REVIEWS

What we can do and what we cannot do with fMRI

Nikos K. Logothetis[1]

Functional magnetic resonance imaging (fMRI) is currently the mainstay of neuroimaging in cognitive neuroscience. Advances in scanner technology, image acquisition protocols, experimental design, and analysis methods promise to push forward fMRI from mere cartography to the true study of brain organization. However, fundamental questions concerning the interpretation of fMRI data abound, as the conclusions drawn often ignore the actual limitations of the methodology. Here I give an overview of the current state of fMRI, and draw on neuroimaging and physiological data to present the current understanding of the haemodynamic signals and the constraints they impose on neuroimaging data interpretation.

Magnetic resonance imaging (MRI) is the most important imaging advance since the introduction of X-rays by Conrad Röntgen in 1895. Since its introduction in the clinic in the 1980s, it has assumed a role of unparalleled importance in diagnostic medicine and more recently in basic research. In medicine, MRI is primarily used to produce structural images of organs, including the central nervous system, but it can also provide information on the physico-chemical state of tissues, their vascularization, and perfusion. Although all of these capacities have long been widely appreciated, it was the emergence of functional MRI (fMRI)—a technique for measuring haemodynamic changes after enhanced neural activity—in the early 1990s that had a real impact on basic cognitive neuroscience research. A recent database (ISI/Web of Science) query using the keywords 'fMRI' or 'functional MRI' or 'functional magnetic resonance imaging' returned over 19,000 peer-reviewed articles. Given that the first fMRI study without exogenous contrast agents was published in 1991, this corresponds to approximately 1,100 papers per year, or over 3 papers per day. This average obscures the actual rate of publications, as in 1992 there were four publications in total, increasing to about eight per day by 2007. About 43% of papers explore functional localization and/or cognitive anatomy associated with some cognitive task or stimulus—constructing statistical parametric maps from changes in haemodynamic responses from every point in the brain. Another 22% are region of interest studies examining the physiological properties of different brain structures, analogous to single-unit recordings; 8% are on neuropsychology; 5% on the properties of the fMRI signal; and the rest is on a variety of other topics including plasticity, drug action, experimental designs and analysis methods.

In humans, fMRI is used routinely not just to study sensory processing or control of action, but also to draw provocative conclusions about the neural mechanisms of cognitive capacities, ranging from recognition and memory to pondering ethical dilemmas. Its popular fascination is reflected in countless articles in the press speculating on potential applications, and seeming to indicate that with fMRI we can read minds better than direct tests of behaviour itself. Unsurprisingly, criticism has been just as vigorous, both among scientists and the public. In fact, fMRI is not and will never be a mind reader, as some of the proponents of decoding-based methods suggest, nor is it a worthless and non-informative 'neophrenology' that is condemned to fail, as has been occasionally argued.

Perhaps the extreme positions on both sides result from a poor understanding of the actual capacities and limitations of this technology, as well as, frequently, a confusion between fMRI shortcomings and potential flaws in modelling the organizational principles of the faculties under investigation. For example, a frequently made assumption is that the mind can be subdivided into modules or parts whose activity can then be studied with fMRI. If this assumption is false, then even if the brain's architecture is modular, we would never be able to map mind modules onto brain structures, because a unified mind has no components to speak of. Even if true, the challenge remains in coming up with the correct recursive decompositions—in each of which any given cognitive capacity, however abstract, is divided into increasingly smaller functional units that are localized to specific brain parts, which in turn can be detected and studied with fMRI. This is not a neuroimaging problem but a cognitive one. Hierarchical decompositions are clearly possible within different sensory modalities and motor systems. Their mapping, which reflects the brain's functional organization, is evidently possible and certainly meaningful beyond any reasonable doubt[1].

Here, I offer an assessment of fMRI methodology itself, leaving aside such epistemological and ontological issues. I take the modular organization of many brain systems as a well established fact, and discuss only how far fMRI can go in revealing the neuronal mechanisms of behaviour by mapping different system modules and their dynamic inter-relationships. In this context the term module captures the classical local neuronal circuits repeated iteratively within a structure (for example, the columns or swirling, slab-like tangential arrangements of the neocortex), as well as the entities within which modules might be grouped by sets of dominating external connections. The often used term functional segregation refers to such specialized and spatially separated modules. Segregated entities that are interconnected might further result in nested distributed systems, the activity of which, often termed functional integration, can only be visualized by large-scale neuroimaging.

The principal advantages of fMRI lie in its noninvasive nature, ever-increasing availability, relatively high spatiotemporal resolution, and its capacity to demonstrate the entire network of brain areas engaged when subjects undertake particular tasks. One disadvantage is that, like all haemodynamic-based modalities, it measures a surrogate signal whose spatial specificity and temporal response are subject to both physical and biological constraints. A more important

[1]Max Planck Institute for Biological Cybernetics, 72076 Tuebingen, Germany, and Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PL, UK.

shortcoming is that this surrogate signal reflects neuronal mass activity. Although this fact is acknowledged by the vast majority of investigators, its implications for drawing judicious conclusions from fMRI data are most frequently ignored. The aim of this review is first to describe briefly the fMRI technology used in cognitive neuroscience, and then discuss its neurobiological principles that very often limit data interpretation. I hope to point out that the ultimate limitations of fMRI are mainly due to the very fact that it reflects mass action, and much less to limitations imposed by the existing hardware or the acquisition methods. Functional MRI is an excellent tool for formulating intelligent, data-based hypotheses, but only in certain special cases can it be really useful for unambiguously selecting one of them, or for explaining the detailed neural mechanisms underlying the studied cognitive capacities. In the vast majority of cases, it is the combination of fMRI with other techniques and the parallel use of animal models that will be the most effective strategy for understanding brain function.

A brief overview of fMRI

The beautiful graphics MRI and fMRI produce, and the excitement about what they imply, often mask the immense complexity of the physical, biophysical and engineering procedures generating them. The actual details of MRI can only be correctly described via quantum mechanics, but a glimpse of the method's foundation can be also afforded with the tools of classical physics using a few simple equations. (See refs 2 and 3 for a comprehensive account of the theoretical and practical aspects of MRI, and ref. 4 for its functional variants.) Here I offer a brief overview that permits an understandable definition of the terms and parameters commonly used in magnetic resonance imaging (see 'MRI and fMRI principles' in the Supplementary Information for a description of the principles and terms of anatomical and functional MRI). Functional activation of the brain can be detected with MRI via direct measurements of tissue perfusion, blood-volume changes, or changes in the concentration of oxygen. The blood-oxygen-level-dependent (BOLD) contrast mechanism[5,6] is currently the mainstay of human neuroimaging.

Critical factors determining the utility of fMRI for drawing conclusions in brain research are signal specificity and spatial and temporal resolution. Signal specificity ensures that the generated maps reflect actual neural changes, whereas spatial and temporal resolution determine our ability to discern the elementary units of the activated networks and the time course of various neural events, respectively. The interpretability of BOLD fMRI data also depends critically on the experimental design used.

Spatiotemporal properties of BOLD fMRI. The spatiotemporal properties of fMRI are covered in some detail in the Supplementary Information. Briefly, spatial specificity increases with increasing magnetic field strength and for a given magnetic field can be optimized by using pulse sequences that are less sensitive to signals from within and around large vessels (see Fig. 1 and 'Spatial and temporal specificity' in the Supplementary Information). Spatiotemporal resolution is likely to increase with the optimization of pulse sequences, the improvement of resonators, the application of high magnetic fields, and the invention of intelligent strategies such as parallel imaging, for example, sensitivity encoding (SENSE) method (see 'Spatial resolution' section in the Supplementary Information).

