory EEG activity and the laminar-specific BOLD signal**. *Proc Natl Acad Sci U S A* 2016, **113**:6761-6766
This study demonstrates that specific EEG frequencies are differentially amplified in correspondence with specific laminar depths in early visual cortex.

38. Moerel M, De Martino F, Uğurbil K, Yacoub E, Formisano E: **Processing complexity increases in superficial layers of human primary auditory cortex**. *Sci Rep* 2019:1-9 http://dx.doi.org/10.1038/s41598-019-41965-w.

39. De Martino F, Moerel M, Ugurbil K, Goebel R, Yacoub E, Formisano E: **Frequency preference and attention effects across cortical depths in the human primary auditory cortex**. *Proc Natl Acad Sci U S A* 2015, **112**:16036-16041.

40. •• Huber L, Handwerker DA, Jangraw DC, Chen G, Hall A, Stüber C, Gonzalez-Castillo J, Ivanov D, Marrett S, Guidi M *et al.*: **High-resolution CBV-fMRI allows mapping of laminar activity and connectivity of cortical input and output in human M1**. *Neuron* 2017, **96**:1253-1263.e7
This is a seminal paper in laminar fMRI, as it clearly showed modulation of M1 by modulation of sensory input.

41. Persichetti AS, Avery JA, Huber L, Merriam EP, Martin A: **Layer-specific contributions to imagined and executed hand movements in human primary motor cortex**. *Curr Biol* 2020, **30**:1721-1725.e3.

42. Yu Y, Huber L, Yang J, Jangraw DC, Handwerker DA, Molfese PJ, Chen G, Ejima Y, Wu J, Bandettini PA: **Layer-specific activation of sensory input and predictive feedback in the human primary somatosensory cortex**. *Sci Adv* 2019, **5**:eaav9053.

43. Koster R, Chadwick MJ, Chen Y, Berron D, Banino A, Düzel E, Hassabis D, Kumaran D: **Big-loop recurrence within the hippocampal system supports integration of information across episodes**. *Neuron* 2018, **99**:1342-1354 http://dx.doi.org/10.1016/j.neuron.2018.08.009.

44. •• Finn ES, Huber L, Jangraw DC, Molfese PJ, Bandettini PA: **Layer-dependent activity in human prefrontal cortex during working memory**. *Nat Neurosci* 2019, **22**:1687-1695
This was one of the first demonstrations of laminar fMRI revealing modulation with a cognitive task in pre-frontal cortex.

45. Sharoh D, van Mourik T, Bains LJ, Segaert K, Weber K, Hagoort P, Norris DG: **Laminar specific fMRI reveals directed interactions in distributed networks during language processing**. *Proc Natl Acad Sci U S A* 2019, **116**:21185-21190 http://dx.doi.org/10.1073/pnas.1907858116.

46. • Lawrence SJD, Formisano E, Muckli L, de Lange FP: **Laminar fMRI: applications for cognitive neuroscience**. *Neuroimage* 2019, **197**:785-791
This is a review article outlining strategies for applying laminar fMRI towards questions in cognitive neuroscience.

47. Larkum ME, Petro LS, Sachdev RNS, Muckli L: **A perspective on cortical layering and layer-spanning neuronal elements**. *Front Neuroanat* 2018, **12**:56.

48. Finn ES, Huber L, Bandettini PA: Higher and deeper: **Bringing layer fMRI to association cortex.** *Prog Neurobiol*. [date unknown], https://doi.org/10.31234/OSF.IO/3XBZW.

49. Stephan KE, Petzschner FH, Kasper L, Bayer J, Wellstein KV, Stefanics G, Pruessmann KP, Heinzle J: **Laminar fMRI and computational theories of brain function**. *Neuroimage* 2019 http://dx.doi.org/10.1016/j.neuroimage.2017.11.001.

50. Van Kerkoerle T, Self MW, Roelfsema PR: **Layer-specificity in the effects of attention and working memory on activity in primary visual cortex**. *Nat Commun* 2017, **8**.

51. Dijkstra N, Bosch SE, van Gerven MAJ: **Shared neural mechanisms of visual perception and imagery**. *Trends Cogn Sci* 2019, **23**:423-434 http://dx.doi.org/10.1016/j.tics.2019.02.004.

52. Jia K, Zamboni E, Kemper V, Rua C, Goncalves NR, Ng AKT, Rodgers CT, Williams G, Goebel R, Kourtzi Z: **Recurrent processing drives perceptual plasticity**. *Curr Biol* 2020, **30**:4177-4187 http://dx.doi.org/10.1016/j.cub.2020.08.016.