Human fMRI can profit a great deal from the use of high-field scanners and by the optimization of the pulse sequences used. Surprisingly, only a minority of the studies in the cognitive sciences seem to exploit the technical innovations reported from laboratories working on magnetic resonance methodologies. Most of the top-cited cognitive neuroscience studies (approximately 70%) were carried out at 1.5 T scanners, 20% were carried out at 3 T scanners, and very few at 2 T or 4 T field strengths. About 87% of all studies used the conventional gradient-echo echoplanar imaging (GE-EPI), whereas the rest used different variants of the spin-echo echoplanar imaging (SE-EPI) sequence. This combination of low magnetic field and traditional GE-EPI is prone to many localization errors. However, as of the beginning of the twenty-first century the percentage of middle-field (3 T) studies has increased, to reach about 56% in 2007. High magnetic fields are likely to dominate magnetic resonance research

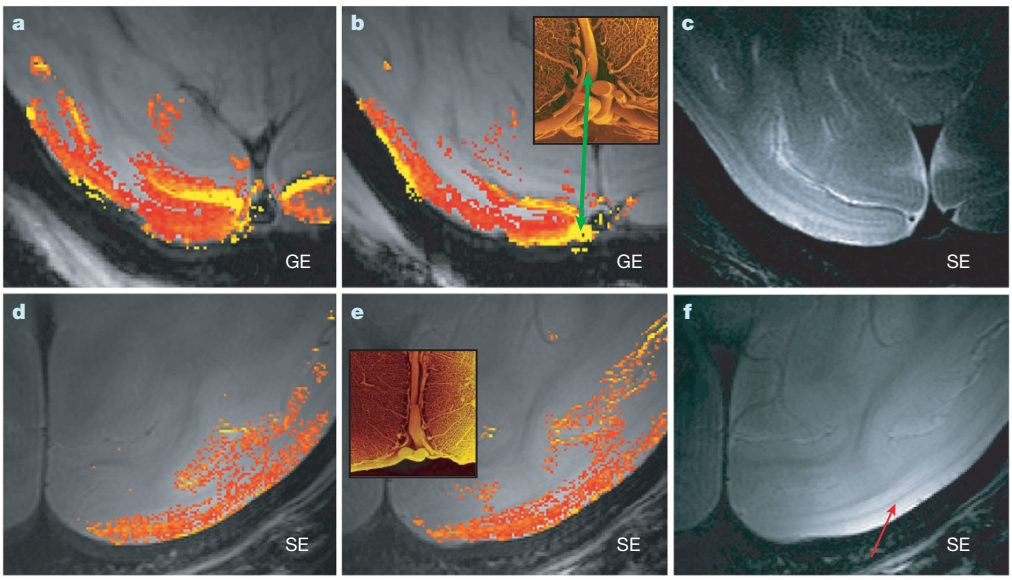

Figure 1 | Specificity of GE-EPI and SE-EPI. Examples of high-resolution GE-EPI and SE-EPI (courtesy J. Goense, MPI for Biological Cybernetics). **a**, **b**, Two slices of GE-EPI demonstrating the high functional signal-to-noise ratio (SNR) of the images, but also the strong contribution of macrovessels. The yellow areas (indicated with the green arrows) are pia vessels, an example of which is shown in the inset scanning electron microscopy image (total width of inset, 2 mm). For the functional images red indicates low and yellow indicates high. In-plane resolution 333 × 333 μm^2; slice thickness 2 mm. **c**, Anatomical scan, SE-EPI, 250 × 188 μm^2, 2 mm slice, with time to echo (TE) and repetition time (TR) 70 and 3,000 ms respectively. **d**, **e**, Two slices of SE-EPI showing the reduction of vascular contribution at the pial side of the cortex. In-plane resolution 250 × 175 μm^2, slice thickness 2 mm. **f**, The anatomical scan is the SE-EPI used for obtaining the functional scans (TE/TR = 48/2,000 ms) but at different greyscale and contrast. The resolution of the anatomical scan permits the clear visualization of the Gennari line (red arrow), the characteristic striation of the primary visual

facilities in the future, and this should definitely improve the quality of data obtained in human magnetic resonance studies. At the same time, high magnetic field scanners are likely to require even tighter interaction between magnetic resonance physicists and application scientists, as the much larger inhomogeneity of both B0 (main static field) and B1 (the field generated by the excitation pulses) at high field will demand a great deal of expertise and experimental skill to achieve the desired image quality.

All in all, MRI may soon provide us with images of a fraction of a millimetre (for example, $300 \times 300\ \mu m^2$ with a couple of millimetres slice thickness or $500 \times 500 \times 500\ \mu m^3$ isotropic), which amount to voxel volumes of about two–three orders of magnitude smaller than those currently used in human imaging (see 'Developments and perspectives' in the Supplementary Information). With an increasing number of acquisition channels such resolution may ultimately be attained in whole-head imaging protocols, yielding unparalleled maps of distributed brain activity in great regional detail and with reasonable—a couple of seconds—temporal resolution. Would that be enough for using fMRI to understand brain function?

The answer obviously depends on the scientific question and the spatial scale at which this question could be addressed—"it makes no sense to read a newspaper with a microscope", as neuroanatomist Valentino Braitenberg once pointed out. To understand the functioning of the microcircuits in cortical columns or of the cell assemblies in the striosomes of basal ganglia, one must know a great deal about synapses, neurons and their interconnections. To understand the functioning of a distributed large-scale system, such as that underlying our memory or linguistic capacities, one must first know the architectural units that organize neural populations of similar properties, and the interconnections of such units. With 10^{10} neurons and 10^{14} connections in the cortex alone, attempting to study dynamic interactions between subsystems at the level of single neurons would probably make little sense, even if it were technically feasible. It is probably much more important to understand better the differential activity of functional subunits—whether subcortical nuclei, or cortical columns, blobs and laminae—and the instances of their joint or conditional activation. If so, whole-head imaging with a spatial resolution, say, of $0.7 \times 0.7\ mm^2$ in slices of 1-mm thickness, and a sampling time of a couple of seconds, might prove optimal for the vast majority of questions in basic and clinical research. More so, because of the great sensitivity of the fMRI signal to neuromodulation (see below and Supplementary Information). Neuromodulatory effects, such as those effected by arousal, attention, memory, and so on, are slow and have reduced spatiotemporal resolution and specificity[7,8].

Designs and analyses. Many studies initially used block designs, reminiscent of earlier positron emission tomography (PET) paradigms. These designs use time-integrated averaging procedures, and usually analyse the data by means of subtraction methods. The central idea is to compare a task state designed to place specific demands on the brain with an investigator-defined control state. Under these conditions, both enhancements and reductions of the fMRI signal are observed. In the early cognitive fMRI studies the prevailing block design was cognitive subtraction, with an emphasis on serial subtraction designs[9]. Such designs rely strictly on pure insertion, which asserts that a single cognitive process can be inserted into a task without affecting the remainder, an assumption that all too often is not tenable (see 'On pure insertion' in the Supplementary Information). Even if an experimental design could satisfy this assumption at the cognitive level, the assumption would be condemned to fail at the level of its neuronal instantiation[10] owing to the highly nonlinear nature of most brain processes. To overcome this kind of problem and ensure better interpretation of the neuroimaging data it is necessary to perform a detailed task analysis to determine subtraction components and their interactions. Yet most neuroimaging studies provide no formal task analysis that would ensure that the particular cognitive process of interest is indeed being isolated by the subtraction[11]. Traditional block designs have excellent functional contrast-to-noise ratio (that is, signal difference between test and control epochs, normalized to the mean signal of all epochs), but they are usually long (from 20 to 60 s), and may be confounded by the general state of arousal of the subject. High-speed fMRI methods, capable of whole-brain imaging with a temporal resolution of a few seconds, enabled the employment of so-called event-related designs[12]. The time course of the response in such experiments is closer to the underlying neural activity.

The block designs discussed so far may reveal differential patterns of activation only in those cases in which different stimulus attributes or different cognitive processes have distinct, non-overlapping spatial organizations. Overlapping networks of neurons subserving different functions are likely to go unnoticed owing to the spatial averaging that characterizes the blocked subtraction paradigms. Functional MRI adaptation designs were conceived as tools that might, at least to some extent, tackle the problem of spatially overlapping neural networks[13]. In this experimental design, a stimulus is presented repeatedly with the expectation that it will eventually induce response adaptation in neurons selective for its various properties. In general, repetition of an identical stimulus does indeed produce a reduction in the fMRI signal. After adaptation, the subject is presented with a stimulus that is varied along one dimension (for example, the direction of a moving pattern or the view of a human face) and the possibility of a response rebound is examined. If the underlying neural representation is insensitive to the changes in the stimulus then the fMRI signal will be reduced, similar to the reduction produced by the repetition of identical stimuli. Alternatively, if the neurons are sensitive to the transformation, the signal will show a clear rebound to its original, pre-adaptation level.

Functional MRI adaptation designs have been widely used in cognitive neuroscience, but they also have shortcomings, as any area receiving input from another region may reveal adaptation effects that actually occurred in that other region, even if the receiving area itself has no neuronal specificity for the adapted property[13]. Moreover, the conclusions of experiments relying on adaptation designs strongly rely on existing electrophysiological evidence, which itself may hold true for one area and not for another[72].

Finally, clever analysis is required to exploit clever design. Most studies so far have used voxel-based conventional analyses of MRI time series from one or more subjects[14]. The approach is predicated on an extension of the general linear model that allows for correlations between error terms owing to physiological noise or correlations that ensue after temporal smoothing. The method is reliable and, when well implemented, offers the best analysis strategy for most studies. Another approach is to take into account the full spatial pattern of brain activity, measured simultaneously at many locations[15]. Such multivariate analyses or pattern-classification-based techniques (decoding techniques) can often detect small differences between two task or stimulus conditions—differences that are not picked up by conventional univariate methods. However, this is not equivalent to saying that they unequivocally reveal the neural mechanisms underlying the activation patterns. The presence, for instance, of voxels selective to two different stimulus attributes could be potentially detected by modern classifiers, yet the existence of two types of patterns does not necessarily imply the existence of two different types of neural populations[72].

What do activation maps represent?

Does the activation of an area mean that it is truly involved in the task at hand? This question implies that we understand what neural activity in a given area would unequivocally show its participation in the studied behaviour. But do we? It is usually alleged that cognitive capacities reflect the 'local processing of inputs' or the 'output' of a region, instantiated in the patterns of action potentials, with their characteristic frequency and timing. In principle, brain structures can be conceptualized as information processing entities, with an input, a local-processing capacity, and an output. Yet, although such a scheme may describe the function of subcortical nuclei, its implementation

in different areas of cortex is anything but straightforward. In fact, we now know that the traditional cortical input–elaboration–output scheme, commonly presented as an instantiation of the tripartite perception–cognition–action model, is probably a misleading over-simplification[16]. Research shows that the subcortical input to cortex is weak; the feedback is massive, the local connectivity reveals strong excitatory and inhibitory recurrence, and the output reflects changes in the balance between excitation and inhibition, rather than simple feedforward integration of subcortical inputs[17]. In the context of this review, the properties of these excitation–inhibition networks (EIN) deserve special attention, and are briefly discussed below.

Feedforward and feedback cortical processing. Brain connectivity is mostly bidirectional. To the extent that different brain regions can be thought of as hierarchically organized processing steps, connections are often described as feedforward and feedback, forward and backward, ascending and descending, or bottom-up and top-down[18]. Although all terms agree on processing direction, endowing backward connections with a role of engineering-type or functional 'feedback' might occasionally be misleading, as under a theoretical generative model perspective on brain function, it is the backward connections that generate predictions and the forward connections that convey the traditional feedback, in terms of mismatch or prediction error signals[19].

In the sensory systems, patterns of long-range cortical connectivity to some extent define feedforward and feedback pathways[20]. The main thalamic input mainly goes to middle layers, whereas second-order thalamic afferents and the nonspecific diffuse afferents from basal forebrain and brain-stem are, respectively, distributed diffusely regionally or over many cortical areas, making synapses mainly in superficial and/or deep layers. Cortical output has thalamic and other subcortical projections originating in layers VI and V, respectively, and corticocortical projections mostly from supragranular layers. The primary thalamic input innervates both excitatory and inhibitory neurons, and communication between all cell types includes horizontal and vertical connections within and between cortical layers. Such connections are divergent and convergent, so that the final response of each neuron is determined by all feedforward, feedback and modulatory synapses[17].

Very few of the pyramid synapses are thalamocortical (less than 10–20% in the input layers of cortex, and less than 5% across its entire depth; in the primary visual cortex the numbers are even lower, with the thalamocortical synapses on stellate cells being about 5%[21]), with the rest originating from other cortical pyramidal cells. Pyramidal axon collateral branches ascend back to and synapse in superficial layers, whereas others distribute excitation in the horizontal plane, forming a strongly recurrent excitatory network[17].

The strong amplification of the input signal caused by this kind of positive feedback loop is set under tight control by an inhibitory network interposed among pyramidal cells and consisting of a variety of GABAergic interneurons[22,23]. These can receive both excitatory and inhibitory synapses on to their somata, and have only local connections. About 85% of them in turn innervate the local pyramidal cells. Different GABAergic cells target different subdomains of neurons[22,24]. Some (for example, basket cells) target somata and proximal dendrites, and are excellent candidates for the role of gain adjustment of the integrated synaptic response; others (for example, chandelier cells) target directly the axons of nearby pyramidal neurons, and appear to have a context-dependent role[25]—they can facilitate spiking during low activity periods, or act like gatekeepers that shunt most complex somatodendritic integrative processes during high activity periods (for example, see up- and down states below). Such nonlinearities might generate substantial dissociations between subthreshold population activity and its concomitant metabolic demand and the spiking of pyramidal cells.

Modules and their microcircuits. A large number of structural, immunochemical and physiological studies, in all cortical areas examined so far, suggested that the functional characteristics of a cortical module are instantiated in a simple basic EIN, referred to as a canonical microcircuit[17] (see also Fig. 2a). Activation of a microcircuit sets in motion a sequence of excitation and inhibition in every neuron of the module, rather than initiating a sequential activation of separate neurons at different hypothetical processing stages. Re-excitation is tightly controlled by local inhibition, and the time evolution of excitation–inhibition is far longer than the synaptic delays of the circuits involved. This means the magnitude and timing of any local mass activation arise as properties of the microcircuits.

Computational modelling suggested that EIN microcircuits, containing such a precisely balanced excitation and inhibition, can account for a large variety of observations of cortical activity, including amplification of sensory input, noise reduction, gain control[26], stochastic properties of discharge rates[27], modulation of excitability with attention[28], or even generation of persisting activity during the delay periods of working memory tasks[29].

The principle of excitation–inhibition balance implies that microcircuits are capable of large changes in activity while maintaining proportionality in their excitatory and inhibitory synaptic conductances. This hypothesis has been tested directly in experiments examining conductance changes during periods of high (up) and low (down) cortical activity. Alternating up states and down states can be readily observed in cerebral cortex during natural sleep or anaesthesia[30], but they can be also induced *in vitro* by manipulating the ionic concentrations in a preparation so that they match those found *in situ*. Research showed that the up state is characterized by persisting synaptically mediated depolarization of the cell membranes owing to strong barrages of synaptic potentials, and a concomitant increase in spiking rate, whereas the down state is marked by membrane hyperpolarization and reduction or cessation of firing[31,32]. Most importantly, the excitation–inhibition conductances indeed changed proportionally throughout the duration of the up state despite large changes in membrane conductance[31,32].

Microcircuits therefore have the following distinct features: (1) the final response of each neuron is determined by all feedforward, feedback and modulatory synapses; (2) transient excitatory responses may result from leading excitation, for example, due to small synaptic delays or differences in signal propagation speed, whereupon inhibition is rapidly engaged, followed by balanced activity[31,32]; (3) net excitation or inhibition might occur when the afferents drive the overall excitation–inhibition balance in opposite directions; and (4) responses to large sustained input changes may occur while maintaining a well balanced excitation–inhibition. In the latter case, experimentally induced hyperpolarization of pyramidal cells may abolish their spiking without affecting the barrages of postsynaptic potentials (see ref. 31 and references therein). It is reasonable to assume that any similar hyperpolarization under normal conditions would decrease spiking of stimulus-selective neurons without affecting presynaptic activity. In visual cortex, recurrent connections among spiny stellate cells in the input layers can provide a significant source of recurrent excitation[26]. If driven by proportional excitation–inhibition synaptic currents, the impact of their sustained activity might, once again, minimally change the spiking of the pyramidal cells. This last property of microcircuits suggests that changes with balanced excitation–inhibition are good candidates for mechanisms adjusting the overall excitability and the signal-to-noise ratio (SNR) of the cortical output. Thus microcircuits—depending on their mode of operation—can, in principle, act either as drivers, faithfully transmitting stimulus-related information, or as modulators, adjusting the overall sensitivity and context-specificity of the responses[28]. Figure 2b summarizes the different types of excitation-inhibition changes and their potential effect on the haemodynamic responses.

This interesting and important driver/modulator distinction was initially drawn in the thalamus[33], in which the afferents in the major sensory thalamic relays were assigned to one of two major classes on the basis of the morphological characteristics of the axon terminals, the synaptic relationships and the type of activated receptors, the

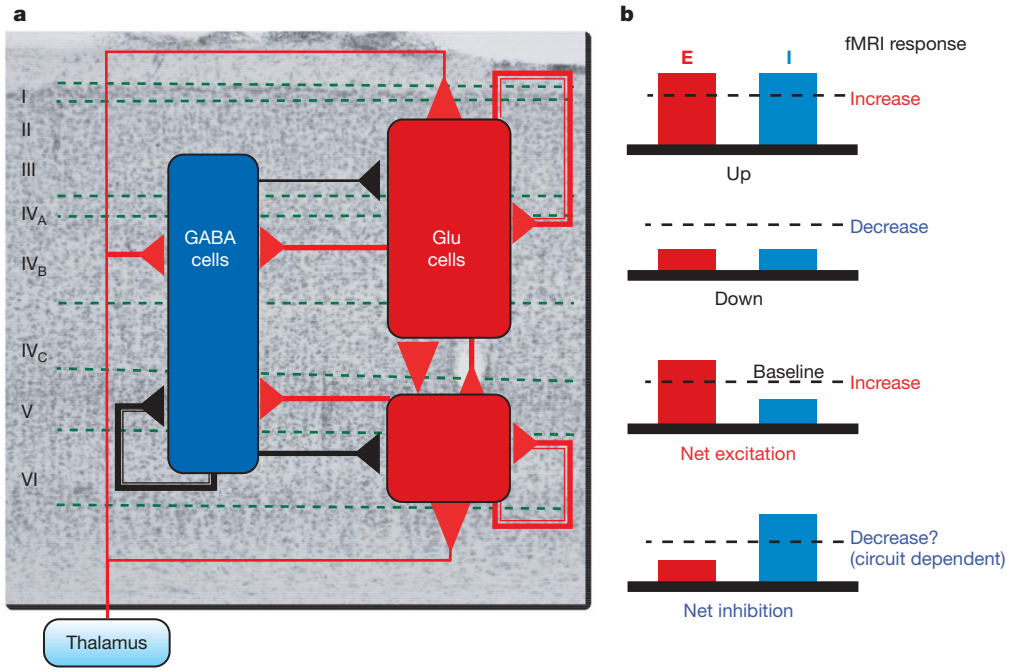

Figure 2 | Principles of excitation–inhibition circuits. a, Model of a canonical cerebral microcircuit (adapted from ref. 71). Three neuronal populations interact with each other: supragranular–granular and infragranular glutamatergic spiny neurons, and GABAergic cells. Excitatory synapses are shown in red and inhibitory synapses in black. All groups receive excitatory thalamic input. The line width indicates the strength of connection. The circuit is characterized by the presence of weak thalamic input and strong recurrence (see text for details). Glu, glutamatergic. **b**, Potential proportional and opposite-direction changes of cortical excitation (E) and inhibition (I). Responses to large sustained input changes may occur while maintaining a well balanced excitation–inhibition (up and down). The commonly assumed net excitation or inhibition might occur when the afferents drive the overall excitation–inhibition balance in opposite directions. The balanced proportional changes in excitation–inhibition activity, which occur as a result of neuromodulatory input, are likely to strongly drive the haemodynamic responses.

degree of input convergence, and the activity patterns of postsynaptic neurons. The same concept also broadly applies to the afferents of the cerebral cortex[34], wherein the thalamic or corticocortical axons terminating in layer IV can be envisaged as drivers, and other feedback afferents terminating in the superficial layers as modulators. It can also be applied to the cortical output, whereby the projections of layer VI back to the primary relays of the thalamus are modulatory, whereas the cortico-thalamo-cortical paths originating in layer V of cortex, reaching higher-order thalamic nuclei (for example, pulvinar), and then re-entering cortex via layer IV, are drivers[33].

The initial information reaching a cortical region is elaborated and evaluated in a context-dependent manner, under the influence of strong intra- and cross-regional cortical interactions. The cortical output reflects ascending input but also cortico-thalamo-cortical pathways, whereas its responsiveness and SNR reflect the activity of feedback, and likely input from the ascending diffuse systems of the brain-stem. The neuromodulation (see 'Neurotransmission and neuromodulation' in Supplementary Information) afforded by these systems, which is thought to underlie the altered states of cognitive capacities, such as motivation, attention, learning and memory, is likely to affect large masses of cells, and potentially induce larger changes in the fMRI signal than the sensory signals themselves.

Excitation–inhibition networks and fMRI. The organization discussed above evidently complicates both the precise definition of the conditions that would justify the assignment of a functional role to an 'active' area, and interpretation of the fMRI maps. Changes in excitation–inhibition balance—whether they lead to net excitation, inhibition, or simple sensitivity adjustment—inevitably and strongly affect the regional metabolic energy demands and the concomitant regulation of cerebral blood flow (CBF) (that is, they significantly alter the fMRI signal). A frequent explanation of the fMRI data simply assumes an increase in the spiking of many task- or stimulus-specific neurons. This might be correct in some cases, but increases of the BOLD signal may also occur as a result of balanced proportional increases in the excitatory and inhibitory conductances, potential concomitant increases in spontaneous spiking, but still without a net excitatory activity in stimulus-related cortical output. In the same vein, an increase in recurrent inhibition with concomitant decreases in excitation may result in reduction of an area's net spiking output, but would the latter decrease the fMRI signal? The answer to this question seems to depend on the brain region that is inhibited, as well as on experimental conditions.

Direct haemodynamic measurements with autoradiography suggested that metabolism increases with increased inhibition[35]. An exquisite example is the inhibition-induced increase in metabolism in the cat lateral superior olive (LSO). This nucleus, which contains the representations of low-, middle- and high-tone frequencies, receives afferents from both ears: over a two-neuron pathway from the ipsilateral ear and over a three-neuron pathway from the contralateral ear. Furthermore, it has no presynaptic axo-axonic endings that might mediate presynaptic inhibition via excitatory terminals. Electrophysiology showed that the LSO afferents from the ipsilateral ear are excitatory whereas the afferents from the contralateral ear are inhibitory. This unusual combination of anatomical and physiological features suggests that if one ear is surgically deafened and the animal is exposed to a high-frequency pure tone, a band of tissue in the LSO on the side opposite to the remaining active ear is subjected to strictly inhibitory synaptic activity without complications by presynaptic inhibition, concurrent lateral excitation, disinhibition/excitation, or other kinds of possibly excitatory action. Under these conditions, maps obtained with [^{14}C]2-deoxyglucose (2DG) autoradiography[36] demonstrated clear increases in metabolism in the contralateral LSO[37], suggesting that the presynaptic activity in that area is sufficient to show strong energy consumption despite the ensuing spiking reduction. Similar increases in metabolism during the reduction of spike rates were observed during long-lasting microstimulation of the fornix, which induces sustained suppression of pyramidal cell firing in hippocampus[38].

In contrast, human fMRI studies reported haemodynamic and metabolic downregulation accompanying neuronal inhibition in motor[39] and visual cortices[40], suggesting that the sustained negative BOLD response (NBR) is a marker of neuronal deactivation. Similarly, combined fMRI and electrophysiological experiments showed a clear correspondence of NBR and decreased population spiking in haemodynamically 'negative' areas in the monkey primary visual cortex[41]. Decreases in blood oxygenation and volume were also found to be co-localized with predominant neuronal inhibition and arteriolar vasoconstriction during somatosensory stimulation in rats[42]. Thus, without understanding the intrinsic correlation between direct or indirect inhibitory activity and concomitant changes in energy metabolism in a given situation, conclusions cannot be drawn. Unfortunately, the few published theoretical estimates of energy budget have not considered the metabolic costs of spikes in interneurons and of the inhibitory postsynaptic potentials (IPSPs) they produce[43]. Modelling of inhibition is unlikely to be straightforward. On the one hand, the density of cortical inhibitory neurons is 10–15 times lower than excitatory neurons[16], and for each one of them the electrochemical gradient, down which Cl^- moves postsynaptically at inhibitory synapses, is weaker than that of Na^+ at excitatory synapses, requiring less energy to pump Cl^- back. In fact, the transport cycles of the cation–chloride co-transporters, which have a key role in intracellular Cl^- regulation, are driven without the direct hydrolysis of ATP, by using the energy from the cation gradients generated by the Na,K-ATPase[44]. On the other hand, inhibitory interneurons are fast spiking[45,46]. For example, the firing of pyramidal cells in hippocampus is 1.4 Hz, whereas that of interneurons in the strata pyramidale and oriens is 15 Hz and 10 Hz, respectively. Similarly, cortical inhibitory interneurons may discharge 2–3 times faster than pyramidal cells[47]. In principle, inhibition may increase or decrease energy consumption depending on the contribution of the aforementioned factors (for a recent comprehensive review on inhibitory neurons and brain metabolism, see ref. 48). Last but not least, neurons directly affect microvessels. Pericytes, the flat, contractile connective-tissue cells, often attached to the abluminal surface of the capillary endothelial cells, might directly alter CBF in response to changes in neural activity[49]. Moreover, a body of evidence suggests that increased activity of single inhibitory interneurons results in precise vasomotor responses in neighbouring brain microvessels, and these contractile or dilatory responses were attributed to arteriole smooth muscle[50].

The diversity of the haemodynamic responses to neural inhibition obtained in different types of experiments is therefore hardly surprising: it is primarily due to the fact that regional inhibition itself might have a number of different causes, including early shunting of the weak cortical input, leading to a reduction of recurrent excitation rather than an increase in summed inhibition; increased synaptic inhibition; shunting of the cortical output through the axo-axonic connections of the chandelier cells; or any combination thereof. In the first case inhibition might result in a clear NBR; in the other two it might reflect the local metabolism increases induced by the unaffected input and its ongoing processing, resulting in fMRI activations. The fMRI responses might further blur the origin of inhibition owing to the direct effects of the latter on the arterioles and microvessels.

Evidently much research is needed to characterize the actual state of an area and its participation in behaviour, but quite independent of this fact, the nature of the EIN suggests that mass action and its surrogate haemodynamics are ambiguous signals, the interpretation of which must be constrained by the concurrent use of other methodologies.

Neurophysiological correlates of the BOLD signal

EIN and mesoscopic neural signals. The active regions of the membrane of a discharging neuron at any given time are considered to act as a current sink, whereas the inactive ones act as a current source for the active regions (see 'Neural signals' in Supplementary Information). The linear superposition of currents from all sinks and sources forms the extracellular field potential measured by microelectrodes. The extracellular field potential captures at least three different types of EIN activity: single-unit activity representing the action potentials of well isolated neurons next to the electrode tip, multiple unit activity reflecting the spiking of small neural populations in a sphere of 100–300 μm radius, and perisynaptic activity of a neural population within 0.5–3 mm of the electrode tip, which is reflected in the variation of the low-frequency components of the extracellular field potential. Multiple unit activity and local field potentials (LFPs) can be reliably segregated by frequency band separation. A high-pass filter cutoff in the range of 500–1,000 Hz is used in most recordings to obtain the multiple unit activity, and a low-pass filter cutoff of approximately 250 Hz to obtain LFP. A large number of experiments have presented data indicating that such a band separation does indeed underlie different neural events (see 'Neural signals' in Supplementary Information).

LFP signals and their different band-limited components (alpha, beta, gamma, and so on) are invaluable for understanding cortical processing, as they are the only signs of integrative EIN processes. In fact, LFPs do not, as initially thought, solely reflect population postsynaptic potentials, but also integrative soma–dendritic processes—including voltage-dependent membrane oscillations and afterpotentials following soma–dendritic spikes—that all together represent the local (perisynaptic) activity in a region (see 'Neural signals' in Supplementary Information). A shortcoming of the LFP is its ambiguity. A change in the power of LFP in a particular frequency band most likely occurs for any mode of operations of the EIN. As most of the excitatory input into an area is local, LFPs will also indirectly reflect some of the postsynaptic effects of pyramidal cell activity. In addition, LFPs have a certain neural-class bias, which in this case is determined by geometry and regional architecture. The arrangement of the pyramidal and Purkinje cells will give rise to large LFP modulations; in contrast, interneurons will contribute only weakly because of their star-shaped dendrites and their geometrical disorder. Finally, inhibitory synapses may occasionally act as 'shunts' for the excitatory currents through low-resistance channels, in which case large synaptic conductance changes may produce little effect in the membrane potential, and result in weak and hard-to-measure multiple unit activity and LFPs.

When individual LFP bands are examined separately, local spiking activity is occasionally found to affect certain frequency bands, whereas that of neuromodulation affects others[51–53]. It is evident that the most useful information will not be derived by one type of signal alone, but rather by the study of relative changes in one signal or the other. Electrophysiological studies examining the individual contributions of different LFP frequency bands, multiple unit activity, and spiking of individual neurons are probably our only realistic chance of gaining insights into the neural mechanisms of haemodynamic responses and their meaning in the context of different cognitive tasks.

Mesoscopic signals and the BOLD signal. The relationship of neocortical LFPs and spiking activity to the BOLD signal itself was examined directly in concurrent electrophysiology and fMRI experiments in the visual system of anaesthetized[54] and alert[55] monkeys. These studies found that the BOLD responses reflect input and intracortical processing rather than pyramidal cell output activity. Initially, both LFPs and spiking seemed to be correlated with the BOLD response, although quantitative analysis indicated that LFPs are better predictors of the BOLD response than multiple-unit or single-unit spiking. The decisive finding leading to the papers' conclusion, however, was not the degree of correlation between the neural and the fMRI responses or the differential contribution of any type of signal into the BOLD responses[55], but rather the striking, undiminished haemodynamic responses in cases where spiking was entirely absent despite a clear and strong stimulus-induced modulation of the field

potentials[54,55]. Similar dissociations between spikes and CBF had been demonstrated earlier and very recently in a number of studies using other techniques[56–58].

The findings are in close agreement with a number of older autoradiography studies, also showing that regional glucose utilization is directly related to neuronal synaptic activity[35]. For example, the greatest 2-DG uptake occurs in the neuropil (that is, in areas rich in synapses, dendrites and axons, rather than in cell bodies). During orthodromic and antidromic electrical microstimulation, only orthodromic microstimulation, which involves presynaptic terminals, increases glucose consumption. Similarly, the highest density of cytochrome oxidase (an enzyme of the respiratory chain) is found in somato-dendritic regions that are adjacent to axon terminals. Finally, as mentioned earlier, presynaptic activity increases metabolism even if the output is inhibited (that is, the spiking activity is abolished).

Despite all this evidence, some discussion still concentrates on the importance of the firing rate of action potentials of projection neurons in the generation of the haemodynamic responses, perhaps stemming from the fact that important early studies of neural correlates of behaviour took the mean spiking rate to be the gold standard for quantifying neuronal activation. These discussions, however, often suffer from a certain amount of contention seeking where none is warranted. In many cases, spikes do indeed correlate with LFPs, and they will also correlate with the BOLD signal. In addition, unusually high correlations between multiple unit activity and BOLD signal (or LFP and multiple unit activity) may result from excessive signal-smoothing owing to sampling rates of several seconds rather than a fraction of a second, as well as inter-subject averaging when simultaneous physiology and fMRI measurements are not possible (see ref. 55 for discussion).

Predicting neural activity from the fMRI signals. Functional MRI signals are presumed to result from changes in the activity of the neuronal populations responsible for the functions in question (for example, stimulus- or task-selective neurons). This assumption is mainly based on decades of electrophysiology research with recordings from isolated single neurons in experimental animals, in which particular sensory stimuli that the animal perceives or tasks that it performs were found to increase the firing rate of certain cells but not of others. The psychologist or cognitive neuroscientist who finds cortical area X to be activated by the task at hand implicitly or explicitly assumes that—if an electrode were placed in the subject's brain—an increase in the spiking rate of those specialized neurons underlying the subject's behaviour would be observed. This might well be true in some cases, but not in all. When attempting to interpret the fMRI signal by modelling, or when comparing the results of human neuroimaging to those obtained in monkey physiology experiments, it is useful to take the following facts into consideration.

In humans, there are about 90,000–100,000 neurons under 1 mm^2 of cortical surface. This number is relatively constant for all structurally and functionally distinct areas, including the somatosensory, temporal, parietal, frontal and motor cortical areas[16,59]. An exception is the primary visual cortex of certain primates, including monkey and human, which has approximately twice as many neurons. The number of cortical neurons under unitary cortical surface is also similar across many species, including mouse, rat, cat, monkey and human. Its small variability is the result of a trade-off between cortical thickness and neural density. The former varies from area to area and from species to species (for example, from mouse to human the cortex becomes approximately three times thicker). Neural density varies inversely to cortical thickness. On average, density is 20,000 to 30,000 neurons per mm^3; it peaks in the primary visual cortex by a factor of 4, and it is minimal in the motor cortex[59,60]. Synaptic density ranges from 0.4 to 1×10^9 per mm^3. Depending on the thickness of the cortex (2–4 mm), the number of synapses beneath 1 mm^2 surface is around 10^9 ($0.8–4 \times 10^9$). Although the number of synapses and the axonal length per neuron increases with increasing cortical thickness[61], the overall length of neuronal processes remains relatively constant, with axonal length being approximately 4 km mm^{-3} and dendrite length 0.4 km mm^{-3}. Overall, synaptic density and the ratio of excitatory to inhibitory synapses also remain constant.

Given these neuro-statistical data, what are the actual contents of a neuroimaging voxel? An examination of the 300 top-cited cognitive fMRI studies suggests that the commonly used in-plane resolution is 9–16 mm^2, for slice thicknesses of 5–7 mm. The average voxel size before any pre-processing of the data is thus 55 μl (or 55 mm^3). Often the effective size is 2–3 times larger due to the spatial filtering that most investigators apply to improve the functional SNR. Less than 3% of this volume is occupied by vessels and the rest by neural elements (see Fig. 3) A typical unfiltered fMRI voxel of 55 μl in size thus contains 5.5 million neurons, $2.2–5.5 \times 10^{10}$ synapses, 22 km of dendrites and 220 km of axons.

This 'large population view' is in contrast to the scope of the traditional microelectrode recordings. It would be nice if we could monitor every relevant neuron in the cortex during intracortical microelectrode recordings, but this is practically impossible. Instead, the typical electrophysiological measurements in behaving animals report only on the properties of most active large neurons that constitute a minority. The strong selection bias during extracellular recordings is partly due to practical limitations (for example, injury or simply size bias[62]) and partly to the physiological properties of neurons and/or the organizational principles of neural networks. In fact, many different types of electrical and optical measurements provide evidence that a substantial proportion of neurons, including the cortical pyramidal cells, might be silent[63]. Their silence might reflect unusually high input selectivity or the existence of decoding schemes relying on infrequent co-spiking of neuronal subsets. Most important for the comparison of neuroimaging and electrophysiology results is the fact that lack of measurable neuronal spiking may not necessarily imply lack of input and subthreshold processing.

A direct analogy between neuronal spiking as measured in animal experiments and the fMRI signal obtained in human recording is thus simply unrealistic and might often lead to incorrect conclusions. It is hardly surprising that most studies so far relying purely on BOLD fMRI have failed to reveal the actual neural properties of the studied area, at least those properties (for example, selectivity to various visual features) that were previously established in electrophysiological studies.

An example is cortical area V5 (or MT) that has been extensively studied in the context of motion processing and perception[64,65]. Electrophysiology has shown that the vast majority of the V5 neurons in monkeys are direction and speed selective. Neuroimaging localized the homologue of area V5 in humans as an area responding stronger to moving than to stationary stimuli. Later studies suggested that human V5 is sensitive to motion direction, and that it may be thought of as containing large populations of directionally selective units, just like its monkey homologue. The studies of directional specificity exploited the phenomenon of motion after-effect induced by motion adaptation. After prolonged exposure to a stimulus moving in one direction, subjects perceive a subsequent static stimulus to move in the opposite direction. It is assumed that motion after-effect is due to the fact that the balance of mutual inhibition (opponency) between detectors for opposite directions of movement is distorted after adaptation. The sensitivity of the detectors selective for the adapting direction is reduced, which in turn releases from inhibition the neurons selective for the opposite direction[66]. Using this phenomenon, human studies demonstrated that the fMRI response to a stationary stimulus was greater when the stimulus was preceded by a motion-after-effect-inducing, unidirectional adaptation, than when preceded by bidirectional adaptation[67]. Given the existing physiology data in the monkey V5, these findings were interpreted as demonstrating that the BOLD signal directly reflects direction-selective spiking activity of the area.

Yet, as I have indicated above, the BOLD signal is primarily affected by changes in excitation–inhibition balance, and this balance may be controlled by neuromodulation more than by the changes in spiking rate of a small set of neurons. In fact, the BOLD signal is strongly modulated by attention[68], and the results of the motion after-effect experiments could, in principle, be due to the fact that a stimulus with illusory motion automatically draws the attention of a subject more compared to a situation in which there is no motion after-effect. This hypothesis turned out to be correct, as a later study—in which balance in attentional load was accomplished by having the subjects perform a concurrent visual task—found no signal differences between the motion after-effect and no motion after-effect conditions[69].

A similar example pertains to the differences in neurophysiological and fMRI responses in the primary visual cortex during different perceptual states. It is known that physiological signals are in general stronger when stimuli are perceived as opposed to when they are not. Intriguingly, in some regions the BOLD response seems to reflect this even more sensitively than physiological measures like spikes and multi-unit activity[70]. An example is the pattern of fMRI activation changes in V1 during binocular rivalry (that is, the perceptual alternations experienced when the two eyes view different stimuli). This phenomenon has been studied extensively psychophysically and also over the last two decades in a series of electrophysiology studies in monkeys[70]. These studies showed that only a small fraction of V1 cells modulate their spiking during the perceptual changes; neuroimaging, on the other hand, demonstrated fMRI-signal modulations that were nearly as large as those obtained during the physical alternation of stimuli[70]. The difference, once again, reflects the fact that neuromodulatory feedback from higher areas can be easily detected by means of fMRI, but not through the measurement of single-unit activity. Interestingly, measurements of subthreshold activity in another study of perceptual multistability revealed perception-related modulations in LFP, despite the unaltered spike rates[53]. Such clear spiking and BOLD signal mismatches appear even in simple experiments probing sensory processing. Simple stimuli, such as those used in the aforementioned studies, are most likely to generate a proportional enhancement in both the afferent and efferent activity of any sensory area. The activation of high-level association areas related to cognitive processing might be more sensitive or even dominated by feedback and neuromodulation, whose differential effect on spiking and haemodynamic responses is utterly unknown.

Conclusions and perspectives

The limitations of fMRI are not related to physics or poor engineering, and are unlikely to be resolved by increasing the sophistication and power of the scanners; they are instead due to the circuitry and

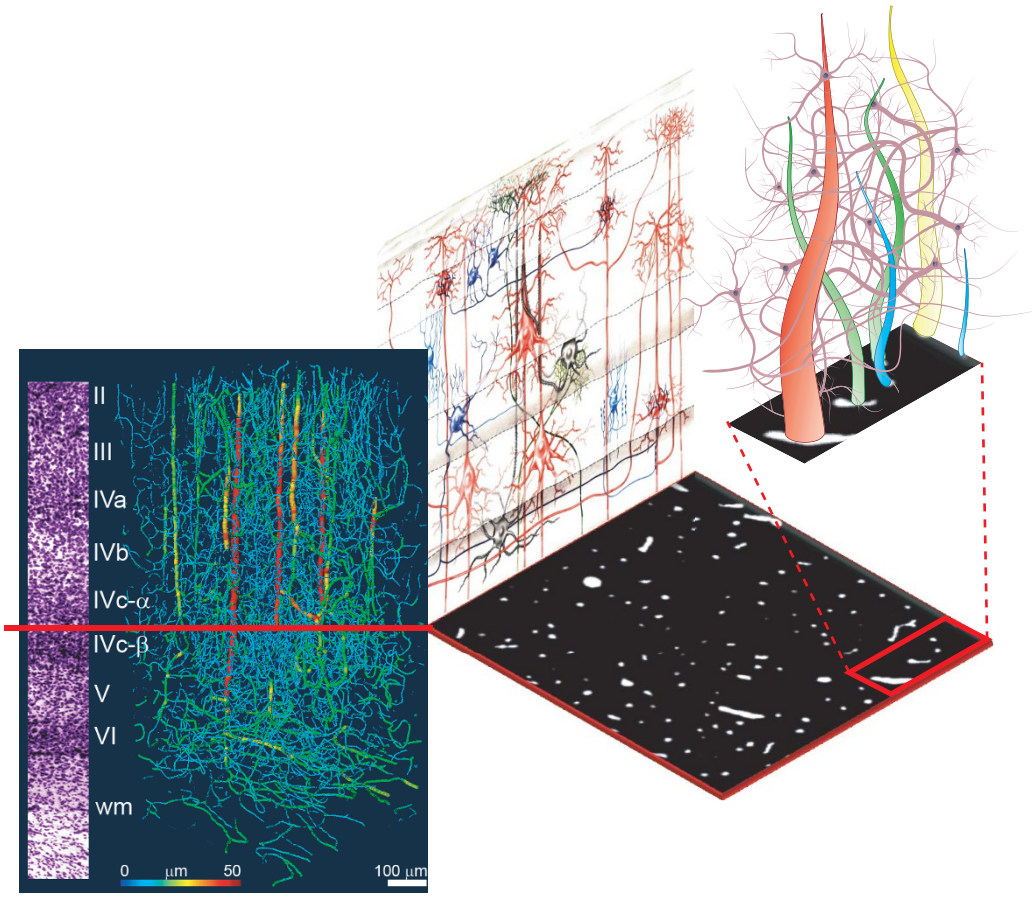

Figure 3 | Neural and vascular contents of a voxel. The left panel demonstrates the relative density of vessels in the visual cortex of monkeys. The dense vascular mesh is displayed by perfusing the tissue with barium sulphate and imaging it with synchrotron-based X-ray microtomography (courtesy B. Weber, MPI for Biological Cybernetics). The vessel diameter is colour coded. Cortical surface without pial vessels is displayed at the top; white matter at the bottom. At the left of the panel is a Nissl slice from the same area, showing the neural density for layers II through to the white matter (wm). Although the density of the vessels appears to be high in this three-dimensional representation, it is actually less the 3% (see section at the right; white spots are cross-sections of vessels). The average distance between the small vessels (capillaries) is about 50 μm. This is approximately the distance that oxygen molecules travel by diffusion within the limited transit time of the blood. The dense population of neurons, synapses and glia occupy the intervascular space, as depicted in the drawing at the top right—a hypothetical distribution of vascular and neural elements in a small section (red rectangle). The drawing in the background shows some of the typical neuronal types (for example, red, large pyramidal cell; dark blue, inhibitory basket cells; light blue, chandelier inhibitory neurons; and grey, stellate cells) and their processes.

functional organization of the brain, as well as to inappropriate experimental protocols that ignore this organization. The fMRI signal cannot easily differentiate between function-specific processing and neuromodulation, between bottom-up and top-down signals, and it may potentially confuse excitation and inhibition. The magnitude of the fMRI signal cannot be quantified to reflect accurately differences between brain regions, or between tasks within the same region. The origin of the latter problem is not due to our current inability to estimate accurately cerebral metabolic rate of oxygen ($CMRO_2$) from the BOLD signal, but to the fact that haemodynamic responses are sensitive to the size of the activated population, which may change as the sparsity of neural representations varies spatially and temporally. In cortical regions in which stimulus- or task-related perceptual or cognitive capacities are sparsely represented (for example, instantiated in the activity of a very small number of neurons), volume transmission (see Supplementary Information)—which probably underlies the altered states of motivation, attention, learning and memory—may dominate haemodynamic responses and make it impossible to deduce the exact role of the area in the task at hand. Neuromodulation is also likely to affect the ultimate spatiotemporal resolution of the signal.

This having been said, and despite its shortcomings, fMRI is currently the best tool we have for gaining insights into brain function and formulating interesting and eventually testable hypotheses, even though the plausibility of these hypotheses critically depends on used magnetic resonance technology, experimental protocol, statistical analysis and insightful modelling. Theories on the brain's functional organization (not just modelling of data) will probably be the best strategy for optimizing all of the above. Hypotheses formulated on the basis of fMRI experiments are unlikely to be analytically tested with fMRI itself in terms of neural mechanisms, and this is unlikely to change any time in the near future.

Of course, fMRI is not the only methodology that has clear and serious limitations. Electrical measurements of brain activity, including invasive techniques with single or multiple electrodes, also fall short of affording real answers about network activity. Single-unit recordings and firing rates are better suited to the study of cellular properties than of neuronal assemblies, and field potentials share much of the ambiguity discussed in the context of the fMRI signal. None of the above techniques is a substitute for the others. Today, a multimodal approach is more necessary than ever for the study of the brain's function and dysfunction. Such an approach must include further improvements to MRI technology and its combination with other non-invasive techniques that directly assess the brain's electrical activity, but it also requires a profound understanding of the neural basis of haemodynamic responses and a tight coupling of human and animal experimentation that will allow us to fathom the homologies between humans and other primates that are amenable to invasive electrophysiological and pharmacological testing. Claims that computational methods and non-invasive neuroimaging (that is, excluding animal experimentation) should be sufficient to understand brain function and disorders are, in my opinion, naive and utterly incorrect. If we really wish to understand how our brain functions, we cannot afford to discard any relevant methodology, much less one providing direct information from the actual neural elements that underlie all our cognitive capacities.

1. Wandell, B. A., Brewer, A. A. & Dougherty, R. F. Visual field map clusters in human cortex. *Phil. Trans. R. Soc. Lond. B* **360,** 693–707 (2005).
   This paper provides a description of human visual field maps and the rationale generating and naming them.
2. Haacke, E. M. *et al. Magnetic Resonance Imaging: Principles and Sequence Design* (John Wiley & Son, New York, 1999).
3. Wood, M. L. & Wehrli, F. W. Principles of magnetic resonance imaging. In *Magnetic Resonance Imaging* 3rd edn (eds Stark, D. D. & Bradley, W.) 1–14 (Mosby, St Louis/Baltimore/Boston/London/Tokyo, 1999).
4. Buxton, R. B. *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques* (Cambridge Univ. Press, Cambridge, UK, 2002).
5. Ogawa, S. & Lee, T. M. Magnetic resonance imaging of blood vessels at high fields: *in vivo* and *in vitro* measurements and image simulation. *Magn. Reson. Med.* **16,** 9–18 (1990).
6. Ogawa, S. *et al.* Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.* **14,** 68–78 (1990).
7. Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70,** 909–919 (1993).
8. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77,** 24–42 (1997).
9. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci.* **1,** 153–170 (1989).
10. Friston, K. J. *et al.* The trouble with cognitive subtraction. *Neuroimage* **4,** 97–104 (1996).
11. Poeppel, D. A critical review of PET studies of phonological processing. *Brain Lang.* **55,** 317–351 (1996).
12. Buckner, R. L. *et al.* Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **93,** 14878–14883 (1996).
13. Krekelberg, B., Boynton, G. M. & van Wezel, R. J. Adaptation: from single cells to BOLD signals. *Trends Neurosci.* 29, 250–256 (2006).
    This paper discusses fMRI adaptation designs.
14. Friston, K. J. *et al.* Analysis of fMRI time-series revisited. *Neuroimage* **2,** 45–53 (1995).
15. Haynes, J. D. & Rees, G. Decoding mental states from brain activity in humans. *Nature Rev. Neurosci.* **7,** 523–534 (2006).
16. Braitenberg, V. & Schuez, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* 2nd edn (Springer, Berlin, 1998).
17. Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27,** 419–451 (2004).
    This paper provides a review of cortical microcircuits.
18. Ullman, S. Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* **5,** 1–11 (1995).
19. Friston, K. A theory of cortical responses. *Phil. Trans. R. Soc. B* **360,** 815–836 (2005).
20. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in primate cerebral cortex. *Cereb. Cortex* **1,** 1–47 (1991).
21. Douglas, R. J. & Martin, K. A. Mapping the matrix: the ways of neocortex. *Neuron* **56,** 226–238 (2007).
22. Freund, T. F. Interneuron diversity series: Rhythm and mood in perisomatic inhibition. *Trends Neurosci.* **26,** 489–495 (2003).
23. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5,** 793–807 (2004).
    This paper is a review on the various types of interneuron.
24. DeFelipe, J. Types of neurons, synaptic connections and chemical characteristics of cells immunoreactive for calbindin-D28K, parvalbumin and calretinin in the neocortex. *J. Chem. Neuroanat.* **14,** 1–19 (1997).
25. Szabadics, J. *et al.* Excitatory effect of GABAergic axo-axonic cells in cortical microcircuits. *Science* **311,** 233–235 (2006).
26. Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269,** 981–985 (1995).
27. Shadlen, M. N. & Newsome, W. T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4,** 569–579 (1994).
28. Chance, F. S., Abbott, L. F. & Reyes, A. D. Gain modulation from background synaptic input. *Neuron* **35,** 773–782 (2002).
29. Brunel, N. & Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* **11,** 63–85 (2001).
30. Steriade, M., Timofeev, I. & Grenier, F. Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* **85,** 1969–1985 (2001).
31. McCormick, D. A., Shu, Y. S. & Hasenstaub, A. Balanced recurrent excitation and inhibition in local cortical networks. in *Excitatory-Inhibitory Balance: Synapses, Circuits, Systems* (ed. Hensch, T.) (Kluver Academic Press, New York, 2003).
32. Haider, B., Duque, A., Hasenstaub, A. R. & McCormick, D. A. Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* 26, 4535–4545 (2006).
    This paper provides a demonstration of the regulation of excitation–inhibition balance changes *in vivo*.
33. Sherman, S. M. & Guillery, R. W. *Exploring the Thalamus and its Role in Cortical Function* 2nd edn (MIT Press, Cambridge, Massachusetts, 2006).
34. Crick, F. & Koch, C. Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* **391,** 245–250 (1998).
35. Jueptner, M. & Weiller, C. Review: does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. *Neuroimage* **2,** 148–156 (1995).
36. Sokoloff, L. *et al.* The [^{14}C]deoxyglucose method for the measurement of local cerebral glucose utilization: Theory, procedure and normal values in the conscious and anesthetized albino rat. *J. Neurochem.* **28,** 897–916 (1977).
37. Nudo, R. J. & Masterton, R. B. Stimulation-induced [^{14}C]2-deoxyglucose labeling of synaptic activity in the central auditory system. *J. Comp. Neurol.* **245,** 553–565 (1986).

38. Ackermann, R. F., Finch, D. M., Babb, T. L. & Engel, J. Jr. Increased glucose metabolism during long-duration recurrent inhibition of hippocampal pyramidal cells. *J. Neurosci.* **4,** 251–264 (1984).
39. Stefanovic, B., Warnking, J. M. & Pike, G. B. Hemodynamic and metabolic responses to neuronal inhibition. *Neuroimage* **22,** 771–778 (2004).
40. Shmuel, A. *et al.* Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron* **36,** 1195–1210 (2002).
41. Shmuel, A., Augath, M., Oeltermann, A. & Logothetis, N. K. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neurosci.* **9,** 569–577 (2006).
42. Devor, A. *et al.* Suppressed neuronal activity and concurrent arteriolar vasoconstriction may explain negative blood oxygenation level-dependent signal. *J. Neurosci.* **27,** 4452–4459 (2007).
43. Attwell, D. & Gibb, A. Neuroenergetics and the kinetic design of excitatory synapses. *Nature Rev. Neurosci.* **6,** 841–849 (2005).
44. Payne, J. A., Rivera, C., Voipio, J. & Kaila, K. Cation-chloride co-transporters in neuronal communication, development and trauma. *Trends Neurosci.* **26,** 199–206 (2003).
45. McCormick, D. A., Connors, B. W., Lighthall, J. W. & Prince, D. A. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J. Neurophysiol.* **54,** 782–806 (1985).
46. Buzsaki, G., Geisler, C., Henze, D. A. & Wang, X. J. Interneuron diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends Neurosci.* **27,** 186–193 (2004).
47. Wang, Y., Gupta, A., Toledo-Rodriguez, M., Wu, C. Z. & Markram, H. Anatomical, physiological, molecular and circuit properties of nest basket cells in the developing somatosensory cortex. *Cereb. Cortex* **12,** 395–410 (2002).
48. Buzsaki, G., Kaila, K. & Raichle, M. Inhibition and brain work. *Neuron* **56,** 771–783 (2007).
49. Peppiatt, C. M. *et al.* Bidirectional control of CNS capillary diameter by pericytes. *Nature* **443,** 700–704 (2006).
50. Hamel, E. Perivascular nerves and the regulation of cerebrovascular tone. *J. Appl. Physiol.* **100,** 1059–1064 (2006).
51. Kayser, C. & Konig, P. Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials. *Eur. J. Neurosci.* **19,** 485–489 (2004).
52. Liu, J. & Newsome, W. T. Local field potential in cortical area MT: Stimulus tuning and behavioral correlations. *J. Neurosci.* **26,** 7779–7790 (2006).
53. Wilke, M., Logothetis, N. K. & Leopold, D. A. Local field potential reflects perceptual suppression in monkey visual cortex. *Proc. Natl Acad. Sci. USA* **103,** 17507–17512 (2006).
54. Logothetis, N. K. *et al.* Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412,** 150–157 (2001).
55. Goense, J. B. M. & Logothetis, N. K. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biol.* **18,** 631–640 (2008).
56. Mathiesen, C., Caesar, K., Akgoren, N. & Lauritzen, M. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J. Physiol.* **512,** 555–566 (1998).
57. Viswanathan, A. & Freeman, R. D. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. *Nature Neurosci.* 10, 1308–1312 (2007).
This paper provides a demonstration of the coupling between $CMRO_2$ and the LFP.
58. Rauch, A., Rainer, G. & Logothetis, N. K. The effect of a serotonin-induced dissociation between spiking and perisynaptic activity on BOLD functional MRI. *Proc. Natl Acad. Sci. USA* **105,** 6759–6764 (2008).
59. Rockel, A. J., Hiorns, R. W. & Powell, T. P. The basic uniformity in structure of the neocortex. *Brain* **103,** 221–244 (1980).
60. Cragg, B. G. The density of synapses and neurones in the motor and visual areas of the cerebral cortex. *J. Anat.* **101,** 639–654 (1967).
61. Schuz, A. & Demianenko, G. P. Constancy and variability in cortical structure. A study on synapses and dendritic spines in hedgehog and monkey. *J. Hirnforsch.* **36,** 113–122 (1995).
62. Logothetis, N. K. & Wandell, B. A. Interpreting the BOLD signal. *Annu. Rev. Physiol.* **66,** 735–769 (2004).
63. Shoham, S., O'Connor, D. H. & Segev, R. How silent is the brain: is there a "dark matter" problem in neuroscience? *J. Comp. Physiol. A* **192,** 777–784 (2006).
64. Born, R. T. & Bradley, D. C. Structure and function of visual area MT. *Annu. Rev. Neurosci.* **28,** 157–189 (2005).
65. Zeki, S. Thirty years of a very special visual area, area V5. *J. Physiol.* **557,** 1–2 (2004).
66. Mather, G., Verstraten, F. A. & Anstis, S. M. *The Motion Aftereffect: a Modern Perspective* (MIT Press, Cambridge, Massachusetts, 1998).
67. Tootell, R. B. H. *et al.* Visual motion aftereffect in human cortical area MT revealed by functional magnetic-resonance-imaging. *Nature* **375,** 139–141 (1995).
68. Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L. & Petersen, S. E. Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* **248,** 1556–1559 (1990).
69. Huk, A. C., Ress, D. & Heeger, D. J. Neuronal basis of the motion aftereffect reconsidered. *Neuron* **32,** 161–172 (2001).
70. Blake, R. & Logothetis, N. K. Visual competition. *Nature Rev. Neurosci.* **3,** 13–21 (2002).
71. Douglas, R. J., Martin, K. A. C. & Whitteridge, D. A canonical microcircuit for neocortex. *Neural Comput.* **1,** 480–488 (1989).
72. Bartels, A., Logothetis, N. K. & Moutoussis, K. FMRI and its interpretations: An illustration on directional sensitivity in area V5/MT. *Trends Neurosci.* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank my co-workers A. Bartels, J. Goense, M. Munk and A.-C. Zappe for discussions; my colleagues P. Hoffmann, C. Koch, K. Martin, A. Schüz, C. Kayser and R. Turner for their insightful comments and suggestions on the latest version of the article; J. Goense, B. Weber and A. L. Keller for providing graphics; and D. Blaurock for language corrections. The work is supported by the Max Planck Society.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to N.K.L. (nikos.logothetis@tuebingen.mpg.de).